

University of Groningen

DTFE

Schaap, Willem Egbert

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:
2007

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Schaap, W. E. (2007). DTFE: the Delaunay Tessellation Field Estimator. [s.n.].

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

RIJKSUNIVERSITEIT GRONINGEN

DTFE

The Delaunay Tessellation Field Estimator

PROEFSCHRIFT

ter verkrijging van het doctoraat in de
Wiskunde en Natuurwetenschappen
aan de Rijksuniversiteit Groningen
op gezag van de
Rector Magnificus, dr. F. Zwarts,
in het openbaar te verdedigen op
vrijdag 19 januari 2007
om 14:45 uur

door

Willem Egbert Schaap

geboren op 28 september 1974
te Emmen

Promotores:
Prof. dr. M.A.M. van de Weijgaert
Prof. dr. T.S. van Albada

Beoordelingscommissie:
Prof. dr. E.A. Bergshoeff
Prof. dr. B.J.T. Jones
Prof. dr. R.H. Sanders

ISBN 90-367-2886-x

Voor mijn ouders

Cover illustration:

DTFE reconstruction of the large scale matter distribution in the universe.

Contents

Chapter 1 Introduction 11

1.1	The hot Big Bang	12
1.2	The large scale structure of the universe	15
1.3	The formation and evolution of cosmic structures	18
1.3.1	Primordial conditions	20
1.3.2	The linear regime	21
1.3.3	The non-linear regime	22
1.4	Current analysis tools and their shortcomings	26
1.5	Goals and outline of this thesis	29

Chapter 2 The Delaunay Tessellation Field Estimator 33

2.1	Introduction	34
2.2	The Delaunay Tessellation Field Estimator	34
2.3	Analysis of a cosmological N -body simulation	36

Chapter 3 Continuous Fields and Discrete Samples: Reconstruction through Tessellations 41

3.1	Introduction	42
3.2	Spatial point processes	45
3.2.1	Homogeneous Poisson point processes	46
3.2.2	Inhomogeneous Poisson point processes	46
3.2.3	Homogeneous binomial random point processes	47
3.2.4	Inhomogeneous binomial random point processes	47
3.3	Voronoi and Delaunay tessellations	47
3.3.1	Voronoi tessellations	48
3.3.2	Delaunay tessellations	49
3.4	Reconstruction of continuous fields	50
3.4.1	Mass-weighted versus volume-weighted reconstructions	51
3.4.2	Reconstruction versus filtering	52

3.5	Multi-dimensional interpolation:	
	tessellation-based field reconstruction schemes	52
3.5.1	Zeroth-order interpolation	53
3.5.2	First-order interpolation	54
3.5.3	Spline interpolation	55
3.5.4	Natural neighbor interpolation	56
3.6	Reconstruction of density fields	56
3.6.1	Grid-based reconstruction schemes	57
3.6.2	SPH-like reconstruction schemes	59
3.7	Tessellation-based reconstruction of density fields:	
	the Voronoi Tessellation Field Estimator	61
3.7.1	The VTFE procedure	62
3.7.2	Mass conservation	62
3.8	The Delaunay Tessellation Field Estimator	64
3.8.1	The DTFE procedure	64
3.8.2	Mass conservation	69
3.8.3	Higher-order density estimators	71
3.9	Effective smoothing kernels and resolution	71
3.9.1	The TSC kernel	72
3.9.2	The SPH kernel	72
3.9.3	The VTFE and DTFE kernels	72
3.9.4	Comparison	73
3.10	Processing of reconstructed fields	76
3.10.1	Profile and image construction	76
3.10.2	Filtering	77
3.10.3	Feature detection	79
3.10.4	Statistical processing	79
3.11	Computational considerations	80
3.12	Summary, discussion and outlook	81
3.A	The walking triangle algorithm	85
3.A.1	The construction of the adjacency matrix	87

Chapter 4 Resolving Complex Geometric Patterns and Hierarchical Substructure in the Cosmic Web 89

4.1	Introduction	90
4.1.1	The cosmic web	91
4.1.2	Analyzing the cosmic web	92
4.1.3	The Delaunay Tessellation Field Estimator	95
4.1.4	The DTFE and cosmic patterns	96
4.1.5	Plan and organization	97
4.2	Substructure and structural hierarchy: the Soneira-Peebles model	98
4.2.1	The Soneira-Peebles model	98
4.2.2	Realizations: singular Soneira-Peebles models	101
4.2.3	Realizations: extended Soneira-Peebles models	104

4.3	Soneira Peebles model: density field reconstructions	104
4.3.1	The $\eta = 4$ singular model	106
4.3.2	The $\eta = 2$ singular model	107
4.3.3	The extended Soneira-Peebles model	107
4.4	Soneira-Peebles model reconstructions: quantitative analysis	108
4.4.1	Singular models: one-point density distribution function	108
4.4.2	Singular models: PDFs	110
4.4.3	Extended models: Kolmogorov-Smirnov tests	115
4.4.4	Extended models: two-point correlation function	117
4.5	Web-like spatial patterns: Voronoi clustering models	119
4.5.1	The asymptotic cosmic web: Voronoi clustering models	120
4.5.2	Voronoi tessellations: the skeleton	121
4.5.3	Geometric clustering	122
4.5.4	Voronoi clustering models: formalism	123
4.5.5	Voronoi clustering models: realizations	125
4.6	Voronoi clustering models: density field reconstructions	126
4.6.1	Filaments	126
4.6.2	Walls and clusters	132
4.7	Voronoi clustering models: shape and morphology analysis	133
4.7.1	Shape and morphology	134
4.8	Voronoi clustering models: topology and volume occupation	136
4.8.1	Density field volume occupation: results	137
4.9	Summary and discussion	139
4.A	Density estimation	145
4.A.1	TSC density field reconstruction procedure	145
4.A.2	SPH density field reconstruction procedure	146
4.A.3	DTFE density field reconstruction procedure	146
4.B	Voronoi clustering model	147
4.B.1	Procedure outline	148
4.B.2	Initial conditions	148
4.B.3	Path within Voronoi cell	149
4.B.4	Path within Voronoi wall	149
4.B.5	Path along Voronoi edge	150
4.B.6	Voronoi vertex	150
4.C	Voronoi wall density and mass distribution	151
4.C.1	Single Voronoi wall	151
4.C.2	Multiple Voronoi walls	152
4.D	Voronoi cell volume distribution	152

Chapter 5 Density Estimators in Particle Hydrodynamics:

DTFE versus SPH 155

5.1	Introduction	156
5.2	DTFE and SPH density estimates	158
5.2.1	SPH density estimate	158

5.2.2	DTFE density estimate	159
5.2.3	Comparison	160
5.3	Case study: two-phase interstellar medium	160
5.3.1	Results	162
5.4	The DTFE particle method	165
5.5	Delaunay tessellations and ‘moving grid’ hydrocodes	166
5.6	Summary and discussion	167

Chapter 6 Foam Dynamics: on Matter and Motions around the Cosmic Foam 171

6.1	Introduction	172
6.2	Case study: a cosmic velocity field	174
6.3	DTFE reconstruction of velocity fields	176
6.3.1	Case study: DTFE reconstruction of a velocity field	177
6.3.2	Virtues and limitations	177
6.3.3	Divergence, shear and vorticity	179
6.4	GIF simulations: DTFE density and velocity field reconstructions	179
6.4.1	Velocity vector fields	180
6.4.2	The continuity equation	183
6.5	Voids	187
6.5.1	Structure and dynamics	187
6.5.2	Constraining cosmological parameters using void dynamics	189
6.5.3	Evolution	193
6.6	Filaments and walls	196
6.6.1	Structure and dynamics	196
6.6.2	Evolution	198
6.7	Shear and vorticity	198
6.7.1	Shear tensor and eigenvalues	200
6.8	Summary and discussion	204

Chapter 7 The Foamy Morphology of the 2dF Galaxy Distribution 207

7.1	Introduction	208
7.2	The data	211
7.3	DTFE: method	211
7.4	DTFE: properties	213
7.4.1	Advantages	213
7.4.2	Limitations	215
7.5	Galaxy surface density field reconstructions	215
7.6	Selection effects	220
7.6.1	Non-uniform sampling	220
7.6.2	Varying redshift completeness	220
7.6.3	Magnitude limit variations	222
7.6.4	Varying radial selection function	222
7.6.5	Itinerary	224

7.7	2-D galaxy surface density field reconstructions	225
7.8	3-D galaxy density field reconstructions	225
7.9	Summary and discussion	233

Chapter 8 Sampling Noise, Significance and Error Estimation 237

8.1	Introduction	238
8.2	Sampling noise and significance	238
8.2.1	Statistical properties of Poisson sampling noise	239
8.2.2	Significance of uniform fields	241
8.2.3	Significance of inhomogeneous fields	245
8.2.4	Significance of filtered fields	248
8.3	Sampling errors	250
8.3.1	Measurement errors: the location of the sampling points	250
8.3.2	Measurement errors: the weight of the sampling points	253
8.3.3	Systematic errors: inhomogeneous sampling	256
8.3.4	Systematic errors: sampling distortions	257
8.4	Summary and discussion	258

Chapter 9 Summary and Outlook 259

9.1	Summary	259
9.1.1	The Delaunay Tessellation Field Estimator	259
9.1.2	An atlas of the nearby universe	261
9.1.3	Numerical simulations of structure formation	261
9.1.4	The cosmic velocity field	261
9.1.5	Evolution and dynamics of the cosmic web	262
9.2	Virtues and limitations	262
9.2.1	Virtues	262
9.2.2	Limitations	265
9.3	Conclusions and outlook	267

Nederlandse samenvatting 269

List of Publications 285

Dankwoord 287

1

Introduction

Throughout the history of mankind people have wondered what the universe looks like, how it began, if it did indeed, and how it evolved into its present state. From the earliest civilizations onwards people have come up with many different, often religiously inspired cosmologies, aimed to explain our very existence and to put our world in a cosmic perspective. Not until the last centuries much progress was made in a scientific sense, be it through observations or theoretical modelling. Indeed, only during the last century a coherent theoretical framework has emerged which could be and for a large part has been tested by observations. In this framework, the *hot Big Bang theory*, the universe does have a beginning, some 13.7 billion years ago, from which point space, time and all energy contained within the universe originate. The very first moment of the Big Bang and the subsequent evolution of the universe during the first 10^{-43} second of its lifetime (the *Planck time*) is not understood in terms of a coherent physical theory. However, we do know that the observable universe emerged as an extremely dense and hot plasma of radiation and matter, confined within a very small volume. From this moment onwards the evolution of the universe as a whole is successfully described by the Big Bang theory.

Even though the Big Bang theory has been very successful in predicting and explaining a large number of observed phenomena there is still a number of unresolved questions of a rather fundamental nature. First there is the issue of *dark matter*. Observational constraints and theoretical considerations have lead us to believe that no more than 15% of the matter in the universe is in the form of ‘regular’ baryonic matter. This means that the remaining 85% must be of some other, non-baryonic kind whose nature is as yet not understood. This matter is only known to exist through its gravitational interaction. As we cannot see this missing matter it is usually referred to as dark matter. The presence of such a large amount of unobserved matter forms a major challenge for present day cosmology.

A further enigma is presented by the presence of a mysterious and elusive all-pervading *dark energy* which is thought to account for 73% of the total energy content of the universe. Even though speculations about the nature of this dark energy are plenty, it basically remains a mystery. This is illustrated by the fact that its theoretically predicted value is an embarrassing factor of 10^{118} larger than observed (see e.g. Weinberg 1989). Curiously, observations indicate that the energy density of matter and the energy density of the dark energy precisely add up

to the *critical energy density*, a very special value for which the geometry of the universe is exactly flat. The reasons for this peculiar coincidence are as yet not understood.

At present the universe contains a wealth of structures on all scales. Examples include the planet Earth we are inhabiting, the stars, the Milky Way as well as other galaxies. On a Megaparsec scale we find the largest structures presently known to us. Here galaxies are grouped into huge and roughly spherical concentrations which may contain up to thousands of galaxies, *galaxy clusters*. These very dense galaxy clusters are interconnected through highly anisotropic filamentary and wall-like structures, which are of intermediate densities. The largest of these may extend over more than a hundred Megaparsec and are called *superclusters*. Galaxy clusters and superclusters together enclose extended, roughly spherical regions which are almost devoid of galaxies, *voids*. Large redshift surveys have revealed that these structures have assembled into a foamy network, the *cosmic web* or the *cosmic foam*, which is pervading the observable universe. The emergence of these structures cannot be explained in the hot Big Bang theory sec, which assumes the universe to be perfectly isotropic and homogeneous. Instead, an extension of Big Bang theory is necessary, namely the postulation that in the early ages of the universe very small density fluctuations were present. Through the continued action of gravity these fluctuations are believed to have slowly evolved into the structures we presently observe. This extension of Big Bang theory is called *gravitational instability theory*.

Although on the basis of gravitational instability theory large numerical simulations of cosmological structure formation have been successful in reproducing the general appearance of the cosmic foam, its structural and morphological properties and evolution still remain largely unexplored. The same holds for the question of when, how and where galaxies have formed and evolved. Both issues involve a large number of physical processes, each acting over very different spatial and temporal scales. Although a large number of smart yet artificial techniques has been invoked to deal with this inherent complexity, current analysis and simulation techniques largely remain unable to resolve and accurately describe all relevant physical processes.

The aim of this thesis is to develop and test a new method for the analysis of highly complex point distributions, such as are encountered in numerical simulations and galaxy redshift surveys. This method, the *Delaunay Tessellation Field Estimator*, is based on the use of the Delaunay tessellation (Delone 1934) of a given spatial point distribution to form the basis of a natural and fully self-adaptive filter for discretely sampled fields. Our aim is for this new method to contribute to the solution of a number of the many questions which are still pervading present day cosmology. The questions addressed in this thesis primarily involve the analysis of the properties of the cosmic web, but the method is suited for addressing some of the more fundamental questions about our cosmological world view. This introduction provides some of the necessary astronomical background for the questions addressed in this thesis.

1.1 The hot Big Bang

Up to the beginning of the 20th century it was believed that the universe was static. In 1915, Albert Einstein proposed his theory of general relativity and the equations that describe the dynamics of the universe. Einstein found in those equations that the universe should be ex-

panding or contracting, something entirely incompatible with the prevailing notion of a static universe.

General relativity is a metric theory of gravity, for which gravity is a manifestation of the local curvature of space. One of the key assumptions of the Big Bang theory is that the universe is homogeneous and isotropic, the *cosmological principle*. Under this assumption, the only possible geometries are those described by the *Robertson-Walker metric*,

$$ds^2 = c^2 dt^2 - a^2(t) \left(\frac{dr^2}{1 - kr^2} + r^2 d\Omega^2 \right). \quad (1.1)$$

Here spherical coordinates (r, θ, ϕ) are used for describing spatial positions and $d\Omega^2 = d\theta^2 + \sin^2\theta d\phi^2$ is the square of the solid angle. The coordinates are comoving with the expansion of the universe, whose spatial extent is proportional to the *expansion factor* $a(t)$. This means that the position \mathbf{r} of a point can be written as $\mathbf{r} = a\mathbf{x}$, where \mathbf{x} is called the *comoving position*. For convenience the expansion factor has been normalized such that at present $a(t_0) = 1$. The *curvature parameter* k parametrizes the global geometry of the universe, commonly referred to as *closed* ($k > 0$), *flat* ($k = 0$) or *open* ($k < 0$).

Given the Robertson-Walker metric, the Einstein equations describe the dynamics of the universe. In 1922 Friedmann managed to write down and solve the equations of motion for general homogeneous and isotropic universe models. These *Friedmann equations* are given by

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3} \left(\rho + \frac{3p}{c^2} \right) + \frac{\Lambda}{3}; \quad (1.2)$$

$$\frac{\dot{a}^2}{a^2} = \frac{8\pi G\rho}{3} - \frac{kc^2}{a^2} + \frac{\Lambda}{3}. \quad (1.3)$$

Here G is Newton's gravitational constant, ρ the mass density, p the pressure and Λ the dark energy or *cosmological constant*.

Following his work, it was the Belgian priest Georges Lemaître who not only independently solved the same equations in 1927, but also realized its physical ramifications. Extrapolating backward in time he saw that an expanding universe should have had a beginning in an extremely hot and dense phase. With some measure of imagination he indicated this primordial state by the name of 'Primeval Atom'. Soon thereafter, in 1929, these theoretical ideas were confirmed when Edwin Hubble discovered that galaxies recede from us with a velocity which increases as they are located at a larger distance. The physical velocity $\dot{\mathbf{r}}$ of a galaxy can thus be written as $\dot{\mathbf{r}} = H\mathbf{r}$, in which $H(t) = \dot{a}/a$ is known as the *Hubble parameter* and $H\mathbf{r}$ as the *Hubble flow*. This universal relation is commonly known as *Hubble's law*. Without doubt this discovery has been one of the greatest scientific revolutions in human history, providing the first clear and conclusive evidence that our universe is not static, but expanding. In the past few years the accuracy of modern space instruments have lead to a convergence of the estimated present expansion rate of the universe to a value of $H_0 \approx 71 \text{ km/s/Mpc}$ (known as the *Hubble constant*).

With the new observational evidence of an expanding universe, a whole new set of questions confronted the astronomical community. What is the nature of this expansion? Will the expansion last forever? Will the universe expand gradually until reaching a dynamical equilibrium and remain in a static state forever? Will the universe reach a critical point from which

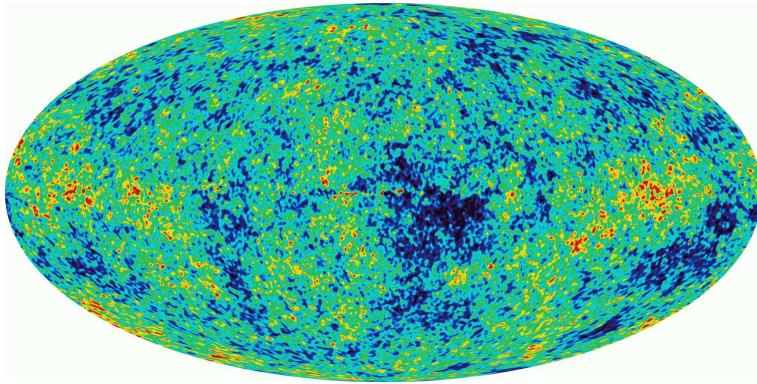


Figure 1.1 — An all-sky image of the infant universe, around three hundred thousand years after the Big Bang, as measured by WMAP. The color differences indicate the tiny temperature fluctuations induced by tiny matter, energy and velocity perturbations which have evolved into cosmic structures observed today. Courtesy: NASA/WMPA Science Team.

it will start to contract until reaching the opposite point of the Big Bang, the *Big Crunch*? The matter and energy content of the universe can be expressed in units of the *critical density*,

$$\rho_C = \frac{3H_0^2}{8\pi G}. \quad (1.4)$$

We may now define the following density parameters:

$$\Omega_m = \frac{\rho_m}{\rho_c}, \quad \Omega_\Lambda = \frac{\Lambda}{3H_0^2}, \quad \Omega_k = -\frac{kc^2}{a_0^2 H_0^2}, \quad (1.5)$$

which describe the density of matter, the energy density of the vacuum and the curvature of the universe. After some algebraic manipulations it follows that Eqn. 1.3 relates the curvature of the universe to the matter and energy density of the universe,

$$\Omega_k = 1 - \Omega_m - \Omega_\Lambda. \quad (1.6)$$

With these definitions the evolution of the universe is completely characterized by the cosmological parameters $(H_0, \Omega_m, \Omega_\Lambda)$. The special case of $\Omega_k = 1$ implies a flat universe. The flat universe with $\Omega_m = 1$ and $\Omega_\Lambda = 0$ is usually referred to as the *Einstein-De Sitter model*.

At present the hot Big Bang model is supported by a large amount of observational evidence. Amongst a range of tests, some four to five have become the most solid pillars of the Big Bang cosmology. Perhaps the most straightforward one is the fact that the Big Bang offers an explanation for a simple observation by Olbers in the early 19th century: the sky at night is dark! Only in a universe with a finite age and with a finite velocity of light this may be understood. Secondly, the discovery of Hubble's law forms a telling confirmation of the reality of an expanding universe. Maybe most tantalizing are two observational results which reach to much earlier epochs of our cosmos. The impressive precision with which the Hot Big Bang theory manages to predict the products of the very early phase of primordial nucleosynthesis of light chemical elements brings us back to the first three minutes of the universe.

The prediction of an isotropic blanket of cosmic thermal radiation, with a precise blackbody spectrum with a temperature of $T \approx 2.725 \text{ } ^\circ\text{K}$, enabled us to turn its discovery in 1965 into the final conclusive proof of the hot Big Bang's reality. The nickname coined by Fred Hoyle in 1950 in a BBC radio show had turned into a honorary title!

Some three minutes after the Big Bang, when the temperature of the universe had cooled down to a few billion degrees, the universe had turned into a gigantic nuclear reactor. The light chemical elements of Deuterium, ^3He lium and ^4He lium, as well as a trace of Lithium, managed to form before the cosmic expansion turned off its 'nuclear power'. The fact that the hot Big Bang offers an explanation for the abundances of these elements in our universe is one of its most convincing victories.

In the subsequent hundreds of thousands of years photons, in close interaction with a tiny residual of electrons and protons, reached equilibrium and distributed their energy into an almost perfect blackbody spectrum. After some 379 000 years the temperature of the universe had cooled to a mere $3000 \text{ } ^\circ\text{K}$, upon which protons and electrons managed to combine into hydrogen atoms, an event called *recombination*. Almost coincidental is the resulting *decoupling* of radiation and matter. No longer scattered by freely floating electrons, photons assumed a long journey along the depths of a virtually transparent universe. These photons of the *cosmic microwave background*, having retained their almost perfect blackbody spectrum, appear to have originated from the *surface of last scattering*. This surface marks the location of the atoms from which the observed cosmic microwave background photons were last dispersed and is shown in Fig. 1.1. Since then, the gradual expansion of the universe goes along with a proportional cooling down of the photon temperature, having reached a present day value of $T \approx 2.725 \text{ } ^\circ\text{K}$.

A continuously rising flood of new cosmological observations, stemming from larger and larger depths in the universe, has lead to a converging consensus regarding the universe we are living in. With impressive accuracy a rather remarkable set of values for the cosmological parameters has emerged, now commonly known by the name of '*concordance model*'[†]. Satellite experiments such as COBE and WMAP, in conjunction with balloon-borne experiments such as Boomerang, have managed to map in great detail the embryonic universe at around the recombination epoch (see e.g. Fig. 1.1). Not only did they show the validity of the hot Big Bang, but they also convinced us that the universe is flat, accurately determined its age to 13.7 billion years and independently confirmed the baryon content predicted by primordial nucleosynthesis.

1.2 The large scale structure of the universe

The cosmological principle states that the universe is isotropic and homogeneous. However, this is an adequate description only when looking at large enough scales ($\gg 100h^{-1} \text{ Mpc}$). At smaller scales the universe contains a wealth of structure and locally appears far from homogeneous or even isotropic. This becomes already clear from looking up to the sky at night, when a bright band of stars can be seen to run across the sky. The stars in this band all belong to the disk of our own Milky Way, a conglomeration of about 200 billion stars. Indeed, all stars in the universe seem to be located in galaxies, gravitationally bound groups of up to hundreds of billions of stars. From very deep optical images like the *Hubble Deep*

[†]A somewhat more modest name would be the *benchmark model* (Ryden 2003).

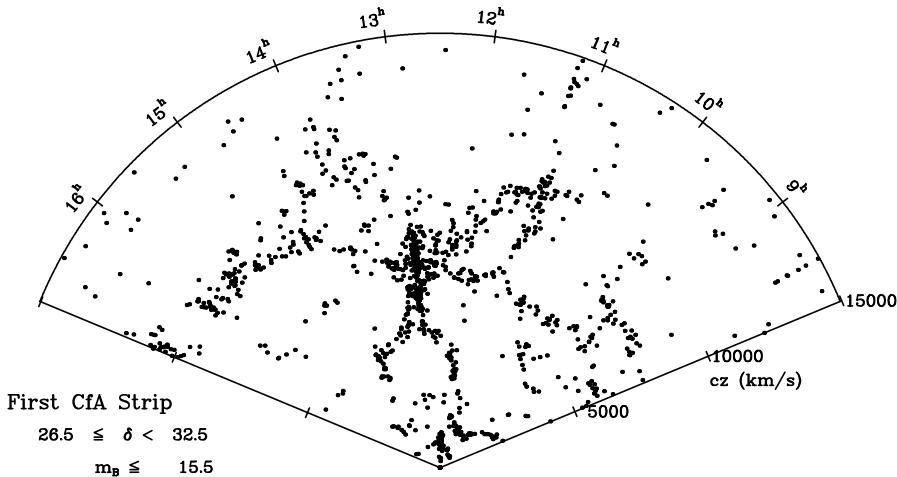


Figure 1.2 — The famous *Slice of the universe* that represents the first set of observations done for the CfA Redshift Survey in 1985 and which clearly showed that galaxies reside in structures on Megaparsec scales. The big structure in the center of the slice is the Coma cluster, while the Virgo cluster can be seen near the origin. Courtesy: Smithsonian Astrophysical Observatory.

Field we have learned that the visible universe contains about 100 billion galaxies. These galaxies are not isolated objects, but tend to cluster together. An example is the Local Group, a system which is dominated by the Milky Way and Andromeda, two giant spiral galaxies, which are both surrounded by a swarm of smaller dwarf and irregular galaxies. Such groups of galaxies are quite common and usually contain some tens of galaxies. In some cases many more galaxies are clustered in very dense and massive gravitationally bound *galaxy clusters*. The most massive of these may contain up to thousands of galaxies within a relatively small volume of only a few Megaparsec size. For instance, in the nearby Virgo and Coma clusters more than a thousand galaxies have been identified within a radius of a mere $1.5h^{-1}$ Mpc around their core. Due to their brightness, galaxy clusters are easily visible out to great distances and their existence has been known for a long time.

It is now twenty years ago that the first CfA2 redshift slice by de Lapparent, Geller & Huchra (1986, see Fig. 1.2) provided us with observational evidence that much larger structures than galaxy clusters exist, and that such structures seem to pervade throughout the observable universe. These extended, anisotropic matter concentrations are called *superclusters*. They usually comprise one or more rich galaxy clusters and a large number of groups and clumps of galaxies. Both our Local Group and the Virgo cluster are members of such a supercluster complex, the Local Supercluster. This a huge flattened concentration of about fifty groups of galaxies in which the Virgo cluster is the dominating and central agglomeration. A more prominent nearby example is the Perseus-Pisces supercluster. Its relative proximity ($\approx 55h^{-1}$ Mpc), its characteristic filamentary morphology and its favourable orientation have made it into one of the best mapped and meticulously studied superclusters. This majestic chain of galaxies has truly impressive proportions, a $5h^{-1}$ Mpc wide ridge of at least $50h^{-1}$ Mpc length, possibly extending out to a total length of $140h^{-1}$ Mpc. Along this major ridge we see a more or less continuous arrangement of high density clusters and

groups, of which the most notable ones are the Perseus cluster itself (Abell 462), Abell 342 and Abell 262. In addition to the presence of such huge filaments the galaxy distribution also contains vast planar assemblies. A striking example is the Great Wall, a huge planar assembly of galaxies with dimensions that are estimated to be of the order of $60 \times 170 \times 5h^{-1}$ Mpc. The Coma cluster of galaxies is its most prominent density enhancement (Geller & Huchra 1989).

Next to the high density clusters and more moderately dense superclusters redshift surveys have also revealed large regions which are almost devoid of galaxies. These *voids* are vast, roughly spherical regions with sizes in the order of $20 - 50h^{-1}$ Mpc. The earliest known example is the Boötes void (Kirshner et al. 1981, 1987), which has a diameter of $60h^{-1}$ Mpc and is almost completely empty (however, see Szomoru 1995).

Since the first CfA redshift slices increasingly larger and deeper redshift surveys have shaped our view of the large scale galaxy distribution. The most important milestone redshift surveys have been the already mentioned CfA survey (de Lapparent et al. 1986), the Las Campanas redshift survey (Schectman et al. 1996), the 2dF galaxy redshift survey (Colless et al. 2001) and the Sloan Digital Sky Survey (Stoughton et al. 2002), while also very deep pencil beam surveys such as by Broadhurst et al. (1990) and the VIMOS VLT Deep Survey (Le Fèvre et al. 2004) have contributed to our current view of the large scale galaxy distribution. In this view galaxy clusters, filamentary and wall-like structures and voids together form an intriguing cellular network which is pervading throughout the whole observable universe. In this foamy network the extended voids are surrounded by walls and filaments, whose nodes are formed by the dense galaxy clusters. This network is often called the *cosmic web* (Bond et al. 1996) or the *cosmic foam* (van de Weygaert et al. 1991, 2002). The cellular geometry of the cosmic web may be readily recognized in current redshift surveys, such as shown in Fig. 1.3, in which the galaxy distribution in the 2dF Galactic Redshift Survey is depicted.

An intriguing property of the cosmic matter distribution is its very high degree of complexity. It is characterized by an assembly of individual elements whose characteristic shapes, sizes and densities are very different. Most of these structures contain a high degree of substructure, and the embedded structures may have quite different properties from the overall structure. For example, filamentary and wall-like superclusters consist of a large number of smaller galaxy groups, each of which has a higher density than the average density of the supercluster as a whole. Zooming in on even smaller scales, galaxies themselves are frequently accompanied by a number of smaller satellites and dwarf galaxies. A well-known example is our own Milky Way, which is surrounded by two relatively large irregular satellites, the Magellanic clouds, as well as by a large number of dwarf galaxies (see e.g. Mateo 1998, Freeman & Bland-Hawthorn 2002). Galaxy clusters also tend to contain all kinds of substructure. An example is the Coma cluster, in which several dominant galaxies are present, each of which may be identified with a cluster core. In addition, subgroups of spiral galaxies with distinct structural and kinematical properties may be identified, possibly corresponding to infalling groups (Colless & Dunn 1996, Beijersbergen 2003). Even voids are known to display substructure. Small voids may be embedded within more extended voids (Regős & Geller 1991, Dubinski et al. 1993, van de Weygaert & van Kampen 1993, Szomoru et al. 1996, Sheth & van de Weygaert 2004).

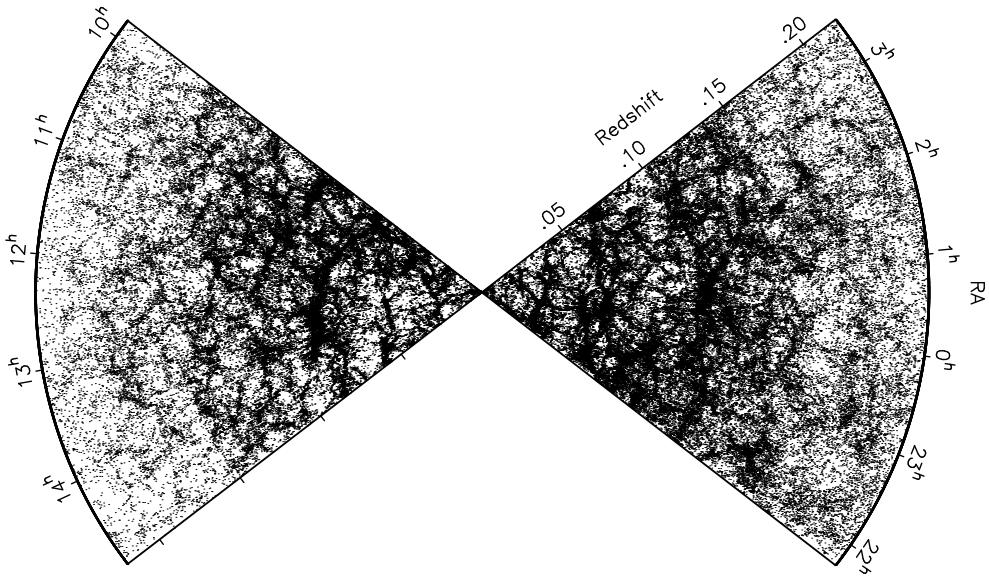


Figure 1.3 — The 2dF galaxy redshift survey. The foamy geometry of the cosmic web is strikingly displayed. Courtesy: the 2dF Galaxy Redshift Survey team.

1.3 The formation and evolution of cosmic structures

The Big Bang theory describes the evolution of the universe as a whole, a smooth and homogeneous medium. Although at sufficiently large scales the universe may indeed be considered homogeneous and isotropic, at smaller scales the universe contains all kinds of structures, which at Megaparsec scales form the cosmic web. This leaves us with the question of how such structures may have emerged in a universe which, according to the cosmological principle, is perfectly isotropic and homogeneous. For this purpose an extension of Big Bang theory is invoked, in which one assumes that at early stages the universe was not perfectly homogeneous, but instead small density fluctuations were about. Subsequently these fluctuations, under the influence of their mutual gravitational interaction, have grown into the wealth of structures which we observe today. The theory which describes the growth of these fluctuations is called *gravitational instability theory*.

The tiny primordial fluctuations are thought to originate from quantum fluctuations in the very early universe. A now standard ingredient of the theory of the evolution of our universe is *inflation* (Guth 1981). Shortly after the Big Bang the universe supposedly experienced a phase transition upon which it assumed a rapid exponential expansion. After this inflationary epoch ended the quantum fluctuations were strongly magnified into small but macroscopic density fluctuations and some regions in the universe contained somewhat less matter than their surroundings, while other regions contained somewhat more matter than their surroundings. Notice that nearly all relevant fluctuations had been blown up to scales larger than the horizon. Once the universe assumed a regular Friedmann expansion, the fluctuations gradually entered the horizon, upon which they started to grow.

Gravitational instability theory describes how these small density fluctuations with respect

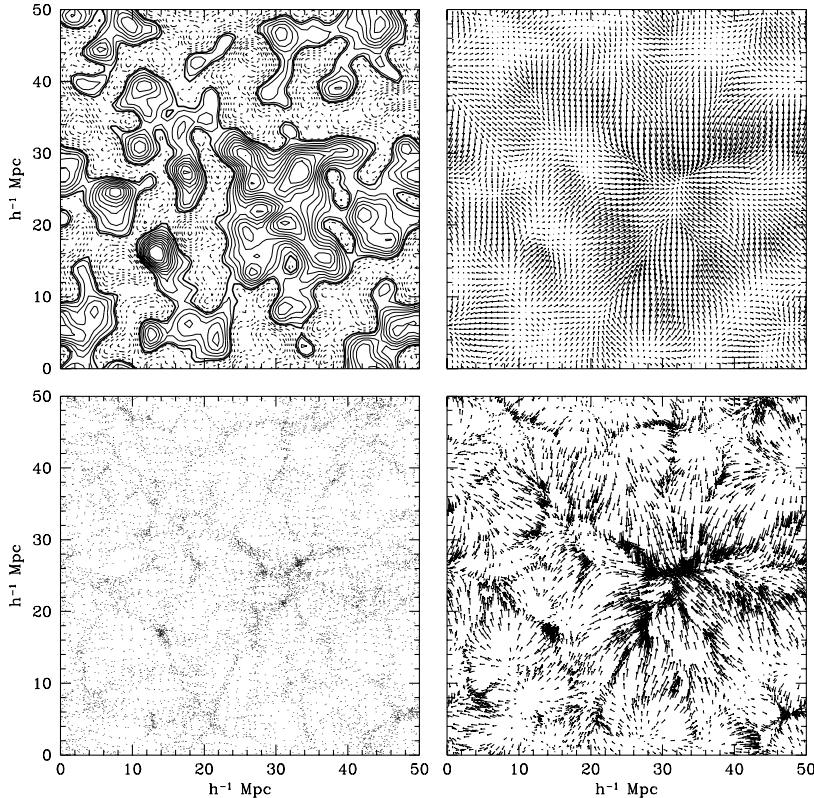


Figure 1.4 — Gravitational instability. Top left-hand frame: contour map of a Gaussian stochastic density field. Top right-hand frame: the resulting gravitational force field. Bottom left-hand frame: the resulting (non-linear) particle distribution. Bottom right-hand frame: vector map of the corresponding velocity field.

to the global cosmic background density and the accompanying tiny velocity perturbations from the general Hubble expansion have grown into the wealth of structures observed today. The process of gravitational instability is illustrated in Fig. 1.4. Given the physical position $\mathbf{r} = a\mathbf{x}$ of a particle, its physical velocity can be written as $\dot{\mathbf{r}} = \dot{a}\mathbf{x} + a\dot{\mathbf{x}} = H\mathbf{r} + \mathbf{v}$, where $\mathbf{v} = a\dot{\mathbf{x}}$ is the *peculiar velocity* of the particle. It is convenient to define the density fluctuation field $\delta(\mathbf{x})$ (Fig. 1.4, top left-hand frame) as

$$\delta(\mathbf{x}) = \frac{\rho(\mathbf{x}) - \rho_b}{\rho_b}. \quad (1.7)$$

In a statistical sense $\delta(\mathbf{x})$ may be considered as a random field, i.e. as a set of random variables, one for each point \mathbf{x} . In this view the universe is a random realization chosen from a statistical ensemble of universes and ρ_b , the background density, represents an average of the ensemble. The resulting total peculiar gravitational acceleration $\mathbf{g}(\mathbf{x})$ at position \mathbf{x} (Fig. 1.4, top right-hand frame) can be written as the sum of the peculiar gravitational attraction exerted by all

matter fluctuations in the universe,

$$\mathbf{g}(\mathbf{x}) = G \rho_b \int d\mathbf{x}' \frac{\delta(\mathbf{x}')(\mathbf{x}' - \mathbf{x})}{|\mathbf{x}' - \mathbf{x}|^3}. \quad (1.8)$$

This relation is basically established through the *Poisson equation*,

$$\nabla^2 \phi = 4\pi G \rho_b a^2 \delta. \quad (1.9)$$

This equation relates the density contrast δ with the gravitational potential perturbation ϕ , from which we obtain $\mathbf{g} = -\nabla\phi/a$. The formation of structure is then ascribed to the gravitational growth of these primordial density and velocity perturbations. Regions containing somewhat more matter than their surroundings exerted a slightly larger than average gravitational pull on their surroundings and started to accrete more matter. Because of this they grew even more overdense than their surroundings, exerting a stronger gravitational pull and attracting even more matter. This started a runaway process which continued until these regions contained so much matter that they collapsed under the influence of their own gravity and became gravitationally bound objects. Conversely, the regions which contained somewhat less matter than their surroundings exerted a smaller than average gravitational pull and therefore matter started to move out of these regions. Their gravitational pull thus became weaker and even more matter streamed out of these regions. Underdense region have thus become gradually emptier. They are thought to correspond with the void-like regions we presently observe. In this way matter migrated from underdense regions towards overdense regions (see bottom frames of Fig. 1.4).

For a pressureless medium the full evolution of this system of coupled cosmic density, velocity and gravity fields is encoded in a closed set of three coupled fluid equations. The Poisson equation (Eqn. 1.9) relates the matter distribution to the gravitational field. The *continuity equation* ensures the conservation of mass. It is given by

$$\frac{\partial \delta}{\partial t} + \frac{1}{a} \nabla \cdot (1 + \delta \mathbf{v}) = 0. \quad (1.10)$$

Finally, the *Euler equation* is the equation of motion which describes the induced matter flows. It is given by

$$\frac{\partial \mathbf{v}}{\partial t} + \frac{1}{a} (\mathbf{v} \cdot \nabla) \mathbf{v} + \frac{\dot{a}}{a} \mathbf{v} = -\frac{1}{a} \nabla \phi. \quad (1.11)$$

1.3.1 Primordial conditions

As we have seen primordial fluctuations originate from quantum fluctuations in the early universe. They were subsequently magnified by inflation and have since then grown by means of gravitational instability. An important property of inflation is that the resulting primordial density field is a Gaussian random field. This means that its set of N -point joint probabilities can be written as

$$\mathcal{P}_N = \frac{\exp\left[-\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \delta_i (M^{-1})_{ij} \delta_j\right]}{[(2\pi)^N (\det M)]^{1/2}} \prod_{i=1}^N d\delta_i, \quad (1.12)$$

in which \mathcal{P}_N is the probability that the density fluctuation field δ has values in the range $\delta(\mathbf{x}_j)$ to $\delta(\mathbf{x}_j) + d\delta(\mathbf{x}_j)$ for each of the $j = 1, \dots, N$ with N an arbitrary integer and $\mathbf{x}_1, \dots, \mathbf{x}_N$

N arbitrary locations in the field (conform Bardeen et al. 1986). The matrix M^{-1} is the inverse of the $N \times N$ covariance matrix M , which is given by

$$M_{ij} = \langle \delta(\mathbf{x}_i)\delta(\mathbf{x}_j) \rangle = \xi(\mathbf{x}_i - \mathbf{x}_j), \quad (1.13)$$

in which the brackets $\langle \dots \rangle$ denote an ensemble average. $\xi(\mathbf{x}_i - \mathbf{x}_j)$ is the autocorrelation function of the density field, which is often simply referred to as the *correlation function* $\xi(\mathbf{r})$. Because the universe is isotropic $\xi(\mathbf{r}) = \xi(r)$. In effect, M is the generalization of the variance σ^2 in a one-dimensional normal distribution. As M is fully determined by the correlation function $\xi(r)$, the same is true for the Fourier transform of correlation function, the *power spectrum* $P(k)$,

$$\langle \delta(\mathbf{x})\delta(\mathbf{x} + \mathbf{r}) \rangle = \int \frac{d\mathbf{k}}{(2\pi)^3} P(k) e^{-i\mathbf{k}\cdot\mathbf{r}}. \quad (1.14)$$

Since it is fully characterized by its second order moment, specifying $P(k)$ determines all the properties of the Gaussian random field.

The shape of the initial power spectrum depends on the precise manner in which the initial fluctuations were generated. The simplest inflationary models predict a *scale-invariant* power spectrum,

$$P(k) \propto k, \quad (1.15)$$

in which the metric fluctuations in the early universe had the same amplitude on all scales. This is also called the *Harrison-Zel'dovich spectrum* and was predicted well before inflation had been suggested (Peebles & Yu 1970, Harrison 1970, Zel'dovich 1972). Other featureless power law models with $P(k) \propto k^n$ are also consistent with inflation.

1.3.2 The linear regime

When the fluctuations are small, it is possible to linearize the equations of motion. The resulting *linear theory* describes the evolution of fluctuations when they are small ($\delta \ll 1$). After some algebraic work a single equation for the density perturbation can be obtained (Peebles 1980),

$$\frac{\partial^2 \delta}{\partial^2 t} + 2\frac{\dot{a}}{a}\frac{\partial \delta}{\partial t} = 4\pi G\rho_b \delta. \quad (1.16)$$

This equation is a linear second-order ordinary differential equation in δ . The general solution consists of two modes which evolve independently. It is given by

$$\delta(\mathbf{x}, t) = A(\mathbf{x})D_1(t) + B(\mathbf{x})D_2(t), \quad (1.17)$$

in which $D_1(t)$ and $D_2(t)$ are the two independent time solutions. They involve one *growing* and one *decaying* solution. As the universe evolves the decaying solution damps away and since the starting fluctuations were tiny anyway it can be ignored. The growing mode solution D_1 depends on the cosmological parameters and is given by

$$D_1(z) = E(z) \int_z^\infty dz' \frac{1+z'}{E^3(z')}, \quad (1.18)$$

in which $E(z)$ is the normalized Hubble function, which is defined by

$$E(z) = [\Omega_m(1+z)^3 + \Omega_k(1+z)^2 + \Omega_\Lambda]^{1/2}. \quad (1.19)$$

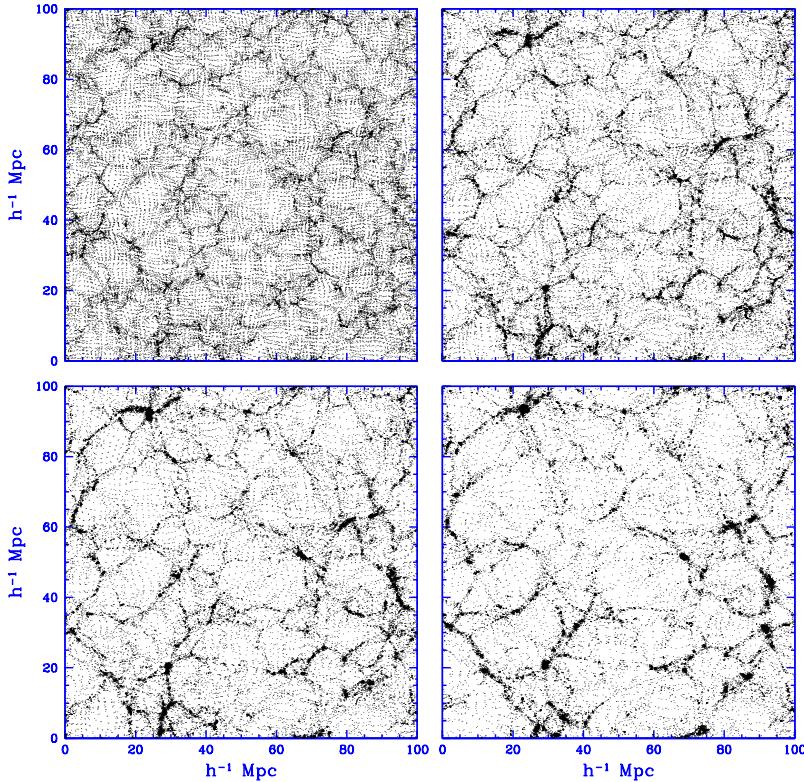


Figure 1.5 — Evolution of structure and development of a cellular morphology in a scenario of structure formation through gravitational instability. Illustrated are four slices, at $a = 0.2, 0.3, 0.5$ and $a = 0.7$, in an SCDM scenario ($\Omega_0 = 1.0, H_0 = 50 \text{ km/s/Mpc}$) from an N -body simulation following 128^3 particles in a $100h^{-1} \text{ Mpc}$ box.

For some special cases an analytic result can be obtained. E.g., in an Einstein-De Sitter universe the growing mode solution is given by $D_1(t) = t^{2/3}$, while in an empty universe it is a constant. The growing mode solution is often referred to as the *linear growth factor*, since density perturbations simply grow proportionally to it: $\delta(\mathbf{x}, t) = D_1(t)\delta(\mathbf{x})$.

1.3.3 The non-linear regime

Once density fluctuations approach unity linear theory ceases to be valid and the governing equations may not be linearized anymore. Due to mode coupling the full non-linear equations are in general too complex to be solved analytically. Beyond the linear regime the evolution of structures has therefore to be followed through computer simulations. An example of such a simulation is shown in Fig 1.5, in which the evolution of large scale structure in a $100h^{-1} \text{ Mpc}$ box is shown. In this figure three important characteristics of the non-linear clustering process are apparent:

- hierarchical clustering;
- anisotropic collapse;
- a complex cellular morphology with extended empty regions.

1.3.3.1 Hierarchical clustering: Press-Schechter theory

Although the full non-linear evolution cannot be understood analytically, we may understand its main characteristics. We have seen that the initial density field is fully determined by its power spectrum $P(k)$, which specifies the amplitude of fluctuations as a function of their spatial scale. In general the initial power spectrum is assumed to be a power law: $P(k) \propto k^n$. The index n determines the balance between small- and large-scale power. In most viable scenarios small-scale fluctuations collapse to form bound objects before larger-scale structures do, resulting in a gradual building-up of successively larger structures by the clumping and merging of smaller-scale structures. This process is called *hierarchical structure formation*. In Fig. 1.6 the hierarchical assembly of a filamentary structure is illustrated.

A powerful model of hierarchical structure formation is *Press-Schechter theory* (Press & Schechter 1974, Bond et al. 1991, Sheth, Mo & Tormen 2001), which describes the sample average characteristics of an emerging population of non-linear objects evolving from a linear field of density fluctuations in the primordial cosmos. Press & Schechter (1974) argued that haloes which have collapsed at a late time can be identified with overdense regions in the initial density field. They described how the assumption that objects form by spherical collapse can be combined with the fact that the initial fluctuation distribution was Gaussian to predict the number density of bound objects that have mass M at time z . For this they assumed that a region collapses at time z if the initial overdensity within it exceeds a critical value, $\delta_{\text{SC}}(z)$. This critical value depends on z , but it is independent of the initial size of the region. The dependence of δ_{SC} on z is given by the spherical collapse model (Gunn & Gott 1972).

Having computed at each cosmic epoch t the fraction $f(M, t)$ of matter fluctuations on a given mass scale M that will have fully collapsed, we may also invert this by computing the cosmic epoch t at which at a particular fixed fraction f_c of the density fluctuations on a given mass scale M has collapsed. Comparing the epoch at which two different mass scales M_1 and M_2 have fully collapsed yields the following expression (see van de Weygaert 2001 for a full derivation):

$$\frac{D(t_{M_1})}{D(t_{M_2})} = \left(\frac{M_2}{M_1} \right)^{(n+3)/2}. \quad (1.20)$$

Here the collapse times have been expressed in terms of the linear growth term. This equation describes the basic differences in the qualitative progression of the hierarchical evolution process for fluctuation spectra with different spectral slopes n . In a pure white noise spectrum in which all scales have the same power ($n = 0$) small-scale clumps fully collapse and virialize before fluctuations of an order of magnitude larger scale start to approach a similar stage. Taking the specific example of an emerging filament, its formation will consist of the gradual assembly of earlier virialized small-scale clumps. Proceeding toward spectra with a more negative slope, in an $n = -2$ scenario a contracting filament will be collapsing while its contents in small-scale clumps have not yet had the opportunity to fully settle. The extreme case of the $n = -3$ scenario represents the asymptotic situation in which fluctuations over

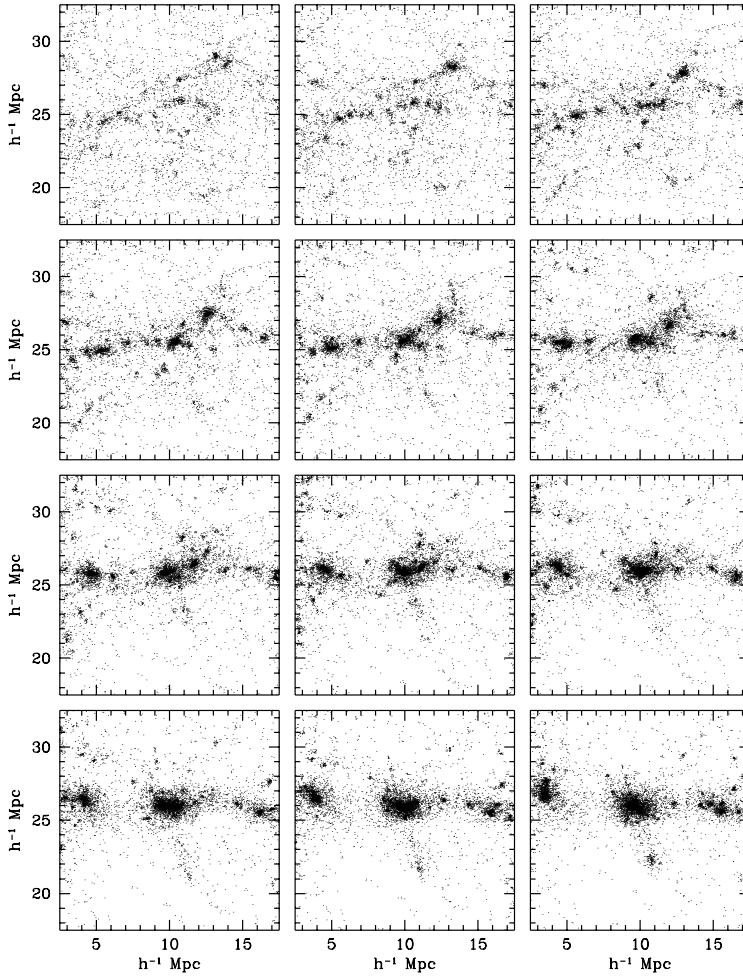


Figure 1.6 — Hierarchical assembly of a filamentary structure. Here we focus in on a small $15 \times 15 h^{-1}$ Mpc region of an N -body simulation of structure formation in an SCDM scenario ($\Omega_0 = 1.0$, $H_0 = 50 \text{ km/s/Mpc}$), from the $100h^{-1}$ Mpc box 128^3 simulation shown in Fig. 1.5. Time is running from top left to bottom right, from $a = 0.1$ to $a = 1.0$. Notice how small-scale clumps aggregate into ever larger haloes, all arranged along a roughly filamentary configuration (oriented horizontally).

the full range of scales undergo contraction and collapse at the same time. Only spectra with $n > -3$ will therefore lead to hierarchical structure formation.

Press-Schechter formalism is based on the assumption that density peaks are perfectly spherically symmetric, but in reality this is not the case. From the statistics of linear Gaussian random fields (e.g. Bardeen et al. 1986) it is known that all peaks have some degree of primordial flattening. This primordial flattening will be amplified by the action of the gravitational force. Sheth, Mo & Tormen (2001) generalized the Press-Schechter formalism to incorporate homogeneous ellipsoidal density perturbations. In this formalism the one-to-one

relation between linear growth factors and perturbation masses (Eqn. 1.20) is not valid anymore. The collapse time now also depends on the primordial shape. Effectively this leads to a diffusion in collapse times along each of the principal directions of a density perturbation. As such, it may happen that clumps with a higher initial overdensity may reach ultimate collapse later than a more moderate overdensity, given a primordial shape which is sufficiently elongated. Also, although an emerging filament will still consist of the gradual assembly of earlier virialized small-scale clumps, these small-scale clumps are not spherically symmetric. Nonetheless, Eqn. 1.20 serves as a valuable guiding principle for the overall appearance of a structure formation scenario.

1.3.3.2 Anisotropic collapse: the Zel'dovich approximation

The second important characteristic of non-linear clustering involves *anisotropic collapse*: structures which at early stages are slightly non-spherical tend to become more and more anisotropic as time evolves. This is clearly visible in Fig. 1.5, in which the emergence of several highly flattened filamentary structures can be observed. This tendency of structure to collapse anisotropically was predicted by Zel'dovich (1970), who explored the non-linear regime by simply assuming that linear conditions remain valid in the early non-linear or *semi-linear* regime. Instead of evaluating the evolution of the density field Zel'dovich used a Lagrangian approach and worked out the initial displacement of a particle from linear theory and assumed that it continued to move in this initial direction. The proper coordinate of a particle is then given by

$$\mathbf{x}(t) = a(t)[\mathbf{q} + D(t)\mathbf{f}(\mathbf{q})], \quad (1.21)$$

in which the coordinates \mathbf{q} are equal to the comoving coordinates at $t = 0$ and $\mathbf{f}(\mathbf{q})$ is the time-independent *displacement field* which is related to the gravitational potential at $t = 0$ (see Peacock 1999). This equation looks like Hubble expansion with some perturbation, which becomes negligible as $t \rightarrow 0$.

The *Zel'dovich approximation* has been very successful in describing a number of properties of the cosmic web. Here we focus on anisotropic collapse. Mass conservation implies that $\rho(\mathbf{x}, t)d\mathbf{x} = \rho(\mathbf{q})d\mathbf{q} = \rho_b d\mathbf{q}$. Hence

$$1 + \delta = \left| \frac{\partial \mathbf{x}}{\partial \mathbf{q}} \right|^{-1} = \frac{1}{(1 - D(t)\lambda_1)(1 - D(t)\lambda_2)(1 - D(t)\lambda_3)}. \quad (1.22)$$

Here the vertical bars denote the Jacobian determinant of the transformation between \mathbf{x} and \mathbf{q} . λ_1, λ_2 and λ_3 are the three eigenvalues of the *strain* or *deformation tensor* $\partial f_i / \partial q_j$. This equation describes the evolution of the density field in the Zel'dovich approximation. It follows that collapse takes place first along the axis defined by the largest positive eigenvalue. For an isolated triaxial ellipsoid perturbation this corresponds to collapse along the shortest axis. Aspherical perturbations thus get enhanced by the action of gravity, which leads to collapse into flattened planar or sheet-like structures known as *pancakes*. The further evolution of the density field will be dictated by the second largest eigenvalue and the respective collapse into filamentary structures. The final collapse along the axis defined by the third eigenvalue will result in the formation of galaxy clusters. At any moment of time the density field will thus contain an array of structures with widely varying anisotropies.

The fact that the Zel'dovich approximation gives a remarkably accurate description of the semi-linear regime is somewhat of a coincidence. The semi-linear phase starts out with the

collapse of structures into planar sheets, making the collapse effectively a one-dimensional contraction for which the Zel'dovich approximation is exact (see Peacock 1999). This argument ceases to hold when sheets of matter cross each other's orbit, at which point the Zel'dovich approximation breaks down.

1.3.3.3 Cellular morphology: tidal fields and voids

The third important characteristic of non-linear clustering is the appearance of a complex cellular geometry, consisting of a foam-like network of filamentary and wall-like structures surrounding extended empty regions. In order to understand this apparently generic outcome, one may follow two approaches.

In a matter-based description one may note that there is a close causal link between the generically anisotropic tidal force fields generated by typical cosmic matter distributions, their impact on the shape of an emerging and evolving structure and the resulting cosmic foam-like structure. Bond, Kofman & Pogosyan (1996) coined the word *cosmic web* in their study, in which they drew attention to their finding that knowledge of the value of the tidal field at a few well-chosen cosmic locations in some region would determine the overall outline of the web-like pattern in that region. Moreover, they showed that the cosmic web is largely defined by the position and primordial tidal fields of rare events in the medium, with the strongest filaments between nearby clusters whose tidal tensors are nearly aligned. The close connection between the local force field and the global matter distribution has since then been confirmed and elucidated by, amongst others, van de Weygaert & Bertschinger (1996).

Conversely, in a void-based description of the evolution of the cosmic matter distribution voids mark the transition scale at which density perturbations have decoupled from the Hubble flow and contracted into recognizable structural features. Voids act as the key organizing element for arranging matter concentrations into an all-pervasive cosmic network. As voids expand matter is squeezed in between them and sheets and filaments form the void boundaries (Icke 1984; van de Weygaert 1991, 2002). This view has been pursued by Regős & Geller (1991), Dubinski et al. (1993), van de Weygaert & van Kampen (1993) and Gottlöber et al. (2003), who have described how voids evolve in numerical simulations of gravitational clustering.

1.4 Current analysis tools and their shortcomings

The hot Big bang and gravitational instability theory provide the physical framework in which we may understand the growth of structure in our universe. We have seen that an analytical description of the emergence of cosmic structure is only possible in the very first linear and semi-linear phases apart from some special, relatively simple physical configurations. This is due to the complexity of the cosmic structure formation process, which is illustrated and underlined by characteristics such as

- hierarchical clustering;
- anisotropic collapse;
- a complex cellular geometry with extended empty regions.

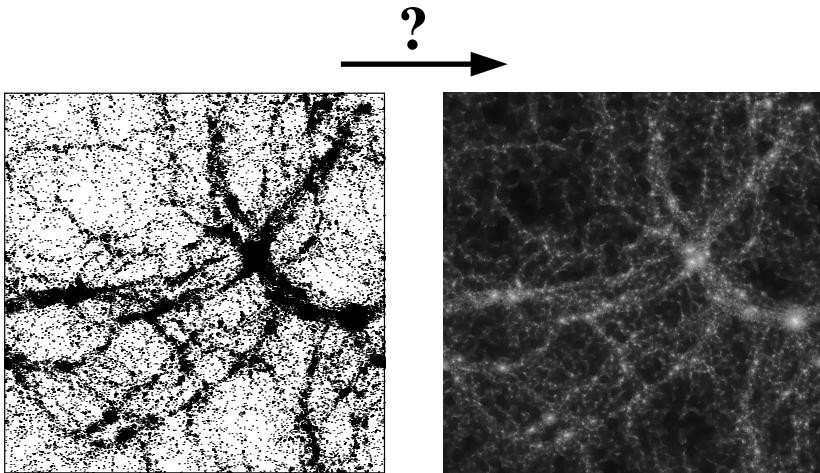


Figure 1.7 — The central problem of this thesis: the reconstruction of a density field (right-hand frame) from a set of points sampling this field (left-hand frame). The example in this figure shows a $10h^{-1}$ Mpc thick slice through a particle distribution resulting from a cosmological N -body simulation. This figure also illustrates another difficulty which one frequently encounters in astronomical applications: the particle distribution is characterized by structures whose shapes and densities are strongly varying within the simulation box. The density field shown in the right-hand frame has been reconstructed with the DTFE, the reconstruction procedure described in this thesis.

Clearly such a high degree of complexity poses a formidable challenge for any quantitative analysis. When studying the formation and evolution of structure in our cosmos we therefore have to resolve to the analysis of galaxy redshift surveys and numerical simulations of cosmic structure formation. Physical insight progresses through the comparison of these experimental observations with predictions of theoretical models. For the case at hand, both observational and numerical experiments involve discretely sampled density (and velocity) fields, while theoretical models involve the corresponding continuous density (and velocity) fields.

Here we are confronted with a central problem in the analysis of galaxy redshift surveys and numerical simulations: the transformation between a discrete set of particles $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ and the corresponding continuous density field $\rho(\mathbf{x})$. This problem is illustrated in Fig. 1.7. The fact that this is not a trivial problem is due to the enormous degree of complexity of the large scale matter distribution. In this thesis we argue that conventional *density* (and *velocity*) *field reconstruction methods* are not able of describing the full complexity of the large scale matter distribution. Instead, most techniques have been specifically devised to describe a single or some of the aspects of the galaxy distribution, but at the same time are completely inappropriate for describing other aspects, often hampering a proper subsequent analysis.

Conventional reconstruction techniques may be of two types. The conceptually most simple ones are non-adaptive, either making use of a fixed grid or a fixed smoothing kernel. In grid-based techniques one calculates the value of the density field at a set of locations defined on a regular grid in accordance with a particular weighting function W , leading to a

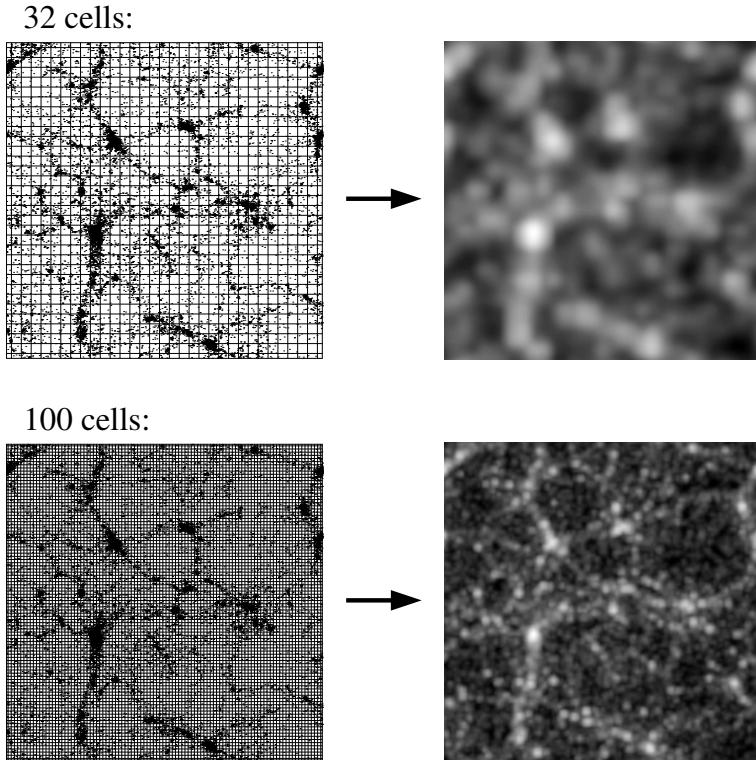


Figure 1.8 — The TSC density field reconstruction procedure. In the TSC procedure the field is overlayed with a fixed regular grid. The number of points inside each grid-cell determines the value of the density field. Clearly visible is how sensitive the reconstructed density field is to the resolution of the grid.

description of the form

$$\widehat{\rho}\left(\frac{\mathbf{n}}{M}\right) = \frac{M^3}{N} \sum_{i=1}^N m_i W\left(\mathbf{x}_i - \frac{\mathbf{n}}{M}\right). \quad (1.23)$$

Here $\mathbf{n} = (n_x, n_y, n_z)$ denotes the grid-cell, M is the number of cells of the grid in each dimension, N is the number of particles and m_i is the mass of particle i . The units have been chosen such that the volume of the box is unity. The weighting function W is chosen such that the resulting density field has the desired resolution and smoothness (e.g. Hockney & Eastwood 1981).

To subsequently obtain the value of the density field at any point in space, one may proceed by applying a smoothing scheme such as given by

$$\widehat{\rho}(\mathbf{x}) = \frac{\sum_{i=1}^N \tilde{\rho}_i W(\mathbf{x} - \mathbf{x}_i)}{\sum_{i=1}^N W(\mathbf{x} - \mathbf{x}_i)}, \quad (1.24)$$

in which $\tilde{\rho}_i$ are the reconstructed values of $\rho(\mathbf{x})$ at the positions \mathbf{x}_i by means of Eqn. 1.23 and $W(\mathbf{x})$ is the adopted filtering function, not necessarily the same as the weighting function used

in Eqn. 1.23 (see e.g. Lombardi and Schneider 2001 for an extensive discussion of this type of smoothing). The smoothing function has a fixed shape and size and is usually spherically symmetric, e.g. a Gaussian.

Grid-based density field reconstructions are illustrated in Fig. 1.8. In this figure some of the disadvantages of non-adaptive procedures are clearly visible. The most important is that the resolution of the grid is arbitrary, while the resulting density field is highly sensitive to its precise value. At low resolutions high density regions are poorly resolved, while anisotropic features are not well recovered. Regions of low density seem to be recovered more accurately. At high resolutions high density regions seem to be better resolved, but regions of low density become hampered by shot-noise effects. Also, extended structures are breaking up into sub-structures. The fixed geometry of the grid also produces other artefacts, an example of which is that the reconstruction of anisotropic features depends on their orientation with respect to the grid.

To overcome the effects of the varying resolution of physical density fields several adaptive procedures have been proposed, for example making use of adaptive grids or filters. A well-known example are SPH-like schemes, in which the mass associated with each particle is smoothed with a spherically symmetric kernel (for a review see Monaghan 1992 and references therein), usually a Gaussian or a spline function, and adding the contributions,

$$\widehat{\rho}(\mathbf{x}) = \sum_{i=1}^N m_i W(\mathbf{x} - \mathbf{x}_i; h_i). \quad (1.25)$$

Here N is the total number of particles, W is the smoothing function and h_i is the smoothing length of particle i . The smoothing length is chosen such that the smoothing volume always contains a certain, user-specified, fixed number of neighbors.

Even though adaptive techniques certainly represent an improvement over conventional grid-based procedures, one may still object that they also involve rather subjective parameters, such as the precise relation between the extent of the filter and the local density, as well as the shape of the filter. Moreover, the rigid geometry of the filter is usually not appropriate for describing highly anisotropic structures.

1.5 Goals and outline of this thesis

In this thesis we have developed a new tool for reconstructing a density or velocity field which has been sampled by an irregularly distributed point set. This tool, the *Delaunay Tessellation Field Estimator* (DTFE), is optimally suited for the analysis of highly complex point distributions such as encountered in the large scale galaxy distributions or in cosmological N -body simulations, in the sense that it automatically adapts to both the density and the geometry of the sampling point distribution without any artificial smoothing.

This thesis consists of two main parts. In the first part, consisting of Chapters 2 to 4, we describe the fundamentals and the properties of the DTFE. We extensively test the DTFE, in particular with respect to point distributions with similar properties as the large scale galaxy distribution. In the second part, consisting of Chapters 5 to 7, several cosmological applications of the DTFE are discussed. In Chapter 8 we discuss how the significance of DTFE reconstructed features may be determined in practical applications and how measurement errors affect a DTFE reconstructed density field.

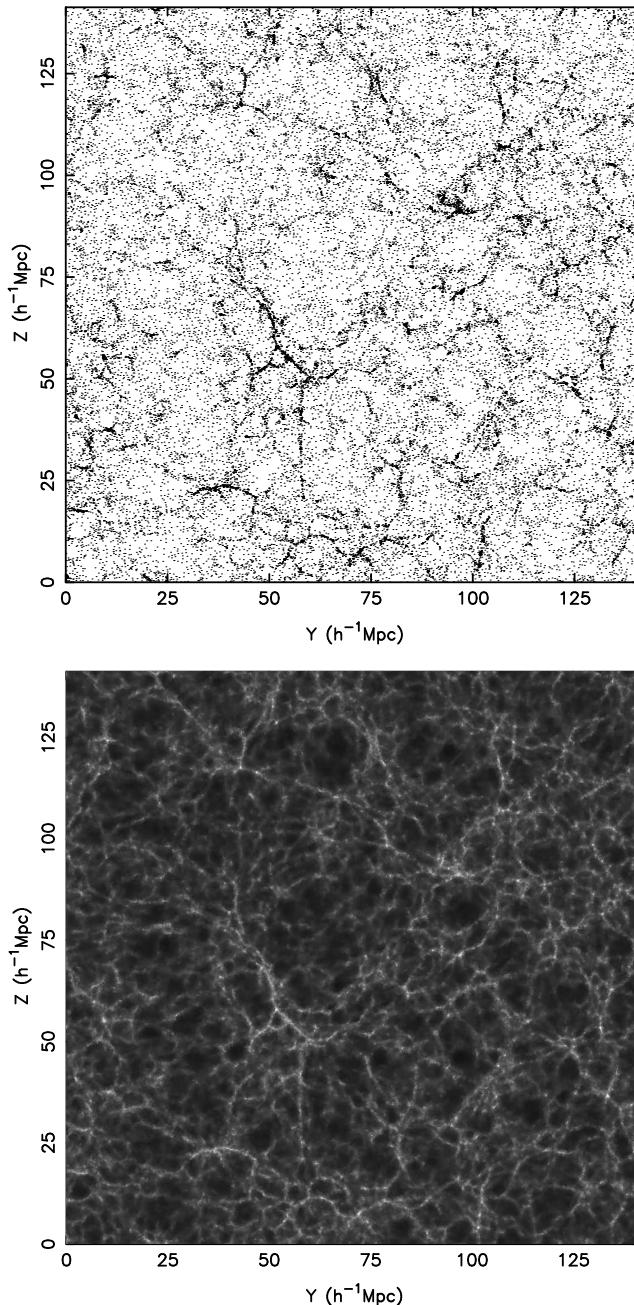


Figure 1.9 — Slice through a GIF N -body simulation (left) and the corresponding DTDE reconstructed density field. Clearly visible is that all the structural details present in the particle distribution have a counterpart in the reconstructed density field.

In Chapter 2 we shortly present the DTFE. Its fundamentals are briefly described and an application to the analysis of a simulation of cosmic structure formation is discussed. In Chapter 3 the fundamentals of the DTFE are extensively discussed. The effective smoothing kernel of the DTFE is compared with that of conventional reconstruction algorithms. In Chapter 4 we explicitly study the performance of the DTFE with respect to complex point distributions whose properties resemble those of the large scale galaxy distribution. In particular, we focus on two crucial aspects for which conventional reconstruction schemes are known to give a suboptimal description: the hierarchy of spatial scales and densities present in the large scale galaxy distribution and its complex cellular geometry, involving structures of widely varying densities and anisotropies.

In Chapter 5 we study the performance of the DTFE density estimation procedure with respect to the SPH density estimation procedure, which forms the heart of many particle-based hydrodynamical simulation algorithms. We discuss how the SPH density estimation recipe in simulation codes can be replaced by the DTFE estimate and to which improvements this will lead. In Chapter 6 we apply the DTFE to a number of high resolution cosmological N -body simulations and show that the DTFE may be used for a simultaneous modelling of both the cosmic density and velocity field. We explicitly model a number of void-like and filamentary regions and show that the dynamics of voids can be used to measure the value of the cosmological constant. In Chapter 7 we apply the DTFE to the 2dF galaxy redshift survey. We present both two-dimensional galaxy surface density and three-dimensional galaxy density maps. In Fig. 1.9 a slice through a Gif N -body simulation is shown with the corresponding DTFE reconstructed density field. All structural details present in the particle distribution have a corresponding counterpart in the slice through the reconstructed density field. The foam-like network of structures which the large scale matter distribution forms is clearly visible and more pronounced than in the particle distribution.

In Chapter 8 we explore the statistical properties of the DTFE, resulting in a discussion of how the significance of reconstructed features may be determined. The effects of several types of errors are also analyzed.

Finally, in Chapter 9 we present a summary of the main findings of this thesis and a brief outlook. We also discuss the main advantages and limitations of the DTFE.

References

- Bardeen J.M., Bond J.R., Kaiser N., Szalay A.S., 1986, ApJ, 304, 15
 Beijersbergen M., 2003, PhD Thesis, University of Groningen, Groningen, The Netherlands
 Bond J.R., Cole S., Efstathiou G., Kaiser N., 1991, ApJ, 379, 440
 Bond J.R., Kofman L., Pogosyan D., 1996, Nature, 380, 603
 Broadhurst T.J., Ellis R.S., Koo D.C., Szalay A.S., 1990, Nature, 343, 726
 Colless M., Dunn A.M., 1996, ApJ, 458, 435
 Colless M. et al., 2001, MNRAS, 328, 1039
 Delone B.N., 1934, Bull. Acad. Sci. USSR: Classe Sci. Mat, 7, 793
 Dubinski J., Da Costa L.N., Goldwirth D.S., Lecar M., Piran T., 1993, ApJ, 410, 458
 Einstein A., 1915, *Die Feldgleichungen der Gravitation*, in ‘Sitzungsberichte der Königlich Preußische Akademie der Wissenschaften’, Berlin
 Freeman K., Bland-Hawthorn J., 2002, ARAA, 40, 487

- Friedmann A., 1922, Zeitschrift für Physik, 21, 326
- Geller M.J., Huchra J.P., 1989, Science, 246, 897
- Gottlöber S., Lokas E.L., Klypin A., Hoffmann Y., 2003, MNRAS, 344, 715
- Gunn J.E., Gott J.R., 1972, ApJ, 176, 1
- Guth A.H., 1981, Phys. Rev. D, 23, 347
- Harrison E.R., 1970, Phys. Rev. D, 1, 2726
- Hockney R., Eastwood J., 1988, *Computer Simulations Using Particles*, Taylor & Francis, Bristol, PA, USA
- Hubble E., 1929, PNAS, 15, 168
- Icke V., 1984, MNRAS, 206, 1
- Kirshner R.P., Oemler A., Schechter P.L., Shectman S.A., 1981, ApJ, 248, L57
- Kirshner R.P., Oemler A., Schechter P.L., Shectman S.A., 1987, ApJ, 314, 493
- de Lapparent V., Geller M.J., Huchra J.P., 1986, ApJ, 302, L1
- Lombardi M., Schneider P., 2001, A&A, 373, 359
- Le Fèvre O. et al., 2004, A&A, 428, 1023
- Lemaître G., 1927, Annals of the Scientific Society of Brussels, 47A, 41
- Mateo M., 1998, ARAA, 36, 435
- Monaghan J.J., 1992, ARAA, 30, 543
- Peacock J.A., 1999, *Cosmological physics*, Cambridge University Press, Cambridge, UK
- Peebles P.J.E., Yu J.T., 1970, ApJ, 162, 815
- Peebles P.J.E., 1980, *The Large-Scale Structure of the Universe*, Princeton University Press, Princeton, USA
- Press W.H., Schechter P., 1974, ApJ, 193, 437
- Regős E., Geller M., 1991, ApJ, 377, 14
- Ryden B., 2003, *Introduction to Cosmology*, Addison Wesley, San Francisco, USA
- Shectman S.A. et al., 1996, ApJ, 470, 172
- Sheth R.K., Mo. H.J., Tormen G., 2001, MNRAS, 323, 1
- Sheth R.K., van de Weygaert R., 2004, MNRAS, 350, 517
- Stoughton C. et al., 2002, AJ, 123, 485
- Szomoru A., 1995, PhD Thesis, University of Groningen, Groningen, The Netherlands
- Szomoru A., van Gorkom J., Gregg M.D., Strauss M.A., 1996, AJ, 111, 2150
- van de Weygaert R., 1991, PhD Thesis, University of Leiden, Leiden, The Netherlands
- van de Weygaert R., van Kampen E., 1993, MNRAS, 263, 481
- van de Weygaert R., 1994, A&A, 283, 361
- van de Weygaert R., Bertschinger E., 1996, MNRAS, 281, 84
- van de Weygaert R., 2002, *Froth across the Universe*, in ‘2nd Hellenic Cosmology Meeting’, Plionis M., Cotsakis S. (eds.), Kluwer Academic Publishers, Dordrecht, NL
- Weinberg S., 1989, Revs. Mod. Physics, 61, 1
- Zel'dovich Y.B., 1970, A&A, 5, 84

2

The Delaunay Tessellation Field Estimator[†]

W.E. Schaap & R. van de Weygaert

ABSTRACT — We introduce the Delaunay Tessellation Field Estimator (DTFE). Its purpose is rendering a volume-covering reconstruction of a density field from a set of discrete data points sampling this field. Reconstructing density or intensity fields from a set of irregularly sampled data is a recurring key issue in operations on astronomical data sets, both in an observational context as well as in the context of numerical simulations. The DTFE is based on the stochastic geometric concept of the Delaunay tessellation generated by the point set. We shortly describe the method and illustrate its virtues by means of an application to an N -body simulation of cosmic structure formation. The DTFE is a fully adaptive method: automatically it probes high density regions at maximum possible resolution, while low density regions are recovered as moderately varying regions devoid of the often irritating shot-noise effects. Of equal importance is its capability to sharply and undilutedly recover anisotropic density features like filaments and walls. The prominence of such features at a range of resolution levels within a hierarchical clustering scenario as the example of the standard CDM scenario is shown to be impressively recovered by the DTFE.

[†]based on Schaap W.E., van de Weygaert R., 2002, A&A, 363, L29

2.1 Introduction

Astronomical observations, physical experiments as well as computer simulations often involve discrete data sets supposed to represent a fair sample of an underlying smooth and continuous field. Conventional methods are usually plagued by one or more artefacts. Firstly, they often involve estimates at a restricted and discrete set of locations – usually defined by a grid – instead of a volume-covering field reconstruction. A problem of a more fundamental nature is that the resulting estimates are implicitly mass-weighted averages, whose comparison with often volume-weighted analytical quantities is far from trivial. For most practical purposes the disadvantage of almost all conventional methods is their insensitivity and inflexibility to the sampling point process. This leads to a far from optimal performance in both high and low density regions, which often is dealt with by rather artificial and ad hoc means.

In particular in situations of highly non-uniform distributions conventional methods tend to conceal various interesting and relevant aspects present in the data. The cosmic matter distribution exhibits conspicuous features like filaments and walls, extended along one or two directions while compact in the other(s). In addition, the density fields display structure of varying contrasts over a large range of scales. Ideally sampled by the data points, appropriate field reconstructions should be set solely and automatically by the point distribution itself. The commonly used methods, involving artificial filtering through for instance grid-size or other smoothing kernels (e.g. Gaussian filters), often fail to achieve an optimal result.

Here we describe a new and fully self-adaptive method based on the Delaunay triangulation of the given point process. After a short description of the fundamentals of our tessellation procedure, we show its convincing performance on the result of an N -body simulation of structure formation, whose particle distribution is supposed to reflect the underlying cosmic density field. A detailed specification of the method, together with an extensive quantitative and statistical evaluation of its performance is presented in Chapters 3, 4 and 8.

2.2 The Delaunay Tessellation Field Estimator

Given a set of field values sampled at a discrete number of locations along one dimension we are familiar with various prescriptions for reconstructing the field over the full spatial domain. The most straightforward way involves the partition of space into bins centered on the sampling points. The field is then assumed to have the – constant – value equal to the one at the sampling point. Evidently, this yields a field with unphysical discontinuities at the boundaries of the bins. A first-order improvement concerns the linear interpolation between the sampling points, leading to a fully continuous field.

In more than one dimension the equivalent spatial intervals of the one-dimensional bins are well-known in stochastic geometry. A point process defines a Voronoi tessellation by dividing space into a unique and volume-covering network of mutually disjunct convex polyhedral cells, each of which comprises that part of multi-dimensional space closer to the defining point than to any of the other (see van de Weygaert 1991 and references therein). These Voronoi cells (see Fig. 2.1) are the multi-dimensional generalization of the one-dimensional bins in which the zeroth-order method approximates the field value to be constant. The natural extension to a multi-dimensional linear interpolation interval then immediately implies the corresponding Delaunay tessellation (Delone 1934). In two dimensions this tessellation consists of a volume-covering tiling of space into triangles (see Fig. 2.1, in three dimensions

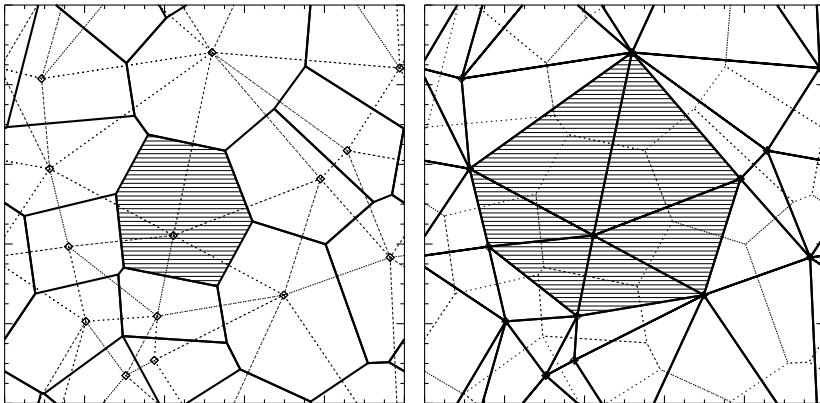


Figure 2.1 — A set of 20 points with their Voronoi (left-hand frame: solid lines) and Delaunay (right-hand frame: solid lines) tessellations. Left-hand frame: the shaded region indicates the Voronoi cell corresponding to the point located just below the center. Right-hand frame: the shaded region is the ‘contiguous Voronoi cell’ of the same point as in the left-hand frame.

these are tetrahedra) whose vertices are formed by three specific points in the dataset. The three points are uniquely selected such that their circumscribing circle does not contain any of the other data points. The Voronoi and Delaunay tessellation are intimately related, being each others dual in that the centre of each Delaunay triangle’s circumcircle is a vertex of the Voronoi cells of each of the three defining points, and conversely each Voronoi cell nucleus a Delaunay vertex (see Fig. 2.1). The favourable properties of the Delaunay tessellation are in fact well-known and have been applied in, amongst others, surface rendering applications such as geographical mapping and various computer imaging algorithms.

Consider a set of N discrete data points in a finite region of D -dimensional space. Having at one’s disposal the field values at each of the $(D+1)$ Delaunay vertices $\mathbf{r}_0, \mathbf{r}_1, \dots, \mathbf{r}_D$ at each location \mathbf{r} in the interior of a D -dimensional Delaunay tetrahedron, the linear interpolation field value is defined by

$$\widehat{f}(\mathbf{r}) = f(\mathbf{r}_0) + \widehat{\nabla f} \Big|_{\text{Del}} \cdot (\mathbf{r} - \mathbf{r}_0), \quad (2.1)$$

in which $\widehat{\nabla f} \Big|_{\text{Del}}$ is the estimated constant field gradient within the tetrahedron. Given the $(D+1)$ field values $f(\mathbf{r}_0), f(\mathbf{r}_1), \dots, f(\mathbf{r}_D)$, the value of the D components of $\widehat{\nabla f} \Big|_{\text{Del}}$ can be straightforwardly computed by evaluating Eqn. 2.1 for each of the D points $\mathbf{r}_1, \dots, \mathbf{r}_D$. This multi-dimensional procedure of linear interpolation has been described by Bernardeau & van de Weygaert (1996) in the context of defining procedures for volume-weighted estimates of cosmic velocity fields. While they explicitly demonstrated that the zeroth-order Voronoi estimator is the asymptotic limit for volume-weighted field reconstructions from discretely sampled field values, they showed the superior performance of the first-order Delaunay estimator in reproducing analytical predictions.

The one factor complicating a trivial and direct implementation of the above procedure in the case of density and intensity field estimates is the fact that the number density of data points itself is the measure of the underlying density field value. Contrary to the case of

velocity fields we therefore cannot start with directly available field estimates at each data point. Instead, we need to define appropriate estimates from the point set itself. Most suggestive would be to base the estimate of the density field at the location \mathbf{r}_i of each point on the inverse of the volume of its Voronoi cell \mathcal{V}_i , $\widehat{\rho}(\mathbf{r}_i) = m/V(\mathcal{V}_i)$. Note that in this we take every data point to represent an equal amount of mass m . The resulting field estimates are then intended as input for the above Delaunay interpolation procedure. However, one can demonstrate that integration over the resulting density field would yield a different mass than the one represented by the set of sampling points (see Chapter 3 for a demonstration). Instead, mass conservation is naturally guaranteed when the density estimate is based on the inverse of the volume of the ‘contiguous’ Voronoi cell \mathcal{W}_i of each data point, $\widehat{\rho}(\mathbf{r}_i) \propto m/V(\mathcal{W}_i)$. The contiguous Voronoi cell of point i is the cell consisting of the union of all $N_{\mathcal{T},i}$ Delaunay tetrahedra $\mathcal{T}_{j,i}$ containing point i as one of its vertices,

$$\mathcal{W}_i = \bigcup_{j=1}^{N_{\mathcal{T},i}} \mathcal{T}_{j,i}. \quad (2.2)$$

Its volume is the sum of the volumes of each of the $N_{\mathcal{T},i}$ Delaunay tetrahedra,

$$V(\mathcal{W}_i) = \sum_{j=1}^{N_{\mathcal{T},i}} V(\mathcal{T}_{j,i}). \quad (2.3)$$

Fig. 2.1 (right-hand frame) depicts an illustration of such a cell. Properly normalizing the mass contained in the reconstructed density field, taking into account the fact that each Delaunay tetrahedron is invoked in the density estimate at $(D + 1)$ locations, we find at each data point the following density estimate,

$$\widehat{\rho}(\mathbf{x}_i) = \frac{(D + 1)m}{V(\mathcal{W}_i)}. \quad (2.4)$$

Having computed these density estimates, we subsequently proceed to determine the complete volume-covering density field reconstruction through the linear interpolation procedure outlined in Eqn. 2.1. Hereafter we will refer to this density field reconstruction procedure as the *Delaunay Tessellation Field Estimator* (DTFE).

2.3 Analysis of a cosmological N -body simulation

Cosmological N -body simulations provide an ideal template for illustrating the virtues of our method. They contain a wide variety of objects, with diverse morphologies, a large reach of densities and spanning over a vast range of scales. They display low density regions, sparsely filled with particles, as well as highly dense and compact clumps, represented by a large number of particles. Moderate density regions typically include strongly anisotropic structures such as filaments and walls.

Each of these features has their own individual characteristics and often these may only be sufficiently highlighted by some specifically designed analysis tool. Conventional methods are usually only tuned for uncovering one or a few aspects of the full array of properties. Contrary to artificial tailor-made methods, which may be insensitive to unsuspected but intrinsically important structural elements, the DTFE is uniquely defined and fully self-adaptive. Its

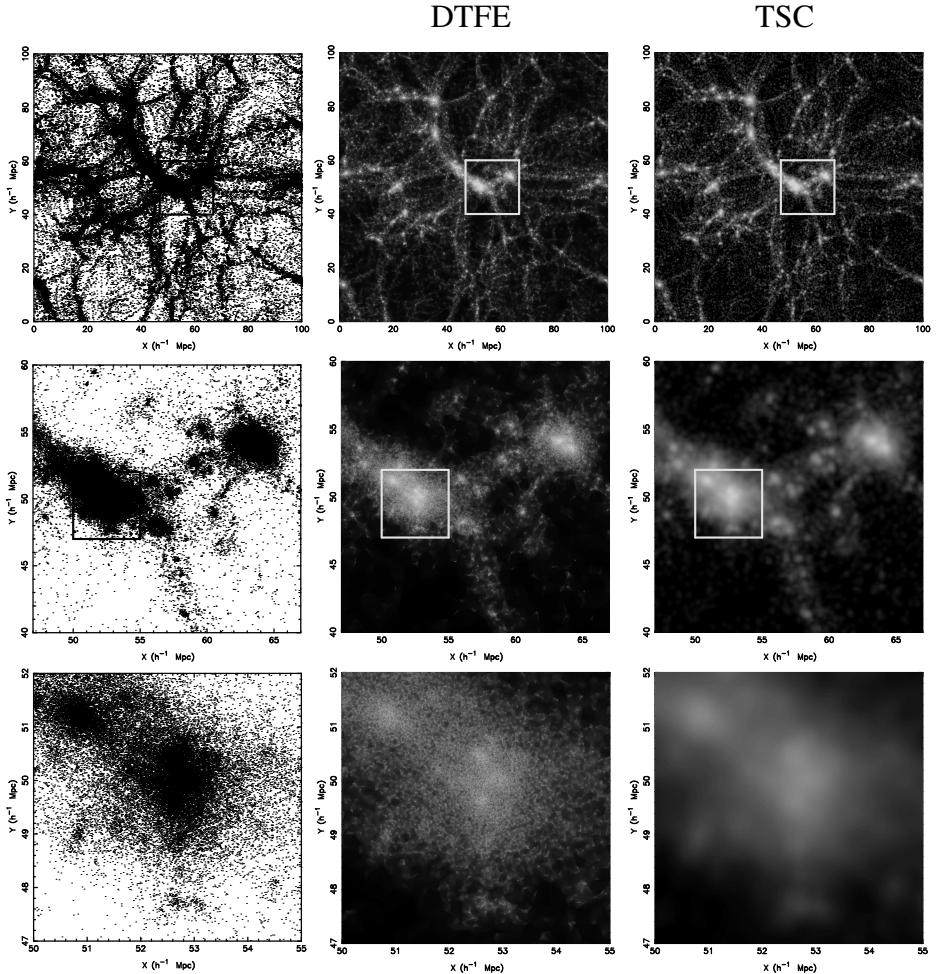


Figure 2.2 — Comparison of the performance of the DTFe with a conventional grid-based TSC method in analyzing a cosmological N -body simulation. Left-hand column: the particle distribution in a $10h^{-1}$ Mpc wide central slice through the simulation box. Central column: the corresponding DTFe density field reconstruction. Right-hand column: the TSC density field reconstruction.

outstanding performance is clearly illustrated by Fig. 2.2. Here we have analyzed an N -body simulation of structure formation in a standard CDM scenario ($\Omega_0 = 1$, $H_0 = 50 \text{ km/s/Mpc}$). It shows the resulting distribution of 128^3 particles in a cubic simulation volume of $100h^{-1} \text{ Mpc}$, at a cosmic epoch at which $\sigma(R_{\text{TH}} = 8h^{-1}\text{Mpc}) = 1$. The figure depicts a $10h^{-1}\text{Mpc}$ thick slice through the center of the box. The left-hand column shows the particle distribution in a sequence of frames at increasingly fine resolution. Specifically we zoomed in on the richest cluster in the region. The right-hand column shows the corresponding density field reconstruction on the basis of the grid-based Triangular-Shaped Clouds (TSC) method, here evaluated on a 518^2 grid. For the TSC method, one of the most frequently applied algorithms, we refer to the description in Hockney & Eastwood (1988). A comparison with other, more elaborate methods which have been developed to deal with the various aspects that we mentioned, of which SPH-based methods are the most common, are presented in Chapters 3, 4 and 5.

A comparison of the left-hand and right-hand columns with the central column, i.e. the DTFE density field reconstruction, reveals the striking improvement rendered by our new procedure. Going down from the top to the bottom in the central column, we observe seemingly comparable levels of resolved detail. The self-adaptive skills of the Delaunay reconstruction evidently succeed in outlining the full hierarchy of structure present in the particle distribution at every spatial scale represented in the simulation. The contrast with the achievements of the fixed grid-based TSC method in the right-hand column is striking, in particular when focus tunes in on the finer structures. The central cluster appears to be a mere featureless blob! In addition, low density regions are rendered as slowly varying regions at moderately low values. This realistic conduct should be set off against the erratic behaviour of the TSC reconstructions, plagued by annoying shot-noise effects.

Fig. 2.2 also bears witness to another virtue of the DTFE. It evidently succeeds in reproducing sharp, edgy and clumpy filamentary and wall-like features. Automatically it resolves the fine details of their anisotropic geometry, seemlessly coupling sharp contrasts along one or two compact directions with the mildly varying density values along the extended direction(s). Moreover, it also manages to deal successfully with the substructures residing within these structures. The well-known poor operation of e.g. the TSC method is clearly borne out by the central right-hand frame. Its fixed and inflexible filtering characteristics blur the finer aspects of anisotropic structures. Such methods are therefore unsuited for an objective and unbiased scrutiny of the foam-like geometry which so pre-eminently figures in both the observed galaxy distribution as well as in the matter distribution in most viable models of structure formation.

Not only qualitatively, but also quantitatively our method turns out to compare favourably with respect to conventional methods. We have carefully scrutinized our method by means of an array of quantitative tests. A full discussion is presented in Chapters 3, 4 and 5. Here we mention the fact that the method recovers the density distribution function over many orders of magnitude. The grid-based methods, on the other hand, only approach the appropriate distribution in an asymptotic fashion and yield reliable estimates of the distribution function over a mere restricted range of density values. Very importantly, on the basis of the continuous density field reconstruction of the DTFE, we obtain an estimate of the density autocorrelation function that closely adheres to the (discrete) two-point correlation function directly determined from the point distribution. Further quantitative assessments are also presented

in Chapter 4. Finally, we should also consider the computational requirements of the various methods. Given a particle distribution, the basic action of computing the corresponding Delaunay tessellation ($O(N \log N)$, van de Weygaert 1991) and the subsequent interpolation steps ($O(N)$) are considerably less CPU intensive than the TSC method ($O(N^2)$). In the case of Fig. 2.1 the DTFE is about a factor of 10 faster. In the present implementation the bottleneck of the DTFE is its memory requirement, which is about a factor of 10 larger than that of the TSC procedure.

The preceding is ample testimony of the promise of tessellation-based methods for the aim of continuous field reconstruction. The presented method, following up on earlier work by Bernardeau & van de Weygaert (1996), may be seen as a first step towards yet more advanced tessellation methods. One suggested improvement will be a second-order method rendering a continuously differentiable field reconstruction, which would dispose of the rather conspicuous triangular patches that form an inherent property of the linear procedure with discontinuous gradients. In particular, we may refer to similar attempts to deal with related problems, along the lines of natural neighbour interpolation (Sibson 1981), such as implemented in the field of geophysics (Sambridge et al. 1995, Braun & Sambridge 1995) and in engineering mechanics (Sukumar 1998). As multi-dimensional discrete data sets are a major source of astrophysical information, we wish to promote tessellation-based methods as a natural instrument for astronomical data analysis.

References

- Bernardeau F., van de Weygaert R., 1996, MNRAS, 279, 693
Braun J., Sambridge M., 1995, Nature, 376, 655
Delone B.N., 1934, Bull. Acad. Sci. USSR: Classe Sci. Mat, 7, 793
Hockney R., Eastwood J., 1988, *Computer Simulations Using Particles*, Taylor & Francis, Bristol, PA, USA
Sambridge M., Braun J., McQueen H., 1995, Geophys. J. Int., 122, 837
Sibson R., 1981, in *Interpreting Multi-Variate Data*, Barnet V. (ed.), Wiley, Chichester, UK
Sukumar N., 1998, PhD Thesis, Northwestern University, Evanston, IL, USA
van de Weygaert R., 1991, PhD Thesis, Leiden University, Leiden, The Netherlands

3

Continuous Fields and Discrete Samples: Reconstruction through Tessellations

W.E. Schaap & R. van de Weygaert

ABSTRACT — We describe the background, the fundamentals and the properties of the Voronoi Tessellation Field Estimator (VTFE) and the Delaunay Tessellation Field Estimator (DTFE), two self-consistent stochastic-geometrical methods for a volume-covering reconstruction of intensity and density fields which are sampled by a discrete set of points. The zeroth-order VTFE is based on the Voronoi tessellation of the given point set, each Voronoi cell representing the multi-dimensional equivalent of a zeroth-order interpolation interval. The first-order DTFE is based on the Delaunay triangulation of the given discrete point set, each Delaunay tetrahedron representing the multi-dimensional equivalent of a linear interpolation interval. The main advantage of reconstruction methods based on tessellations over conventional approaches is that they are intrinsically self-adaptive and do not make use of artificial smoothing. We compare the effective smoothing kernels of the VTFE and DTFE with conventional fixed grid-based and adaptive SPH-like procedures and show that the VTFE and DTFE have a superior spatial resolution. We compare the computational demands with those of conventional reconstruction procedures. Finally, we summarize our results and discuss possible applications and future extensions.

3.1 Introduction

In analyzing astronomical observations or numerical simulations one is often confronted with discrete data-sets which are assumed to represent a fair sample of an underlying smooth and continuous field. Several methods exist for reconstructing this field from the data, but they are subject to a number of problems. Firstly, they are often restricted to a discrete set of grid-points, while one really wants a fully volume-covering reconstruction. Secondly, unlike analytical calculations they do not correspond to volume-weighted quantities. Thirdly, they tend to be a combination of the two different mathematical procedures of reconstruction and smoothing. The combination is such that the separate procedures are not well-defined. Fourthly, conventional methods are usually inflexible with respect to the local sampling density. This results in a less than optimal coverage of high density regions, while low density regions are dominated by shot-noise effects. Finally, these methods are also insensitive to the geometry of the distribution of data points, usually involving a smoothing kernel with a fixed shape.

For the above reasons one has to be very careful in choosing the appropriate method for the type of problem one is dealing with. After all, the properties of the reconstructed field will in general depend on the adopted method and it is not a priori clear how it has been affected by any of the above issues. In an attempt to address this problem, Bernardeau and van de Weygaert (1996) developed schemes based on the Voronoi and Delaunay tessellation of a set of data-points, specifically directed toward interpolating the velocity divergence field. The basis of these schemes is formed by the Voronoi and Delaunay tessellation (Dirichlet 1850, Voronoi 1908, Delone 1934) of a given discrete point set.

The Voronoi tessellation corresponding to a particular point set consists of the covering of space by convex regions Π_i , the Voronoi cells. Each Voronoi cell Π_i is associated with a point n_i such that any point in Π_i is closer to n_i than to any other point. The Voronoi cells may therefore be considered as the natural influence regions of the points in the data-set. An example of a two-dimensional point distribution and its corresponding Voronoi tessellation is shown in Fig. 3.1. The Delaunay tessellation is the straight-line dual of the Voronoi tessellation. Each Delaunay tetrahedron is defined – in three-dimensional space – by a collection of four points in the data-set, corresponding to its vertices. The four points are selected such that their circumscribing sphere does not contain any other data-point. The complete set of Delaunay tetrahedra forms a unique and volume-covering framework of mutually disjoint cells. The minimum circumsphere definition of the Delaunay tetrahedra ensures that the network of Delaunay cells forms the multi-dimensional equivalent of linear interpolation intervals. In Fig. 3.1 an example of a Delaunay tessellation in two dimensions is shown as well.

Indeed, the ‘minimum triangulation’ property of Delaunay tessellations has been known and abundantly applied within the context of several practical situations, ranging from for instance geographical mapping to spatial surface representations in various computer imaging applications. However, in astronomical and astrophysical contexts the concept of Delaunay tessellations has not been widely used, despite the fact that a great many studies are based on observations which by their nature represent a discretely sampled probe of an underlying smooth distribution. In other fields of applied sciences, ranging from for example materials science to geophysics, considerable work has already been done on the implementation of interpolation schemes which are based on the concept of tessellations (see Okabe 2000 for

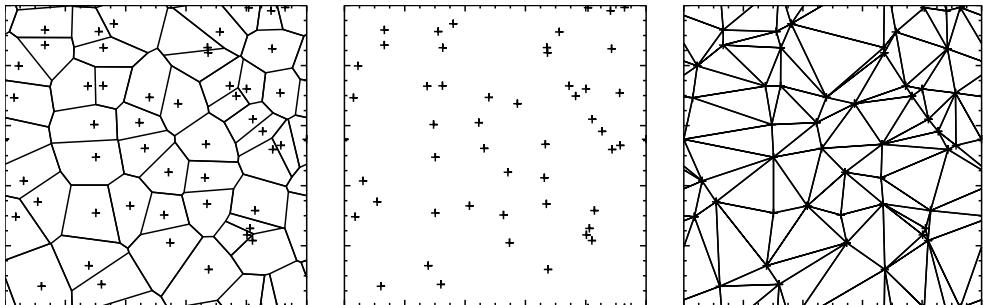


Figure 3.1 — A set of 50 points (central frame) and the corresponding Voronoi (left-hand frame) and Delaunay (right-hand frame) tessellations. Periodic boundary conditions are assumed.

a review). In particular enticing have been the first results in the use of tessellations as self-adaptive computational grids instead of the inflexible solid grid schemes that are used in many numerical codes for following the evolution of natural systems which are described by one or more partial differential equations. An impressive example concerns the numerical modeling of tectonic and surface processes by Braun and Sambridge (1995). This example shows how Delaunay tessellations may be successfully used in problems in which the underlying field varies over orders of magnitude and/or contains caustic features. In the context of astrophysics Whitehurst (1995) implemented a two-dimensional hydrodynamical code based on the Delaunay tessellation of the fluid elements. Because the tessellation automatically adapts to changes in density and geometry, he was able to show that this code is able to accurately follow strong shock fronts and complex gas flows.

A first step towards the exploitation of Delaunay tessellations in the context of astronomical data analysis involved the determination of some statistical aspects of large scale cosmic velocity fields. Observationally, the cosmic velocity field can only be probed through the measurement of the peculiar velocities at the site of galaxies, resulting in a discretely sampled velocity field. Likewise, N -body simulations of structure formation, one of the major instruments of current cosmological research, yield a discrete probe of a supposedly smooth and continuous background velocity field. The schemes developed by Bernardeau & van de Weygaert (1996) on the basis of the Voronoi and Delaunay tessellations were specifically directed towards interpolating the velocity divergence field. The fact that these schemes recover volume-weighted quantities, as the analytic calculations, and that they yielded a significantly better agreement with the results of those calculations than the conventional grid-based procedures formed a convincing argument for the viability of the analytical quasi-linear approximation schemes. While offering transparent and superior estimates of the velocity divergence field, the complicated geometrical computations and memory-demanding nature of the computer implementation has as yet prevented widespread application and further development of tessellation-based interpolation methods.

An extra complication is introduced when attempting to develop similar tessellation estimators of fields which explicitly involve the density or intensity itself. For non-density and intensity fields in principle any reasonable sampling of the field suffices, because the density of the sampling process bears no direct relation with the field. Conversely, in the case of den-

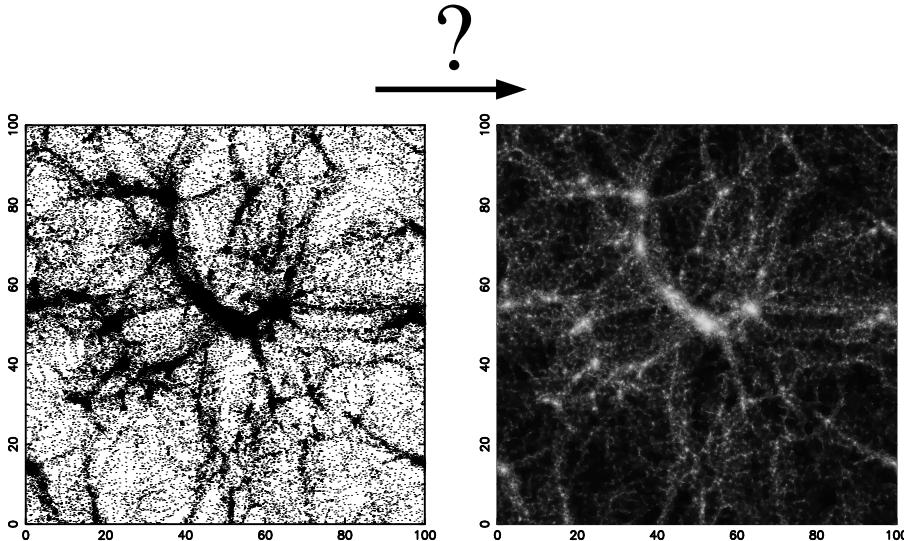


Figure 3.2 — The central problem of this chapter: the reconstruction of a density field (right-hand frame) from a set of points sampling this field (left-hand frame). The example in this figure shows a $10h^{-1}$ Mpc thick slice through a particle distribution resulting from a cosmological N -body simulation. This figure also illustrates a difficulty which frequently occurs in astronomical applications: the particle distribution is characterized by structures whose shapes and densities are strongly varying within the simulation box. The density field shown in the right-hand frame has been reconstructed with the Delaunay Tessellation Field Estimator, the reconstruction procedure described in this chapter.

sity and intensity fields, the density of sampling locations is a direct measure of the quantity to be estimated. This introduces the extra constraint of mass or intensity conservation, namely that the integrated mass or intensity of the reconstructed field should be equal to the total mass or intensity of the sampling points. In Fig. 3.2 an example of this type of problem is depicted.

Several authors have suggested that the inverse area of the Voronoi cells corresponding to a given point set may be used as an estimate of the sampling density at the locations of the points (Brown 1965, Ord 1978, Duyckaerts et al. 1994). In astronomy, this estimate of the local density has been first applied by Ebeling & Wiedenmann (1993) in the context of developing a source detection algorithm. Ramella et al. (2001), Kim et al. (2002) and Marinoni et al. (2002) have applied this technique to the problem of identifying clusters in imaging and redshift surveys. These studies were however limited to a raw estimate of the local sampling density at the location of the points sampling the density field. Here we are interested in the problem of reconstructing a fully volume-covering density field from a set of points sampling this field.

In Chapter 2 we have introduced such a reconstruction scheme which successfully deals with the constraint of mass conservation. In this chapter we work out the definitions of the *Voronoi Tessellation Field Estimator* (VTFE) and the *Delaunay Tessellation Field Estimator* (DTFE), following up on the short introductory description of the DTFE in Chapter 2 and test their performance with respect to a set of more conventional methods.

The DTFE has been developed within the context of the analysis of cosmological N -

body simulations, and incited by our specific interest in studying high density filamentary and wall-like regions (see also Fig. 3.2). Their high density is restricted to flattened or elongated regions, which usually leads to a dilution of the corresponding high density contrast when using the conventional grid-based methods or other roughly direction-independent density estimators. Conventional grid-based methods are also not capable of dealing with the large dynamic range of the cosmic density field, leading to a resolution which is optimal only around the average density. Finally, the occurrence of anisotropic structures in the cosmic density field often coincide with caustic features in the velocity field, which conventional methods are unable to recover. The performance of the DTFE with respect to the presence of hierarchical substructure and anisotropic structures is discussed in Chapter 4, while the DTFE reconstruction of the cosmic velocity field is described in Chapter 6.

This chapter is organized as follows. In section 3.2 we describe the mathematical concept of spatial point processes. Section 3.3 gives an overview of the fundamentals of Voronoi and Delaunay tessellations. In section 3.4 we describe currently available techniques for reconstructing continuous fields from a set of discretely sampled data points. In section 3.5 we discuss multi-dimensional interpolation techniques. In section 3.6 we discuss conventional procedures to reconstruct density and intensity fields. In section 3.7 we describe tessellation-based reconstruction schemes. In section 3.8 we describe the DTFE. Section 3.9 compares the effective smoothing kernel of the DTFE to conventional methods. In section 3.10 we discuss a number of possible processing procedures. In section 3.11 we compare the computation demands of the DTFE with those of conventional reconstruction procedures. Finally, in section 3.12 we summarize and discuss our results.

3.2 Spatial point processes

A spatial point process Φ may be described as a stochastic model determining the location of points $\{\mathbf{x}_i\}$ in some set $X \in \mathbb{R}^D$, with D the dimension of space (Matthes et al. 1978, Ripley 1981, 1988, Diggle 1983, Karr 1986, Daley & Vere-Jones 1988, Cressie 1993, Stoyan et al. 1995, Babu & Feigelson 1996, Martinez & Saar 2001). Any particular set of locations determined by the spatial point process is referred to as a realization of the spatial point process. To be a spatial point process Φ has to satisfy the following two conditions:

1. Φ is *locally finite*: any bounded subset of X contains a finite subset of points for any particular realization;
2. Φ is *simple*: no particular realization contains multiple points.

An important quantity describing a spatial point process is its corresponding intensity field $\lambda(\mathbf{x})$, which describes the local intensity of points. The intensity field is defined by

$$\lambda(\mathbf{x}) = \lim_{V(dY) \rightarrow 0} \frac{\langle \phi(dY) \rangle}{V(dY)}. \quad (3.1)$$

Here ϕ refers to a realization of the point process Φ and $\phi(Y)$ denotes the number of points in this realization in an infinitesimal volume dY with volume $V(dY)$. $\langle \phi(dY) \rangle$ denotes the ensemble average of $\phi(Y)$.

Below we describe a few important examples of point processes.

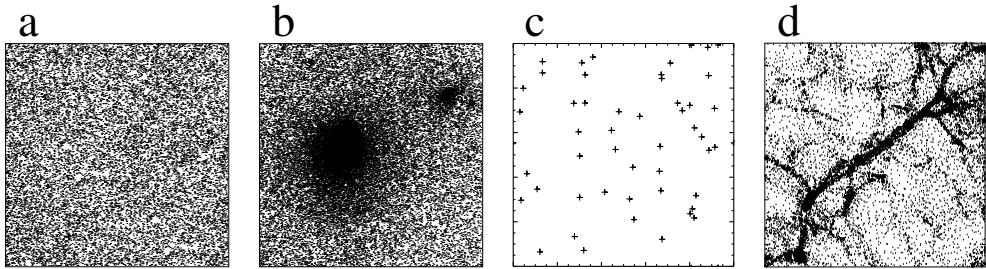


Figure 3.3 — Examples of spatial point processes. (a) Homogeneous Poisson point process. The observed background emission in a small part of the sky. The intensity of the observed radiation is constant across the map. The total number of observed photons during a given time is not constant, but varies according to Poisson statistics. (b) Inhomogeneous Poisson point process. The observed emission from two galaxies in the Coma cluster of galaxies. The intensity of the observed radiation varies across the map. The total number of observed photons during a given time is not constant, but varies according to Poisson statistics. (c) Homogeneous binomial random point process. The position of 50 coins thrown in a box. The probability of a coin ending up at a certain position is the same across the box. The total number of coins in the box is fixed. (d) Inhomogeneous binomial random point process. N -body simulation of filamentary structure in the large scale galaxy distribution. The density of simulation particles varies as a function of position in the simulation box. The total number of simulation particles is fixed.

3.2.1 Homogeneous Poisson point processes

First consider a Poisson point process. For a Poisson point process counts in disjoint set are stochastically independent. The simplest Poisson point process is one whose intensity field does not depend on position, but instead is constant: $\lambda(\mathbf{x}) = \lambda$. Such a process is called a homogeneous Poisson point process. An example is shown in Fig. 3.3, frame (a).

The number of points N in any region of space X with volume $V(X)$ has a Poisson distribution with mean λV . This distribution is given by

$$P[N(X) = k] = \frac{[\lambda V(X)]^k}{k!} e^{-\lambda V(X)}, \quad k = 0, 1, \dots. \quad (3.2)$$

Here $P[N(X) = k]$ is the probability that region X contains k points.

3.2.2 Inhomogeneous Poisson point processes

In practice most interesting applications do not involve a constant intensity field. A Poisson point process with an intensity field which varies as a function of position is called an inhomogeneous Poisson point process. An example is shown in Fig. 3.3, frame (b).

The number of points N in any region of space X with volume $V(X)$ has a Poisson distribution with mean $\Lambda(X)$. This distribution is given by

$$P[N(X) = k] = \frac{[\Lambda(X)]^k}{k!} e^{-\Lambda(X)}, \quad k = 0, 1, \dots. \quad (3.3)$$

Here $\Lambda(X)$ is the integrated intensity in region X defined by

$$\Lambda(X) = \int_X d\mathbf{x} \lambda(\mathbf{x}). \quad (3.4)$$

Many astronomical applications involve an underlying inhomogeneous Poisson point process. Examples include the X-ray intensity field of a galaxy cluster and the galactic density field number in a particular region of the universe as observed in a galaxy redshift survey.

3.2.3 Homogeneous binomial random point processes

Some applications involve point processes in which a fixed number of points is distributed accordingly to some intensity field $\lambda(\mathbf{x})$. Such a process is called a binomial random point process. A binomial random point process is not Poissonian, since counts in disjoint sets are stochastically dependent.

First consider the most simple kind of binomial point process, in which the intensity field does not depend on position, but instead is constant: $\lambda(\mathbf{x}) = \lambda$. Such a process is called a homogeneous binomial random point process. An example is shown in Fig. 3.3, frame (c).

The number of points N in any region of space $X \in W$ has a binomial distribution, which is given by

$$P[N(X) = k] = \binom{N}{k} (p_X)^k (1 - p_X)^{N-k}, \quad k = 0, 1, \dots. \quad (3.5)$$

Here p_X is the probability that a point ends up in region X and is given by

$$p_X = \frac{V(X)}{V(W)}. \quad (3.6)$$

3.2.4 Inhomogeneous binomial random point processes

Again, in practice most interesting applications do not involve a constant intensity field. A binomial random point process with an intensity field which varies as a function of position is called an inhomogeneous binomial random point process. An example is shown in Fig. 3.3, frame (d).

The number of points N in any region of space $X \in W$ is still given by eqn. 3.5, but here p_X is given by

$$p_X = \frac{\int_X d\mathbf{x} \lambda(\mathbf{x})}{\int_W d\mathbf{x} \lambda(\mathbf{x})}. \quad (3.7)$$

Given their fixed number of simulations particles, numerical simulations of for example galaxy and cosmological structure formation are astronomical examples which involve inhomogeneous binomial random fields. Notice that for large numbers of points the distinction between Poisson and binomial point processes becomes academic. In the rest of this chapter we will therefore also refer to binomial random point processes as Poisson point processes.

3.3 Voronoi and Delaunay tessellations

The Voronoi tessellation and the Delaunay tessellation are among the most fundamental and useful geometric constructions defined by a point distribution in a region of D -dimensional space. Voronoi tessellations were first introduced in astronomy by Kiang (1966), who described some properties of Voronoi tessellations of uniform point distributions. The first systematic study in the context of astronomy of the statistical properties of two-dimensional

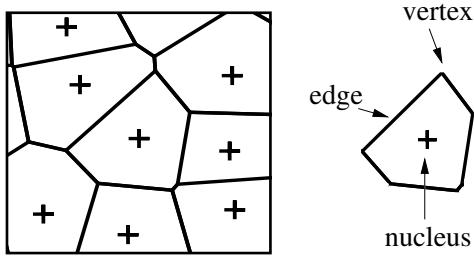


Figure 3.4 — Magnification of the central region of the Voronoi tessellation shown in Fig. 3.1 (left-hand frame) and the different elements of a Voronoi cell (right-hand frame).

Voronoi tessellations was done by Icke & van de Weygaert (1987). Subsequently, they extended their analysis to three dimensions in van de Weygaert & Icke (1989) and van de Weygaert (1991, 1994), following on the completion of a three-dimensional geometrical Voronoi algorithm. In describing the definitions and fundamental properties of Voronoi and Delaunay tessellations, we follow van de Weygaert (1991).

3.3.1 Voronoi tessellations

The *Voronoi tessellation* of a set of nuclei is a subdivision of space in convex regions Π_i , where each region Π_i is associated with a nucleus n_i , such that any point in Π_i is closer to n_i than to any other point. The regions Π_i are called *Voronoi regions* or *Voronoi cells*. Formally, a Voronoi cell Π_i is defined as the set of points \mathbf{x} in space (Dirichlet 1850, Voronoi 1908) for which

$$\Pi_i = \{\mathbf{x} \mid d(\mathbf{x}, \mathbf{x}_i) < d(\mathbf{x}, \mathbf{x}_j) \quad \forall j \neq i\}, \quad (3.8)$$

where $d(\mathbf{x}_i, \mathbf{x}_j)$ is the distance between \mathbf{x}_i and \mathbf{x}_j . An example of a two-dimensional point distribution and its corresponding Voronoi tessellation is given in Fig. 3.1. The central region of this Voronoi tessellation is also shown in Fig. 3.4. In this figure one can see that each Voronoi cell is the intersection of a finite number of many open-half spaces, each being delimited by the perpendicular bisector of the segments joining \mathbf{x}_i with each of the other \mathbf{x}_j 's. Consequently, for all nuclei \mathbf{x}_i the Voronoi cells are convex polyhedra.

Each Voronoi cell is defined by one particular nucleus n_i in the complete set of points. In two dimensions the one-dimensional *Voronoi edges* delimit the boundaries of the cells (see Fig. 3.4). Each Voronoi edge Λ_{ij} is defined by two nuclei n_i and n_j and corresponds to the region of space consisting of points which have an equal distance to n_i and n_j . The pair of nuclei n_i and n_j is called a *contiguous pair* and the members of the pair are said to be *contiguous* to each other. In an analogous fashion, the zero-dimensional *Voronoi vertices* mark the limits of edges (see Fig. 3.4). Each Voronoi vertex V_{ijk} is defined by three nuclei n_i , n_j and n_k and corresponds to the one point equidistant to them and closer to them than to any other nucleus in the sample. In other words, the vertex is the center of the circumsphere (the circumcenter) of the triangle defined by the three nuclei.

In Fig. 3.5 a three-dimensional Voronoi tessellation is shown. It contains an additional element with respect to two-dimensional Voronoi tessellations. The boundaries of the cells are now delimited by two-dimensional *Voronoi walls*. Each Voronoi wall Σ_{ij} is defined by two nuclei n_i and n_j and corresponds to the region of space consisting of points which have an equal distance to n_i and n_j . The one-dimensional Voronoi edges mark the limits of Voronoi walls, while the zero-dimensional Voronoi vertices mark the limits of Voronoi edges.

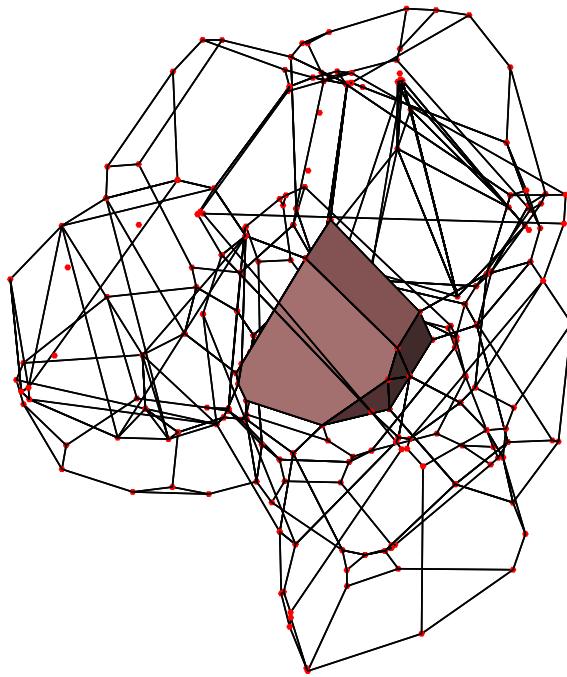


Figure 3.5 — Three-dimensional Voronoi tessellation. The central Voronoi cell is surrounded by its wire-frame depicted Voronoi neighbours. The boundaries of the cells are the polygonal Voronoi walls. The wire edges represent the Voronoi edges. The Voronoi vertices, indicated by dots, are located at each of the two tips of a Voronoi edge, each of them located at the center of the circumsphere of a corresponding set of four nuclei. Courtesy: Jacco Dankers.

3.3.2 Delaunay tessellations

In two dimensions each set of nuclei n_i , n_j and n_k corresponding to a vertex defines a unique triangle, which is known as a *Delaunay triangle*. The three constituent nuclei of each triangle are contiguous to each other. From the set of Voronoi vertices we can thus define an additional dual space-filling tessellation, the *Delaunay tessellation* (Delone 1934), which consists of the set of Delaunay triangles. In three dimensions this tessellation consists of tetrahedra and in a space of any arbitrary dimension D of ‘hyper-triangles’ which are $(D + 1)$ -tuples defined by $(D + 1)$ vertices. An example of a two-dimensional point distribution and its corresponding Delaunay tessellation is shown in Fig. 3.1 (right-hand frame), while in Fig. 3.6 the relation between the Voronoi and the Delaunay tessellation is illustrated.

Delaunay triangles have some important properties. The circumcenter of a Delaunay triangle, i.e. the center of the circumsphere, is a vertex of the corresponding Voronoi tessellation (see Fig. 3.6). This may be immediately appreciated from the definition of the Voronoi tessellation: the nuclei forming the Delaunay triangles are equidistant from the corresponding Voronoi vertex. By definition Delaunay triangles satisfy the ‘empty circumcircle criterion’, meaning that inside the interior of the circumcircle of each Delaunay triangle there are no

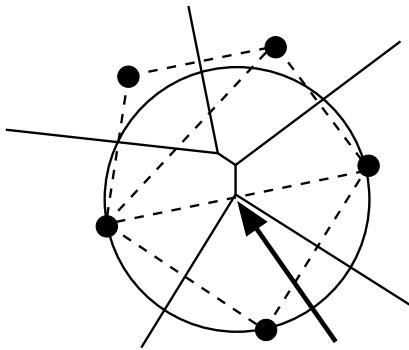


Figure 3.6 — Relation between the Voronoi and Delaunay tessellation of a point set. The straight solid lines indicate the Voronoi tessellation of the five sampling points, while the dashed lines indicate their Delaunay tessellation. The circle is the circumcircle of the lowest Delaunay triangle and illustrates that its center coincides with the vertex of a Voronoi cell (indicated by the arrow). The circle also illustrates that inside the interior of a Delaunay triangle no other sampling points are present.

other points from the defining point set (see Fig. 3.6). This follows from the fact that if an additional nucleus exists inside the circumcircle, it is nearer to the circumcenter than the other nuclei at the surface of the circumcircle. This would imply that the center cannot be the common vertex of the Voronoi polyhedra of the nuclei on the surface of the circumcircle. Finally, among all possible triangulations of a given point distribution, the Delaunay triangles are objects of minimal size and elongation (Lawson 1977, Okabe et al. 2000). All these properties are valid for any number of dimensions.

In general, the Delaunay hyper-triangles consist of $(D + 1)$ vertices, corresponding to the $(D + 1)$ Voronoi polyhedra which meet at each vertex. In principle, however, it is possible that a larger number of Voronoi polyhedra meet at a vertex. Such vertices are called *degenerate*. An example of this occurs for nuclei which are distributed over a regular lattice. A two-dimensional example of this is shown in figure 3.7, in which a circum-circle is shown which contains four nuclei. If a Voronoi tessellation contains degenerate vertices, the corresponding Delaunay tessellation is (locally) not unique. Note however that the measure of such degenerate point sets in a generic stochastic point distribution is zero.

To calculate the Voronoi and Delaunay tessellation corresponding to a given set of data-points, we have used a geometrical algorithm which has been developed by van de Weygaert & Icke (1989) and van de Weygaert (1991, 1994). This algorithm has the advantage that one may use periodic boundary constraints, making it for example ideally suited for analyzing cosmological simulations.

3.4 Reconstruction of continuous fields

A frequently occurring problem in the analysis of astronomical data is the reconstruction of a presumably smooth, continuous field $f(\mathbf{x})$ from a set of discretely sampled data-points. Several methods exist to deal with this problem. The conventional and most commonly used approach is to smooth the data with some filtering function into a continuous map. According to this procedure the estimated value of the field $f(\mathbf{x})$ at the position \mathbf{x} is equal to a weighted sum over all N measured data points:

$$\widehat{f}(\mathbf{x}) = \frac{\sum_{i=1}^N \tilde{f}_i W(\mathbf{x} - \mathbf{x}_i)}{\sum_{i=1}^N W(\mathbf{x} - \mathbf{x}_i)}, \quad (3.9)$$

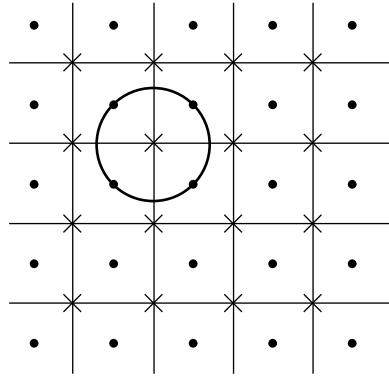


Figure 3.7 — Degenerate Voronoi vertices. Shown is a set of 25 points and the corresponding Voronoi tessellation (the latter is denoted by the solid lines). The crosses indicate the positions of the vertices of the Voronoi cells. All these are degenerate: at each vertex four Voronoi cells meet. This may be seen by the circle whose boundary contains four points.

in which \tilde{f}_i are the measured values of $f(\mathbf{x})$ at the positions \mathbf{x}_i and $W(\mathbf{x})$ is the adopted filtering function (see e.g. Lombardi and Schneider 2001 for an extensive discussion of this type of smoothing). The smoothing function has a fixed shape and size and is usually spherically symmetric. An example of such a spherically symmetric smoothing functions is a Gaussian.

This method has the advantage that it is conceptually simple and the fact that it is so frequently used may give the impression that it produces reasonable results. However, in practice situations may occur in which the above scheme will not work satisfactorily. In particular, variations of the underlying field on a smaller scale than that of the smoothing function will not be recovered. Implicitly, when the flattening of anisotropic features along one or two of their dimensions has a smaller scale than the filter function, their geometry will get smoothed out in accordance with the shape and extent of this filter function.

3.4.1 Mass-weighted versus volume-weighted reconstructions

It is instructive to consider this problem from a more mathematical point of view, first discussed by Bernardeau and van de Weygaert (1996) for the specific case of the cosmic velocity field. If we apply the smoothing scheme (Eqn. 3.9) to a set of field samples \tilde{f}_i , the reconstructed field $\hat{f}(\mathbf{x})$ can be written as:

$$\begin{aligned}\hat{f}(\mathbf{x}) &= \frac{\sum_{i=1}^N \tilde{f}_i W(\mathbf{x}-\mathbf{x}_i)}{\sum_{i=1}^N W(\mathbf{x}-\mathbf{x}_i)} = \\ &= \frac{\int d\mathbf{y} f(\mathbf{y}) W(\mathbf{x}-\mathbf{y}) \sum_{i=1}^N \delta_D(\mathbf{y}-\mathbf{x}_i)}{\int d\mathbf{y} W(\mathbf{x}-\mathbf{y}) \sum_{i=1}^N \delta_D(\mathbf{y}-\mathbf{x}_i)} = \\ &= \frac{\int d\mathbf{y} f(\mathbf{y}) W(\mathbf{x}-\mathbf{y}) \rho(\mathbf{y})}{\int d\mathbf{y} W(\mathbf{x}-\mathbf{y}) \rho(\mathbf{y})}.\end{aligned}\quad (3.10)$$

Here $f(\mathbf{x})$ denotes the true, underlying field, while $\rho(\mathbf{x})$ is the volume density of the sampling points. From this relation we can immediately conclude that the reconstructed field corresponds to the underlying field, smoothed by a filtering function $W(\mathbf{x})$ and weighted by the density of the sampling points. In other words, these conventional estimators produce ‘mass-weighted’ estimates of $f(\mathbf{x})$. This certainly complicates the comparison with analytical

calculations, for the simple reason that an integral over stochastic functions correspond to volume-weighted quantities,

$$f_{\text{volume}}(\mathbf{x}) = \frac{\int d\mathbf{y} f(\mathbf{y}) W(\mathbf{x}-\mathbf{y})}{\int d\mathbf{y} W(\mathbf{x}-\mathbf{y})}. \quad (3.11)$$

This difference reflects the fact that we only have knowledge of the underlying field at the location of the sampling points. From Eqn. 3.11 it is straightforward to see that only in the asymptotic limit of a uniform and infinitely fine sampling, corresponding to a sampling point density of unity everywhere, the reconstructed field is equal to the volume-weighted velocity field. This implies that it is not possible to make any firm statements on the underlying field values – or on related quantities – on the basis of a finite and irregularly distributed set of field samples.

3.4.2 Reconstruction versus filtering

One can make another fundamental objection to the reconstruction procedure described by Eqn. 3.9, namely that it is a combination of the following two closely related, yet distinct mathematical procedures:

1. Reconstruction;
2. Filtering.

However, the combination is such that the separate procedures are not well-defined. This leads to the above described problems related to confusing mass-weighted results with volume-weighted results. In fact, the reconstructed field is given by Eqn. 3.10, while the smoothed ‘true’ underlying field is given by Eqn. 3.11. We therefore advocate a better defined procedure, in which reconstruction and filtering are well-defined and distinct elements of the method. This yields the following expression for the filtered reconstructed field:

$$\widehat{f}_{\text{filtered}}(\mathbf{x}) = \frac{\int d\mathbf{y} \widehat{f}(\mathbf{y}) W(\mathbf{x}-\mathbf{y})}{\int d\mathbf{y} W(\mathbf{x}-\mathbf{y})}, \quad (3.12)$$

while the reconstructed field $\widehat{f}(\mathbf{x})$ is determined by the adopted reconstruction procedure. This approach isolates the effects of the uncertainty of the field values in between the sampled locations to the reconstruction procedure.

3.5 Multi-dimensional interpolation: tessellation-based field reconstruction schemes

A mathematical procedure for reconstructing volume-covering fields without making use of smoothing is interpolation. Effectively, we may define the interpolated field value at any arbitrary point \mathbf{x} in space as the linear combination of the N known field values at the location of the sampling points:

$$f(\mathbf{x}) = \sum_{i=1}^N \alpha_i(\mathbf{x}) f(\mathbf{x}_i), \quad (3.13)$$

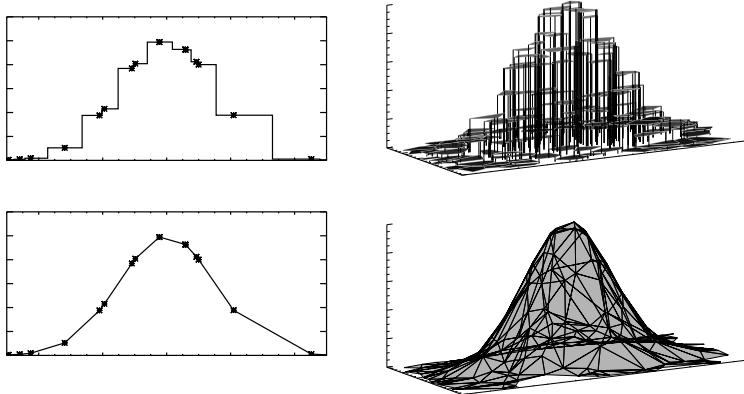


Figure 3.8 — Overview of zeroth-order (top row) and first-order (bottom row) interpolation in one (left-hand column) and two (right-hand column) dimensions.

in which $\alpha_i(\mathbf{x})$ is the weight of sampling point i at position \mathbf{x} . The weights are normalized such that

$$\sum_{i=1}^N \alpha_i(\mathbf{x}) = 1 \quad \forall \mathbf{x}. \quad (3.14)$$

This formalism is equivalent to the general smoothing formalism described by Eqn. 3.9. However, the essential difference between interpolation and smoothing is that in the case of interpolation one is not free to chose an arbitrary weighting function. In particular, the interpolated field value at the location of a sampling point should be equal to the field sample at that location. This requirement puts the following constraints on the weighting functions:

$$\alpha_i(\mathbf{x}_i) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases} \quad (3.15)$$

Moreover, loosely speaking the weighting functions are zero outside the region of influence of the nearest sampling points, such that the sum in Eqn. 3.13 includes only these nearest points. A vast literature exists on various interpolation schemes, ranging from simple zeroth-order schemes to complex higher-order schemes, of which spline interpolation is a particularly noteworthy example. A few specific interpolation options are the following ones.

3.5.1 Zeroth-order interpolation

In one dimension the simplest way of estimating the value of a function of one variable which is sampled at a number of distinct locations, is by dividing the space into bins centered on the sampling points and to assume that the function is constant inside these bins, the constant being equal to the function value at the sampling point (a graphical illustration is shown in the top left-hand frame of Fig. 3.8). Basically these bins form the one-dimensional Voronoi tessellation of the set of sampling points. The multi-dimensional generalization of this is constructing the Voronoi tessellation of the sampling points and to assume that inside each

Voronoi cell the function is constant (see the top right-hand frame of Fig. 3.8). Mathematically, let a function f be measured at a set of locations \mathbf{x}_i and let \mathcal{V}_i denote the Voronoi cell corresponding to point i . The weighting functions $\alpha_i(\mathbf{x})$ are then given by

$$\alpha_i(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x} \in \mathcal{V}_i \\ 0 & \text{if } \mathbf{x} \notin \mathcal{V}_i. \end{cases} \quad (3.16)$$

This approach has the disadvantage that the interpolated field is discontinuous on the boundaries of the Voronoi cell, which results in an unphysical field.

3.5.2 First-order interpolation

In one dimension the most straightforward way to obtain a continuous field is by linearly interpolating between the sampling points (a graphical illustration is shown in the bottom left-hand frame of Fig. 3.8). In this way the resulting function is continuous everywhere, but its derivative or gradient is discontinuous at the boundaries of the bins. One may generalize this approach to more dimensions by covering the space with a space-filling network of hyper-triangles, the sampling points being the vertices of these hyper-triangles. Inside the hyper-triangles we define the function as the linear interpolation between the values at the vertices (see the bottom right-hand frame of Fig. 3.8).

The interpolation procedure can be mathematically formulated as follows. Consider a point \mathbf{x} inside hyper-triangle j which consists of vertices $\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_D$, in which D is the dimension of the space. Then we define the value of a function $\hat{f}(\mathbf{x})$ at location \mathbf{x} by

$$\hat{f}(\mathbf{x}) = f(\mathbf{x}_0) + \left. \nabla \hat{f} \right|_j \cdot (\mathbf{x} - \mathbf{x}_0). \quad (3.17)$$

The D components of the gradient $\nabla \hat{f}$ of the function \hat{f} inside hyper-triangle j may be calculated by solving the following set of D equations:

$$\begin{aligned} f(\mathbf{x}_1) &= f(\mathbf{x}_0) + \left. \nabla \hat{f} \right|_j \cdot (\mathbf{x}_1 - \mathbf{x}_0) \\ &\vdots \\ f(\mathbf{x}_D) &= f(\mathbf{x}_0) + \left. \nabla \hat{f} \right|_j \cdot (\mathbf{x}_D - \mathbf{x}_0). \end{aligned} \quad (3.18)$$

Thus, the $(D+1)$ points which belong to each hyper-triangle are exactly adequate to fix the D values of the field gradient ∇f . Note that given a triangulation of space, it is straightforward to determine the interpolation for M different fields $\{f^1, f^2, \dots, f^M\}$.

This procedure may also be formulated in terms of the weighting functions $\alpha_i(\mathbf{x})$. The resulting equations, however, are rather cumbersome. Consider hyper-triangle m consisting of the $(D+1)$ vertices $\mathbf{x}_{m0}, \mathbf{x}_{m1}, \dots, \mathbf{x}_{mD}$. The only weights which are non-zero in this triangle are $\alpha_{m0}, \alpha_{m1}, \dots, \alpha_{mD}$. Each of these weights may be represented by a plane in $(D+1)$ -dimensional (\mathbf{x}, α) -space. Weight α_i corresponds to a plane which crosses $(\mathbf{x}_i, 1)$ and $(\mathbf{x}_j, 0)$ for $j \neq i$. Let the vertices of hyper-triangle m be ordered such that $m0 = i$. Then inside m the plane corresponding to weight α_i is given by

$$(\mathbf{x}, \alpha) = (\mathbf{x}_i, 1) + t_1(\mathbf{x}_{m1} - \mathbf{x}_i, 1) + t_2(\mathbf{x}_{m2} - \mathbf{x}_i, 1) + \dots + t_D(\mathbf{x}_{mD} - \mathbf{x}_i, 1), \quad (3.19)$$

in which $0 \leq t_1, t_2, \dots, t_D \leq 1$. From this equation it follows that

$$\alpha_i(\mathbf{x}) = 1 + t_1 + t_2 + \dots + t_D. \quad (3.20)$$

To solve this equation for $\alpha_i(\mathbf{x})$ at a particular location \mathbf{x} one has to solve Eqn. 3.19 for t_1, t_2, \dots, t_D . This formalism is equivalent to the formalism in terms of the gradient ∇f described by Eqns. 3.17 and 3.18. From Eqn. 3.17 it follows that ∇f is related to the gradient of the weights through

$$\widehat{\nabla f}(\mathbf{x}) = \sum_{i=1}^N \nabla \alpha_i(\mathbf{x}) f(\mathbf{x}_i). \quad (3.21)$$

Evidently the choice of interpolation hyper-triangles determines the quality and accuracy of the multi-dimensional linear interpolation. Bernardeau and van de Weygaert (1996) argue that Delaunay tetrahedra form the ideal choice of hyper-triangles for this purpose. The reasons for this are twofold. Firstly, the Delaunay tetrahedra are objects of compactness and minimal size and elongation (see the discussion in section 3.3). Secondly, since the Delaunay tetrahedra are the straight-line dual of the Voronoi polyhedra, they are the natural choice for multi-dimensional linear interpolation intervals. This is because linear interpolation requires a definition of neighbor intervals. The natural definition of neighbor points in the multi-dimensional situation is that the two points share a Voronoi wall, meaning that they are contiguous to each other. This suggest that Delaunay tetrahedra are a natural choice for multi-dimensional linear interpolation intervals.

3.5.3 Spline interpolation

In principle, one may generalize the above schemes to interpolation procedures of higher order. Such schemes yield reconstructed fields which are differentiable up to a higher order. Here we discuss two particularly noteworthy examples, spline interpolation and natural neighbor interpolation.

Spline interpolation (Schoenberg 1946a, b) is based on interpolating between sampling points by means of higher-order polynomials. The coefficients of the polynomial are determined ‘slightly’ non-locally, such that a global smoothness in the interpolated function is guaranteed up to some order of derivative. The order of the interpolating polynomials is arbitrary, but in practice cubic splines are most widely used. Cubic splines produce an interpolated function that is continuous through the second derivative. To obtain a cubic spline interpolation for a dataset of $N + 1$ points N separate cubics are needed. Each of these cubics should have each end point match up exactly with the end points to either side. At the location of these points the two adjacent cubics should also have equal first and second derivatives. A full mathematical derivation can be found in e.g. Gerald & Wheatley (1999) or Press et al. (1992).

Spline interpolation is a widely used procedure. Equalising the derivatives has the effect of making the resulting interpolation appear smooth and visually pleasing. For this reason splines are for example frequently used in graphical applications. Splines can provide extremely accurate results when the original sample rate is notably greater than the frequency of fluctuation in the data. Splines however cannot deal very well with large gaps in the dataset. Because the gap between two points is represented by a cubic, these result in peaks or troughs in the interpolation. Another objection to splines is that they are artificially defined constructs.

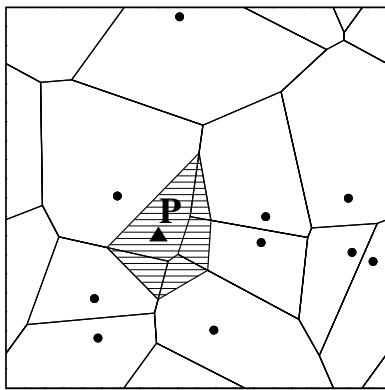


Figure 3.9 — Second-order Voronoi cell. Shown is a box containing 10 sampling points (solid dots) and the corresponding Voronoi tessellation. The shaded region shows the second-order Voronoi cell corresponding to point P . Periodic boundary conditions were imposed.

3.5.4 Natural neighbor interpolation

Natural neighbor interpolation (Sibson 1981) is based on the concept of the *second-order Voronoi cell*. Given a set of field samples $\{\mathbf{x}_i\}$ and its corresponding Voronoi tessellation $\{\mathcal{V}_i\}$ one may construct a new Voronoi cell $\mathcal{V}_{\mathbf{x}}$ around an arbitrary point \mathbf{x} in space, which consists of the region of space closer to \mathbf{x} than to any of the sampling points $\{\mathbf{x}_i\}$ (see Fig. 3.9). $\mathcal{V}_{\mathbf{x}}$ is called the second-order Voronoi cell corresponding to \mathbf{x} in order to stress the difference with the original Voronoi cells $\{\mathcal{V}_i\}$. The interpolated field value at any location \mathbf{x} is defined as a weighted average over all field samples (Eqn. 3.13). The weights $\alpha_i(\mathbf{x})$ are determined by the relative volume each Voronoi cell coincides with the second-order Voronoi cell corresponding to location \mathbf{x} :

$$\alpha_i(\mathbf{x}) = \frac{V(\mathcal{V}_i \cup \mathcal{V}_{\mathbf{x}})}{V(\mathcal{V}_{\mathbf{x}})}. \quad (3.22)$$

Here V denotes the volume of a region of space.

The fact that the second-order Voronoi cell only coincides with Voronoi cells of nearby sampling points ensures the locality of this interpolation procedure. One may show that natural neighbor interpolation yields continuously differentiable interpolated fields, except at the location of the sampling points themselves (Sibson 1981). Natural neighbor interpolation has been successfully implemented in several applications in diverse fields such as computational geometry, geophysics and solid state physics (see e.g. Sambridge et al. 1995, Braun & Sambridge 1995, Sukumar 1998, Boissonat & Cazals 2002).

3.6 Reconstruction of density fields

The basic issue which we wish to address is that of recovering the underlying continuous intensity or density field from a discrete point set. For these fields one does not have a set of field samples, which is usually the case for other discretely sampled continuous fields such as temperature or velocity fields. In the case of density and intensity fields these field samples have to be determined from the sampling point distribution itself. A further complication is the requirement that the total mass (energy) contained in the density (intensity) field should be equal to the mass (energy) of the sampling points.

An example is the calculation of the intensity of light falling on a CCD-camera. In this

case one does not have a set of measured intensities, but this has to be derived from the distribution of the received photons. Another example occurs in N -body simulations, in which one is interested in the density field. This field is represented by a discrete set of particles and the density field has to be derived from the distribution of the simulation particles.

The intensity $\lambda(\mathbf{x})$ of a spatial point process is defined by Eqn. 3.1:

$$\lambda(\mathbf{x}) = \lim_{V(dY) \rightarrow 0} \frac{\langle \phi(dY) \rangle}{V(dY)}. \quad (3.23)$$

Unfortunately we cannot simply use this equation to obtain the intensity field corresponding to the point process, since in general we are dealing with one particular realization ϕ . A direct application of Eqn. 3.1 to this single realization would yield a reconstructed field which can be written as a sum of delta functions:

$$\widehat{\lambda}(\mathbf{x}) = \sum_{i=1}^N \delta(\mathbf{x} - \mathbf{x}_i), \quad (3.24)$$

in which the sum is over all N sampling points in ϕ . A stochastic sampling of λ by $M > 1$ points would have a non-zero measure of two or more points ending up at the same location, in direct contradiction with the requirement of simpleness.

For this reason one has to use another approach. Since the ensemble average cannot be determined one usually relies on taking the average over a particular finite local volume instead. In this way the problem of only having a single realization is circumvented but at the cost of a lower resolution. The main difference between reconstruction procedures is their prescription of the local volume over which the average is taken, which effectively determines the balance between the resolution of the reconstructed field and its susceptibility to sampling noise.

In the rest of this chapter we explicitly refer to the case of a density field represented by a set of particles, but evidently the same arguments hold for an intensity field represented by a set of photons. For the purpose of reconstructing a density field a number of methods are used in astronomical applications, which can be divided into two main groups. The first group is that of the grid-based schemes, in which field estimates are confined to the locations of a fixed and regular grid, which is in principle defined independently from the point distribution itself. The second group is that of what we refer to as ‘SPH-like’ schemes, in which the locations of the interpolation are confined to or defined by the point distribution itself.

3.6.1 Grid-based reconstruction schemes

First consider the grid-based methods, in which one calculates the value of the density field at a set of locations defined on a regular grid. The mass of each particle is spread over this grid in accordance with a particular weighting function W , leading to a description of the form

$$\widehat{\rho}\left(\frac{\mathbf{n}}{M}\right) = \frac{M^3}{N} \sum_{i=1}^N m_i W\left(\mathbf{x}_i - \frac{\mathbf{n}}{M}\right). \quad (3.25)$$

Here $\mathbf{n} = (n_x, n_y, n_z)$ denotes the grid-cell, M is the number of cells of the grid in each dimension, N is the number of particles and m_i is the mass of particle i . The units have been chosen

such that the volume of the box is unity. The weighting function W is chosen such that the resulting density field has the desired resolution and smoothness.

In these methods both the grid and the weighting function are usually fixed and closely related in scale. The grid-size is in principle arbitrary but often set such that on average each grid-cell contains one particle. The shape and size of the weighting function are dependent on the adopted method.

Well-known examples of grid-based reconstruction algorithms are the nearest grid-point (NGP) scheme, in which the mass of a particle is fully assigned to its nearest grid-point, the cloud in cell (CIC) scheme, in which the mass of a particle is linearly interpolated to its closest surrounding grid-points, and the triangular-shaped clouds (TSC) scheme, a quadratic scheme in which the mass of a particle is distributed over all grid-points within a distance of at most two grid-cells. In these schemes the weighting functions are given by a product of one-dimensional weighting functions,

$$W(\mathbf{x}) = \prod_{i=1}^D w(x_i), \quad (3.26)$$

in which D is the dimension of the space and the one-dimensional weighting functions are defined as follows (see also Hockney and Eastwood 1981).

Nearest Grid-Point (NGP):

$$w(x_i) = 1, \quad M |\delta x_i| \leq \frac{1}{2}. \quad (3.27)$$

Cloud in Cell (CIC):

$$w(x_i) = 1 - M |\delta x_i|, \quad M |\delta x_i| \leq 1. \quad (3.28)$$

Triangular-Shaped Clouds (TSC):

$$w(x_i) = \begin{cases} \frac{3}{4} - M^2 \delta x_i^2, & M |\delta x_i| \leq \frac{1}{2} \\ \frac{1}{2} \left(\frac{3}{2} - M |\delta x_i| \right)^2, & \frac{1}{2} \leq M |\delta x_i| \leq \frac{3}{2}. \end{cases} \quad (3.29)$$

From these specifications it follows that the NGP reconstructed density field attains the highest spatial resolution. It is also the coarsest with the low density regions being most severely affected by shot noise. The TSC prescription produces the smoothest fields and is least plagued by shot noise effects, at the cost of a lower spatial resolution. Important to note is that in the zeroth order NGP scheme the density is discontinuous when a particle moves across a cell boundary. In the first order CIC scheme, the density is continuous, but its first derivative is discontinuous. In the second-order TSC scheme, both the density and its first derivative are continuous.

By using these grid-based methods one obtains the values of the density field at the set of grid-points. To subsequently obtain the value of the density field at any point in space,

one can proceed by applying the smoothing scheme given by Eqn. 3.9 for any arbitrary, well-behaved smoothing function W . Examples are the weighting functions in Eqns. 3.27-3.29, corresponding to different kinds of sophistications. The estimation of density values through grid-based methods is subject to similar problems as the smoothing scheme (Eqn. 3.9) is. Grid-based methods are not sensitive to variations in the density field on scales smaller than the grid-size. They are also not successful in recovering anisotropic and caustic features present in the particle distribution.

Grid-based methods have found a widespread application in large scale structure research. In particular, they do play an essential role in particle mesh (PM) N -body codes, a major workhorse for numerical cosmological simulations (see Bertschinger 1998 for a review and references therein). In these codes the mesh is used for the evaluation of the density field, which is needed to solve the Poisson equation. Several authors have tried to improve on the flexibility and adaptability with respect to the particle density by using adaptive grids (see Bertschinger 1998 and references therein). The basic idea is to overlay the original grid with a finer grid at places where the particle density is higher than a certain specified value, such that the resulting resolution is adapted to the local density. Although this procedure certainly improves on the resolution issue and may work fine for very localized problems, it does not address the problems connected to anisotropies. Also, a large number of variables have to be set in a subjective way, such as the mesh size and shape, as well as the condition at which a finer grid is overlaid.

3.6.2 SPH-like reconstruction schemes

An alternative approach, in which the density field is directly reconstructed at every point in space, are SPH-like schemes. These find their origin in particle-based hydrodynamical simulation codes (see Monaghan 1992 for a review and references therein). Such simulations may contain density gradients over orders of magnitude and may involve compression and strong shock waves. In the context of cosmological simulations, the evolutions of systems through gravitational collapse may naturally lead to such situations. A simple grid-based approach will not suffice, as it will be totally unable to resolve the relevant physical processes there where they will be most likely to occur, namely in the regions of highest density. In SPH-like schemes, the gas is represented by individual gas elements. In a Lagrangian fashion, the simulations follow the path and development of each fluid element. This then leads to a formalism wherein the gas is represented by a discrete set of particles, each of which is characterized by its location, velocity and a set of hydrodynamical quantities.

In SPH-like schemes the density field has to be recovered from the particle distribution representing the gas elements. At any point in space this is done by smoothing the mass associated with each particle with a spherically symmetric kernel (although some authors have developed alternative kernels, such as the ellipsoidal kernels introduced by Shapiro et al. 1996), usually a Gaussian or a spline function, and adding the contributions:

$$\widehat{\rho}(\mathbf{x}) = \sum_{i=1}^N m_i W(\mathbf{x} - \mathbf{x}_i; h). \quad (3.30)$$

Here N is the total number of particles, W is the smoothing function and h is the smoothing length which varies as a function of the position \mathbf{x} . The smoothing length is chosen such that

the smoothing volume always contains a certain, user-specified, fixed number of neighbors. A commonly used choice for the kernel function is the spherically symmetric spline kernel of Monaghan & Lattanzio (1985), which is given by

$$W(r, h) = \frac{10}{7\pi h^2} \begin{cases} 1 - \frac{3}{2}\left(\frac{r}{h}\right)^2 + \frac{3}{4}\left(\frac{r}{h}\right)^3, & 0 \leq \frac{r}{h} \leq 1 \\ \frac{1}{4}\left[2 - \left(\frac{r}{h}\right)\right]^3, & 1 < \frac{r}{h} \leq 2 \\ 0, & \frac{r}{h} > 2. \end{cases} \quad (3.31)$$

Two approaches may be distinguished for the determination of the local smoothing length:

1. the scatter approach, in which the smoothing length is set at the location of each particle;
2. the gather approach, in which this length is set at the location at which one wants to determine the value of the density field.

In this way the resolution of the resulting density field is automatically adapted to the local density of particles.

Despite its spatial adaptivity the number of neighbors which one requires to be contained in the smoothing volume still has to be specified by the user and therefore requires a level of arbitrariness. The same holds for the shape of the kernel function. An important objection for most SPH implementations is that the shape of the kernel is fixed and spherical, which results in anisotropic and caustic features being smeared out. In an attempt to improve on this, Shapiro et al. (1996) and Owen et al. (1998) developed a scheme using ellipsoidal kernels, in which the axis ratios of the ellipsoid is determined by the local geometrical distribution of the simulation particles. Here one may object that the resulting method is rather contrived and the specific functional shape of the smoothing kernel and the smoothing length have to be set subjectively.

Another important objection is that the conventional SPH reconstruction procedure, which is a combination of the gather- and scatter-approach, is not self-consistent in the sense that it does not conserve mass: the mass corresponding to the reconstructed field is different from the total mass of the points sampling this field. The reason for this is that in the gather-prescription the contribution to the density from a particular sampling point at a particular location depends on the positions of the other sampling points (see also Hernquist & Katz 1989).

SPH-like schemes for reconstructing the density field from N -body simulations have been introduced by Colberg (1998) and Jenkins et al. (1998). They employ a Gaussian filter to redistribute the mass of the particles in a cell onto a mesh. The smoothing radius is determined by the number of particles in the cell and its surroundings. If this number is bigger than a user-specified value, (in three dimensions typically in the order of 20-30), then no smoothing is applied. If it is smaller, then the surrounding cells are also taken into account. This procedure is repeated until the smoothing volume contains at least the user-specified number of particles.

In summary, we have seen that although a number of tools to construct continuous density and intensity fields from irregularly sampled data do exist, they are all subject to a number of systematic problems. These problems may have a large impact on the analysis of the data, in particular if the data varies over orders of magnitude or if it contains anisotropic or caustic

features. Cosmological N -body simulations and cosmological redshift surveys involve both these issues. Although a number of extensions and modifications for the described methods seek to solve some of these issues, in general they seem rather contrived. Moreover, they do tend to involve a large number of parameters which have to be specified by the user, and hence represent a certain level of arbitrariness.

In addition, the conventional tools to reconstruct density and intensity fields are also subject to the fundamental objection that they involve a fuzzy combination of the two different mathematical procedures of smoothing and reconstruction in a single step.

3.7 Tessellation-based reconstruction of density fields: the Voronoi Tessellation Field Estimator

To improve on this situation we have developed a class of methods to reconstruct density or intensity fields which do not make use of artificial filtering and which do not involve parameters which have to be specified beforehand. These methods are based on the Voronoi and Delaunay tessellations of the sampling point distribution. This approach is motivated by the successful implementations of locally adaptive interpolation schemes based on tessellations.

The density reconstruction procedure is divided into three main steps:

1. the construction of the Voronoi or Delaunay tessellation;
2. the estimation of the density at the locations of the sampling points;
3. the interpolation of the estimated density values to the complete volume of space using the multi-dimensional interpolation schemes described in section 3.5.

It is important to note that in these methods the reconstruction of the field and any processing or analyzing procedures, such as image and profile construction, filtering or feature detection are distinct and well-defined procedures. Although for most applications crucial as here the ‘science’ of the particular application comes in, these post-processing procedures are not an ingredient of the reconstruction procedure. They may, if desired, be applied to the reconstructed density field after the reconstruction procedure has been completed. Processing procedures are discussed in more detail in section 3.10

Another important note is that once the tessellation has been constructed, the interpolation procedure (step 3) may at the same time be used to reconstruct other dynamical fields which have been sampled at the location of the sampling points. For example, in a cosmological N -body simulation the cosmic velocity field is sampled at the location of the simulation particles. This velocity field may therefore be reconstructed simultaneously with the density field. An important advantage of this approach is that the reconstructed velocity field will have the same resolution as the density field.

The crucial element of a density reconstruction procedure is the choice of the density estimator. Not only does its spatial extent determine the resolution of the reconstructed field, it should be chosen in such a way that the total mass corresponding to the density field equals the total mass of the sampling points.

First consider the simplest case, a zeroth-order interpolation scheme (see section 3.5.1). In this scheme, the *Voronoi Tessellation Field Estimator* (VTFE), the Voronoi tessellation corresponding to the distribution of sampling points is constructed. The reconstructed field is assumed to be constant inside each of the Voronoi cells, the constant being equal to the field value at the location of the sampling point defining each individual Voronoi cell. This field

value is estimated as the inverse of the volume of the Voronoi cell. The VTNE results in a density field which is discontinuous at the boundaries of each Voronoi cell.

This way of estimating the density at the location of sampling points has first been suggested by Brown (1965) and Ord (1978). In astronomy, the Voronoi density estimator has been first applied by Ebeling & Wiedenmann (1993) in the context of developing a source detection algorithm. Ramella et al. (2001), Kim et al. (2002) and Marinoni et al. (2002) have applied this technique to the problem of identifying clusters in imaging and redshift surveys. These studies were however limited to a raw estimate of the local sampling density at the location of the points sampling the density field. Here we include the Voronoi density estimator in a self-consistent scheme, the VTNE, to reconstruct a fully volume-covering density field from a set of points sampling this field.

3.7.1 The VTNE procedure

The three main steps of the VTNE are listed below. These steps are graphically illustrated in Fig. 3.10.

1. Construct the Voronoi tessellation corresponding to the point distribution;
2. Estimate the value of the density field $\widehat{\rho}_i$ at the locations $\{\mathbf{x}_i\}$ of the sampling points using the Voronoi density estimator:

$$\widehat{\rho}_i = \frac{m_i}{V(\mathcal{V}_i)}. \quad (3.32)$$

Here m_i denotes the mass of sampling point i and $V(\mathcal{V}_i)$ the volume of Voronoi cell \mathcal{V}_i . Result: set of estimated density field values $\{\widehat{\rho}_i\}$;

3. Interpolate these values to any other point in space by assuming the density field is constant inside each Voronoi cell. Result: reconstructed density field $\widehat{\rho}(\mathbf{x})$.

3.7.2 Mass conservation

Given an interpolation scheme the density estimator should guarantee that the total mass corresponding to the density field equals the total mass of the sampling points. The total mass M corresponding to a density field $\rho(\mathbf{x})$ is given by the integral of the density field over space,

$$M = \int d\mathbf{x} \rho(\mathbf{x}). \quad (3.33)$$

According to this scheme the density is constant in each Voronoi cell, so that this integral can be written as a sum over all Voronoi cells,

$$M = \sum_{i=1}^N V(\mathcal{V}_i) \widehat{\rho}_i. \quad (3.34)$$

Here N is the total number of sampling points and $\widehat{\rho}_i$ is the density inside Voronoi cell \mathcal{V}_i which is equal to the density estimate at the location of sampling point i . From Eqn. 3.34 it may be appreciated that the total mass M is equal to the total mass of the sampling points

VTFE reconstruction procedure:

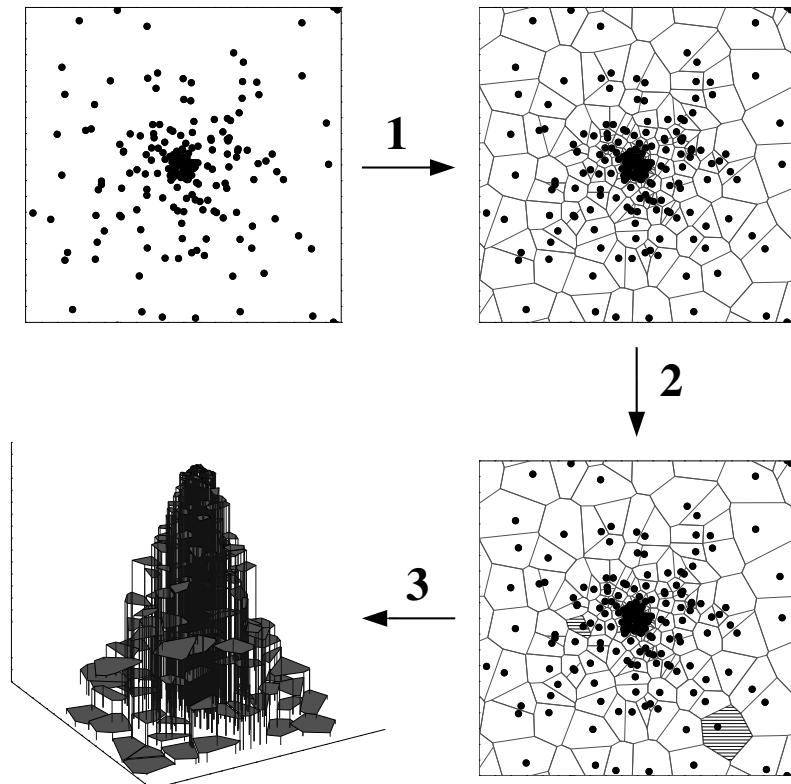


Figure 3.10 — Overview of the zeroth-order VTFE reconstruction procedure. Given a point distribution (top left), one has to construct its corresponding Voronoi tessellation (top right), estimate the density at the position of the sampling points by taking the inverse of the area of their corresponding Voronoi cells (bottom right) and finally to assume that the density is constant within each Voronoi cell, resulting in a volume-covering discontinuous density field (bottom left).

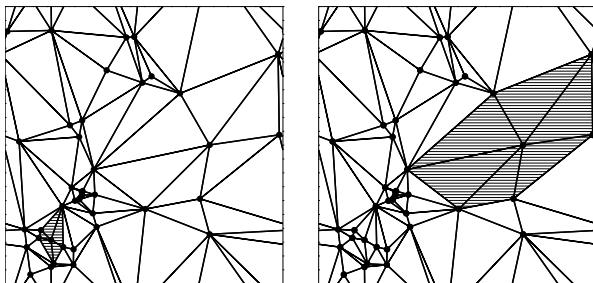


Figure 3.11 — Contiguous Voronoi cells. Left-hand frame: the contiguous Voronoi cell corresponding to a point in a high density environment is denoted by the shaded area. Right-hand frame: the same for a point in a low density environment. In the figure it can be seen that the local point sampling density is inversely proportional to the area of the contiguous Voronoi cell.

$\sum_{i=1}^N m_i$ if the density at the location of the sampling points is defined as the inverse of the volume of their corresponding Voronoi cells weighted by their mass:

$$\widehat{\rho}(\mathbf{x}_i) = \frac{m_i}{V(\mathcal{V}_i)}. \quad (3.35)$$

From Eqns. 3.34 and 3.35 it may be appreciated that the mass of each sampling point is fully distributed over its corresponding Voronoi cell, showing that the VTFE is a strictly local reconstruction scheme.

3.8 The Delaunay Tessellation Field Estimator

The VTFE results in a reconstructed field which is constant inside the Voronoi cells corresponding to the sampling points and discontinuous at the boundaries of these cells, which evidently yields an unphysical reconstructed field. A continuous density field may be obtained by the linear interpolation scheme (see section 3.5.2). In this scheme, the *Delaunay Tessellation Field Estimator* (DTFE), the Delaunay tessellation corresponding to the distribution of sampling points is constructed. The reconstructed field is assumed to vary linearly inside each of the Delaunay tetrahedra. The DTFE results in a density field is volume-covering and continuous, but whose derivative is discontinuous at the boundaries of the Delaunay triangles.

In the DTFE procedure the density field has to be estimated at the location of the sampling points. Below we show that the zeroth-order Voronoi density estimator $\widehat{\rho}(\mathbf{x}_i) = m_i/V(\mathcal{V}_i)$ is not a valid choice, as it would not ensure that the total mass corresponding to the reconstructed field is equal to the total mass of the sampling points. The density field at the location of a sampling point should instead be defined as the inverse of the volume of its surrounding Delaunay tetrahedra, the union of which we will refer to as the *contiguous Voronoi cell* \mathcal{W} of the sampling point. Fig. 3.11 illustrates this concept.

3.8.1 The DTFE procedure

The three main steps of the DTFE are listed below. These steps are graphically illustrated in Figs. 3.12 to 3.14 and summarized in Fig. 3.15.

1. Construct the Delaunay tessellation corresponding to the point distribution;

Step 1: from set of points to Delaunay tessellation

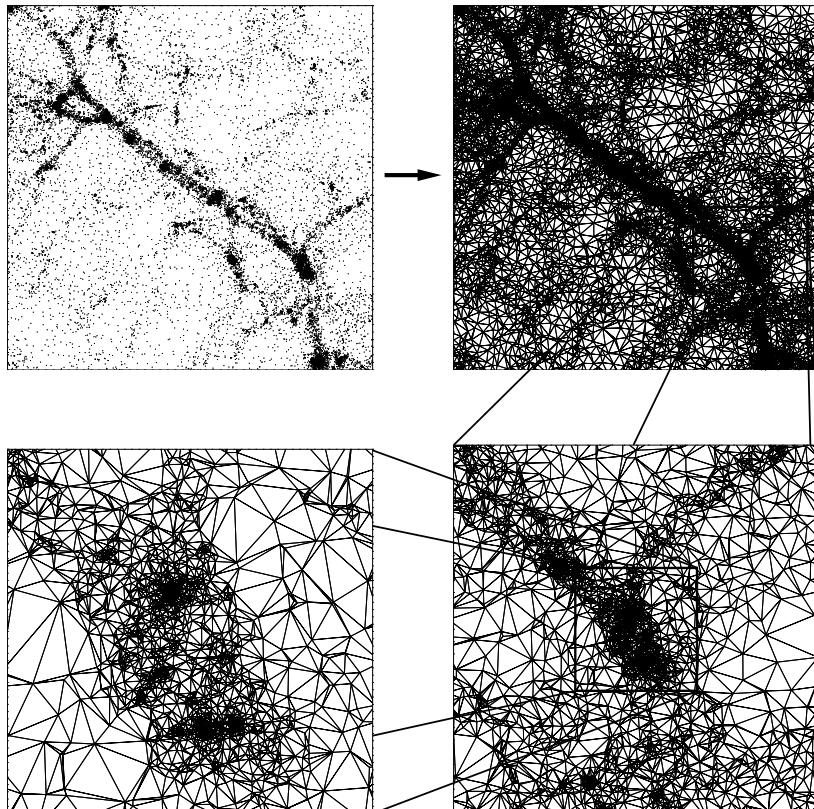


Figure 3.12 — Step 1 of the DTFE reconstruction procedure. Given a point distribution (top left), one constructs the corresponding Delaunay tessellation (top right). The bottom frames illustrate the adaptive properties of the Delaunay tessellation. These frames zoom in on the regions indicated by the squares.

Step 2: from tessellation to set of density estimates

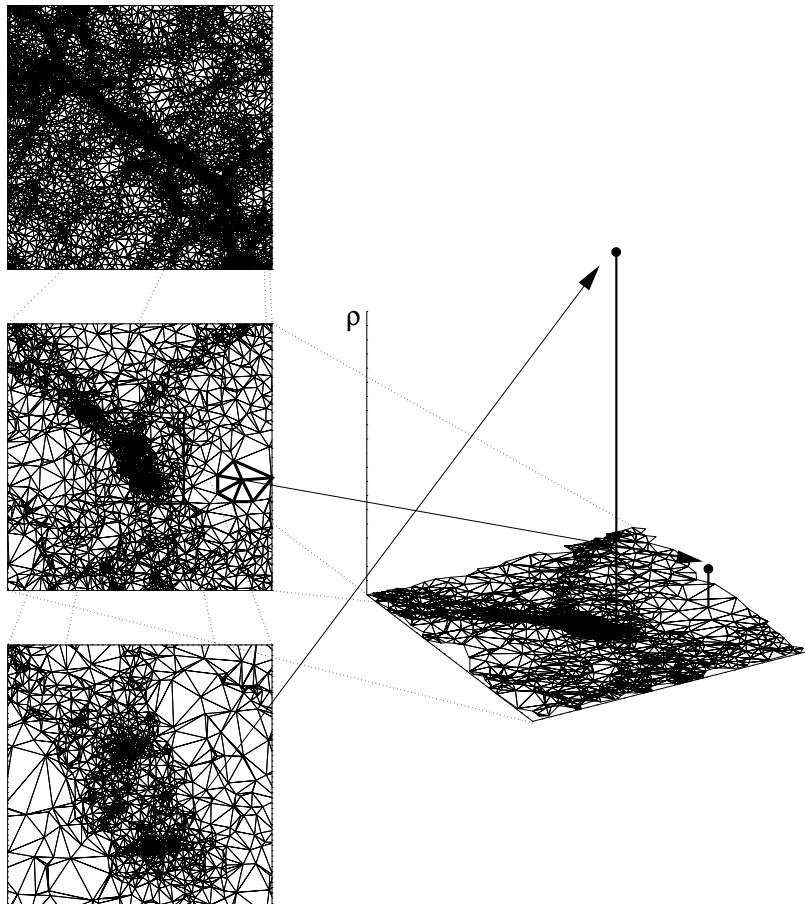


Figure 3.13 — Step 2 of the DTTE reconstruction procedure. The value of the density field is estimated at the location of each sampling point by determining the inverse of the area of their contiguous Voronoi cells. In the central and bottom left-hand frames this is illustrated for both a low and high density environment. The resulting density estimates are shown in the right-hand frame.

Step 3: from set of density estimates to volume-covering density field

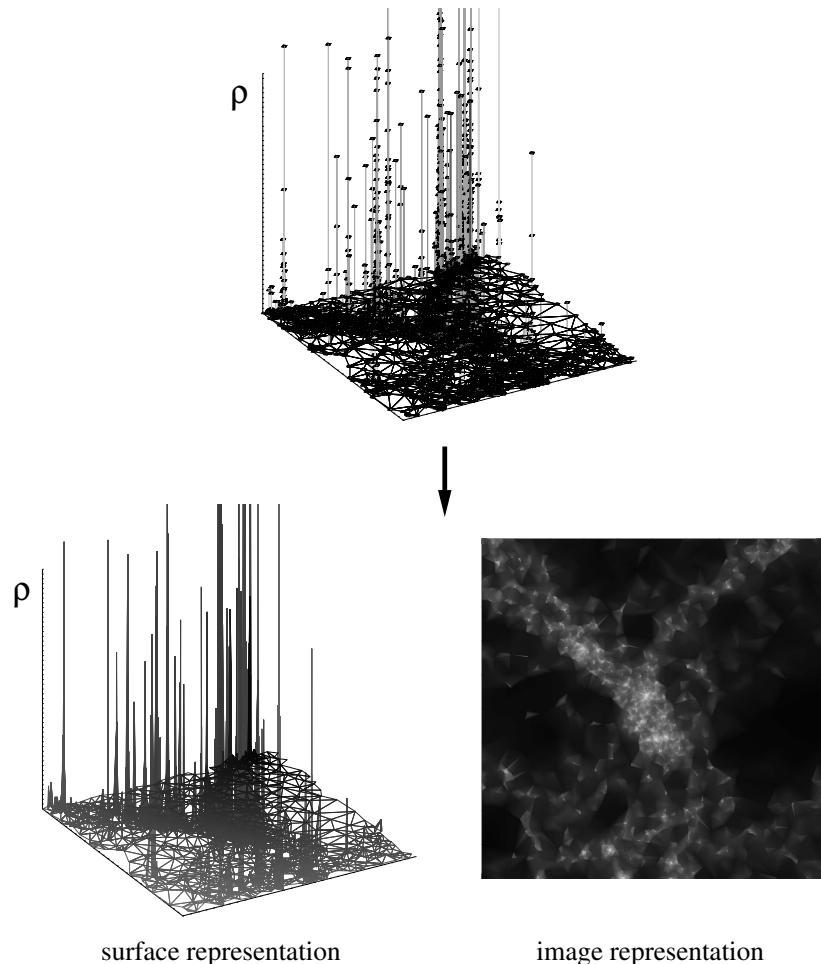


Figure 3.14 — Step 3 of the DTFE reconstruction procedure. The volume-covering density field is determined by linearly interpolating the density estimates of step 2 to any other point in space. In the bottom frames a surface plot and an image of the resulting density field are shown.

DTFE reconstruction procedure:

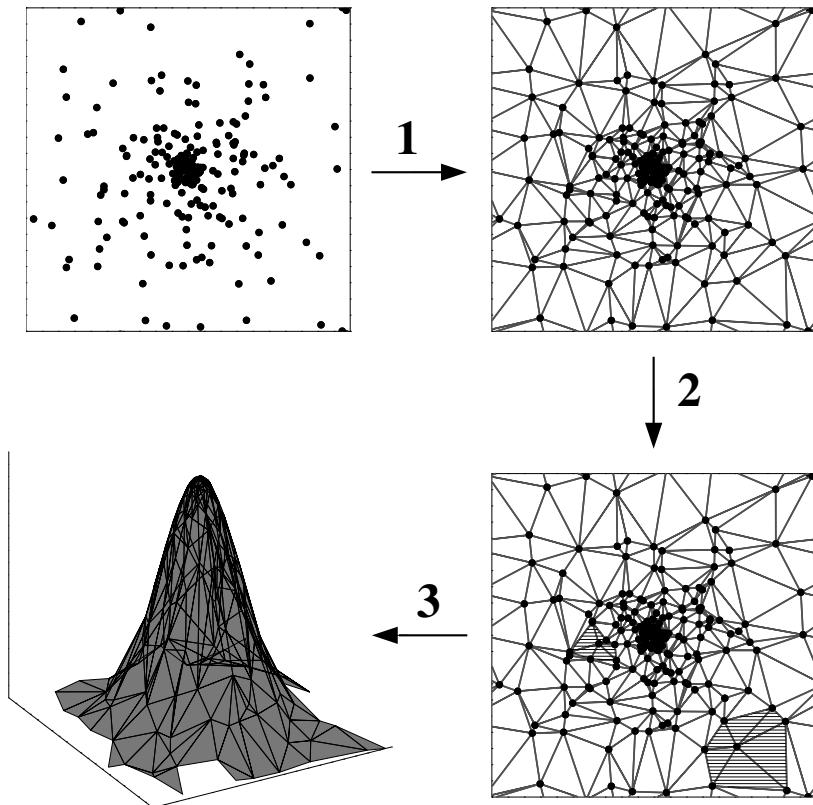


Figure 3.15 — Overview of the first-order DTFE reconstruction procedure. Given a point distribution (top left), one has to construct its corresponding Delaunay tessellation (top right), estimate the density at the position of the sampling points by taking the inverse of the area of their corresponding contiguous Voronoi cells (bottom right) and finally to assume that the density varies linearly within each Delaunay triangle, resulting in a volume-covering continuous density field (bottom left).

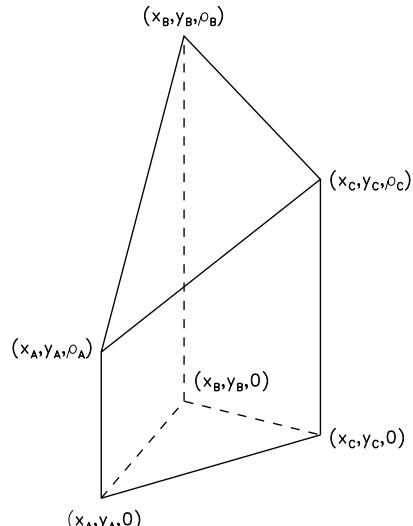


Figure 3.16 — The polyhedron with a triangular base to which each Delaunay triangle in two dimensions corresponds in (\mathbf{x}, ρ) -space. The volume of this polyhedron is equal to the mass corresponding to the triangle.

2. Estimate the value of the density field $\widehat{\rho}_i$ at the locations $\{\mathbf{x}_i\}$ of the sampling points using the Delaunay density estimator:

$$\widehat{\rho}(\mathbf{x}_i) = \frac{(D+1)m_i}{V(\mathcal{W}_i)}. \quad (3.36)$$

Here D is the dimension of space, m_i denotes the mass of sampling point i and $V(\mathcal{W}_i)$ the volume of contiguous Voronoi cell \mathcal{W}_i . Result: set of estimated density field values $\{\widehat{\rho}_i\}$;

3. Interpolate these values to any other point in space by assuming the density field varies linearly inside each Delaunay triangle. Result: reconstructed density field $\widehat{\rho}(\mathbf{x})$.

3.8.2 Mass conservation

The integral in Eqn. 3.33 is equal to the volume below the linearly varying ρ -surface in (\mathbf{x}, ρ) -space. In this space each Delaunay hyper-tetrahedron is the base hyper-plane of a polyhedron. In Fig. 3.16 an example of such a polyhedron in two dimensions is shown.

The total mass corresponding to the density field may therefore be written as the sum of the hyper-volume of these polyhedra,

$$M = \sum_{i=1}^{N_T} V(\Pi_i). \quad (3.37)$$

Here N_T is the total number of Delaunay hyper-tetrahedra in the tessellation and $V(\Pi_i)$ is the hyper-volume of polyhedron Π_i . This volume may be written as the average density at the vertices of Delaunay hyper-tetrahedron \mathcal{T}_i times the hyper-volume of this hyper-tetrahedron

$$V(\Pi_i) = \frac{1}{D+1} (\rho_{i0} + \rho_{i1} + \dots + \rho_{iD}) V(\mathcal{T}_i). \quad (3.38)$$

Here points $\{i0, i1, \dots, iD\}$ is the set of vertices of the hyper-tetrahedron S_i . $V(\mathcal{T}_i)$ denotes the hyper-volume of \mathcal{T}_i . The total mass corresponding to the density field may therefore be written as

$$M = \frac{1}{D+1} \sum_{i=1}^{N_T} (\rho_{i0} + \rho_{i1} + \dots + \rho_{iD}) V(\mathcal{T}_i). \quad (3.39)$$

This sum can be re-ordered and taken over each point instead of over each hyper-tetrahedron,

$$M = \frac{1}{D+1} \sum_{i=1}^N \rho_i \sum_{j=1}^{N_{T,i}} V(\mathcal{T}_{j,i}). \quad (3.40)$$

Here $\mathcal{T}_{j,i}$ denotes the set of hyper-tetrahedra with point i as one of their vertices and $N_{T,i}$ is the number of such hyper-tetrahedra. The set $\mathcal{T}_{j,i}$ is the contiguous Voronoi cell \mathcal{W}_i of point i . Eqn. 3.40 may therefore be written as

$$M = \frac{1}{D+1} \sum_{i=1}^N \rho_i V(\mathcal{W}_i). \quad (3.41)$$

From this equation it may be appreciated that the total mass M is equal to the total mass of the sampling points $\sum_{i=1}^N m_i$ if the density at the location of the sampling points is defined as

$$\widehat{\rho}(\mathbf{x}_i) = (D+1) \frac{m_i}{V(\mathcal{W}_i)}. \quad (3.42)$$

The density at the locations of a sampling points should therefore be taken equal to the inverse of its contiguous Voronoi cell times some normalization constant which only depends on the dimension of space. This normalization factor corresponds to the number of times each hyper-tetrahedron is used in obtaining an estimate of the density at the location of a sampling point.

From Eqn. 3.41 it follows that the mass corresponding to each sampling point is distributed fully over its contiguous Voronoi cell, showing that the DTFE is a strictly local reconstruction scheme. One can see this from the fact that a particular sampling point only contributes to the mass of those triangles for which the sampling point is a vertex.

It is important to note that the zeroth-order Voronoi density estimator $\widehat{\rho}_i = m_i/V(\mathcal{V}_i)$ (Eqn. 3.35) cannot be used in combination with the first-order interpolation scheme described in section 3.5.2. The reason for this is that this combination would not ensure that the total mass corresponding to the reconstructed field is equal to the total mass of the sampling points. From Eqn. 3.41 it follows that this would only be true if the following condition holds:

$$V(\mathcal{V}_i) = \frac{V(\mathcal{W}_i)}{D+1}. \quad (3.43)$$

In words, this would be the case if the volume of each Voronoi cell is equal to the volume of its corresponding contiguous Voronoi cell divided by a constant which depends on the dimension of space. This is, however, in general not the case. In Fig. 3.17 a counter-example in two dimensions is shown for which $V(\mathcal{W}_i) = 2V(\mathcal{V}_i)$ instead of $V(\mathcal{W}_i) = 3V(\mathcal{V}_i)$.

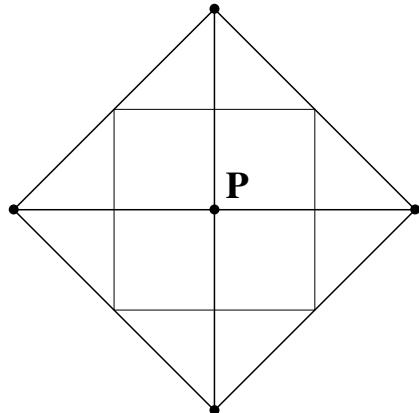


Figure 3.17 — Demonstration that the area of the contiguous Voronoi cell corresponding to a point is not necessarily exactly $(D + 1)$ times larger than the area of its ‘regular’ Voronoi cell. In the example shown here ($D = 2$) the area of the contiguous Voronoi cell corresponding to the central point is twice as large as the area of its regular Voronoi cell.

3.8.3 Higher-order density estimators

In principle it is possible to generalize the above zeroth- and first-order reconstruction procedures to higher order. In this way the reconstructed field would be differentiable up to a higher order. The extension of density field reconstruction schemes to a higher-order schemes based on splines and natural neighbor interpolation is currently in progress for astronomical purposes. However, it is not a trivial task because one needs a recipe for the estimation of the density at the location of the sampling points which guarantees that the total mass corresponding to the reconstructed field is equal to the total mass of the sampling points. The implementation of a scheme based on natural neighbor interpolation is further complicated by the fact that unlike in two dimensions, it has not yet been efficiently implemented in three dimensions.

3.9 Effective smoothing kernels and resolution

In essence, a density reconstruction scheme distributes the mass corresponding to a sampling particle over a region of space according to a distributing function $\mathcal{F}_i(\mathbf{x})$, whose shape generally may be different for each particle i :

$$\rho(\mathbf{x}) = \sum_{i=1}^N m_i \mathcal{F}_i(\mathbf{x}), \quad (3.44)$$

with

$$\int d\mathbf{x} \mathcal{F}_i(\mathbf{x}) = 1 \quad \forall i. \quad (3.45)$$

We will refer to generic shape of the function \mathcal{F} as the effective smoothing kernel corresponding to a reconstruction scheme. It is interesting to compare the effective smoothing kernels of the VTFE and DTFE with those of conventional reconstruction methods.

3.9.1 The TSC kernel

The value of the effective smoothing kernel of a grid-based method at the location of a sampling point follows from Eqn. 3.25:

$$\mathcal{F}_{\text{grid}}(\mathbf{x}_i) = \frac{M^3}{N} W\left(\mathbf{x}_i - \frac{\mathbf{n}}{M}\right). \quad (3.46)$$

The precise shape of the smoothing kernel on non-grid locations depends on the adopted interpolation scheme. For our comparison study we have used the smoothing scheme given by Eqn. 3.9, in combination with the TSC kernel function.

3.9.2 The SPH kernel

From Eqn. 3.30 it may be seen that the effective smoothing kernel of an SPH-like reconstruction scheme is simply its smoothing function $W(\mathbf{x}; h)$:

$$\mathcal{F}_{\text{SPH},i}(\mathbf{x}) = W_i(\mathbf{x}; h), \quad (3.47)$$

in which h is the smoothing length which varies as a function of the position \mathbf{x} . The shape of the smoothing kernel depends on the adopted SPH scheme. The standard shape of a spherically symmetric spline kernel is given by Eqn. 3.31.

3.9.3 The VTFE and DTFE kernels

The effective smoothing kernels of the VTFE and DTFE may be derived from the formalism described in section 3.5. The definition of a multi-dimensionally interpolated density field (Eqn. 3.13) is:

$$\rho(\mathbf{x}) = \sum_{i=1}^N \alpha_i(\mathbf{x}) \rho(\mathbf{x}_i), \quad (3.48)$$

in which $\alpha_i(\mathbf{x})$ is the weight of sampling particle i at position \mathbf{x} . Plugging in the VTFE definition of the density field at the location of the sampling particles yields the following expression for its effective smoothing kernel:

$$\mathcal{F}_{\text{VTFE},i}(\mathbf{x}) = \alpha_i(\mathbf{x}) V(\mathcal{V}_i)^{-1}, \quad (3.49)$$

in which \mathcal{V}_i is the Voronoi cell of sampling particle i .

From this equation one may observe that the difference between the effective smoothing kernels \mathcal{F}_i and the interpolation weights α_i is a normalization factor (which is different for each individual particle): the interpolation weights are normalized such that their sum is unity at each position, while the smoothing kernels are normalized such that the integral of each individual kernel over space is unity.

From Eqn. 3.16 it follows that the effective smoothing kernels of the VTFE are given by

$$\mathcal{F}_{\text{VTFE},i}(\mathbf{x}) = \begin{cases} V(\mathcal{V}_i)^{-1} & \text{if } \mathbf{x} \in \mathcal{V}_i \\ 0 & \text{if } \mathbf{x} \not\in \mathcal{V}_i \end{cases} \quad (3.50)$$

A similar reasoning as for the VTFE yields the following expression for the effective DTFE smoothing kernel:

$$\mathcal{F}_{\text{DTFE},i}(\mathbf{x}) = \begin{cases} (D+1) \alpha_i(\mathbf{x}) V(\mathcal{W}_i)^{-1} & \text{if } \mathbf{x} \in \mathcal{W}_i \\ 0 & \text{if } \mathbf{x} \not\in \mathcal{W}_i \end{cases} \quad (3.51)$$

in which \mathcal{W}_i is the contiguous Voronoi cell of sampling point i . The weights $\alpha_i(\mathbf{x})$ may be derived from Eqns. 3.19 and 3.20.

3.9.4 Comparison

We have compared the shape of the effective smoothing kernels in a number of distinct characteristic geometric configurations. In Fig. 3.18 we show the shape of the effective kernels for the following three environments: (1) a spherically symmetric Gaussian peak (left-hand section), (2) an anisotropic filamentary object (central section) and (3) a low density void (right-hand section). These three environments are characteristic elements of the large scale galaxy distribution.

In the left-hand section of Fig. 3.18 the effective smoothing kernel corresponding to the most central point in a Gaussian peak is shown. The peak has been sampled by 50 points, shown in the upper left-hand frame of the figure. In this frame the basic underlying geometric structure of the reconstruction procedures are also shown. The grid corresponds to the TSC reconstruction procedure. Its size has been chosen such that the peak density corresponds to 5 times the background density. The circle corresponds to the SPH reconstruction procedure and refers to the radius h as defined in Eqn. 3.30. Only particles within a distance of $2h$ of a location contribute to the value of the density field at that location. Finally, also the Voronoi cell and the contiguous Voronoi cell are shown which are used in the VTFE and DTFE procedures. In the subsequent frames the smoothing kernels corresponding to the different reconstruction procedures are shown. Note that the units of the vertical axis are different in each of these frames: the DTFE plot runs to $z = 800$, the SPH plot to $z = 80$ and the TSC plot to $z = 20$. Since the kernels are all normalized so that their volume integral is unity, it follows that the VTFE and DTFE kernels are the most localized ones, while the TSC kernel is the most extended one. This can also be appreciated from the illustration: the DTFE kernel extends only to the neighboring particles, while here the TSC kernel smears the mass of the most central point out over the whole structure. The extent of the SPH kernel is somewhere in between.

In the central section of Fig. 3.18 the kernels of a central point in a filamentary structure are shown. The number of points has been chosen such that the number of grid-cells for the TSC reconstruction procedure is the same as in the case of the Gaussian peak. This amounts to a peak density of the filament of about 3 times the background density. The VTFE and DTFE kernels for the filament look very similar to those in the Gaussian peak. This reflects the fact that the central densities in the filament and in the Gaussian peak have the same order of magnitude. The SPH kernel, on the other hand, is more extended than in the previous case. This may be appreciated both from the circle in the upper left-hand frame as well as from the kernel itself. The reason for this is the rigid circular geometry of the SPH kernel, which together with the condition of a fixed number of particles and the relatively lower density in y -direction leads to a kernel extending over a larger area. By construction the TSC kernel is the same as in the Gaussian peak, extending over a relatively large area and with a poor spatial resolution.

The kernels for the void are illustrated in the right-hand section of Fig. 3.18. The void is represented by a low density inner region surrounded by a high density ring. The void itself has a density which is equal to the background density. The number of sampling points has been set such that the grid-size in the TSC reconstruction procedure is the same as in

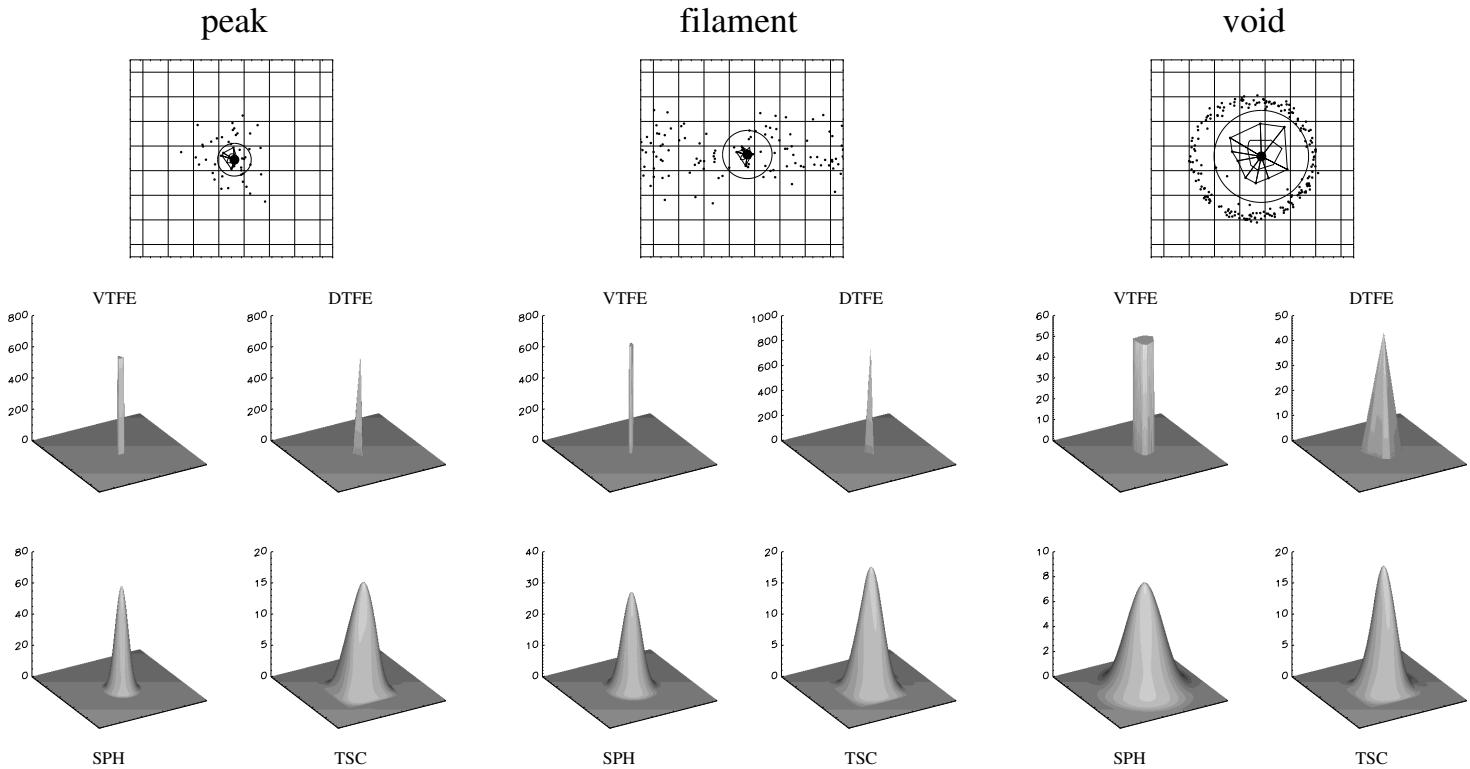


Figure 3.18 — Comparison of the effective smoothing kernels of the VTFE, DTFE, SPH and TSC reconstruction procedure inside three characteristic environments: a Gaussian peak (left-hand section), an anisotropic filament (central section) and a low density void (right-hand section). In the top row the corresponding point distributions are shown, with the basic underlying geometric structures of the reconstruction procedures overlaid. The thick dot denotes the most central point for which the smoothing kernel has been determined. The grid corresponds to the TSC procedure. The circle refers to the radius h as defined in Eqn. 3.30. Only points within a distance of $2h$ of a location contribute to the value of the density field at that location. The Voronoi cell and the contiguous Voronoi cell correspond to the DTFE and VTFE procedures. In the bottom frames the effective smoothing kernels are shown. Note that the units of the vertical axis are different in each frame.

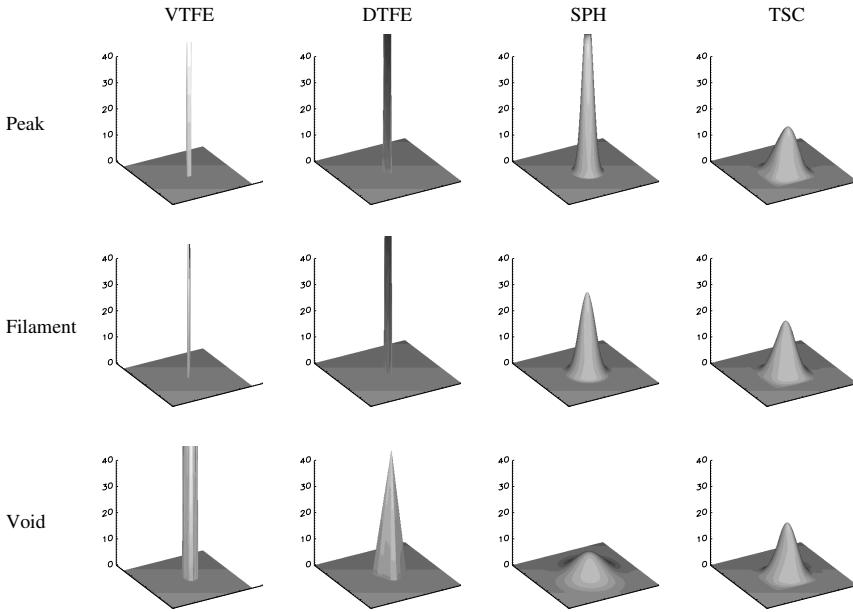


Figure 3.19 — Comparison of the effective smoothing kernels of the VTFE, DTFE, SPH and TSC reconstruction procedure for the most central point in the three different structures shown in Fig. 3.18. From top to bottom: Gaussian peak, filamentary structure and void-like structure. From left to right: DTFE, SPH and TSC. The scale is equal in all frames.

the previous cases. Here the VTFE and DTFE smoothing kernels are much more extended, reflecting the lower density inside the void. Note that the mass of the most central point is smeared out over a region which lies entirely within the void. This is not the case for the SPH smoothing kernel, which is non-zero inside a circular region with a radius twice as large as the circle displayed in the upper left frame. This means that part of the mass of the most central point in the void is added to the density of the filamentary structure surrounding the void! Conversely, part of the mass of the points inside this surrounding filamentary structure is smeared out into the outer regions of the void. This effect concerns many particles and is therefore substantially larger, causing a relatively high density in the outer regions of the void. This process and its repercussions are studied in more detail in Chapter 5. Even though the point distributions in Fig. 3.18 look very different, the effective TSC smoothing kernel is the same. The reason for this is that the effective resolution is set by the grid, not by the local density. The geometry of the grid also causes the TSC kernel to include artificial anisotropies. In other words, the TSC smoothing kernel does not adapt to the local environment.

To be better able to appreciate how the effective smoothing kernels adapt to the local point distribution, we have plotted in Fig. 3.19 all kernels with the same units on the z -axis. In the upper row the kernels correspond to the Gaussian peak, in the central row to the filamentary structure and in the bottom row to the void-like structure. Evidently, the TSC kernel is the same for all configurations. The resolution is set by the total number of sampling points. It is roughly equal to two to three grid-cells (linearly), depending on its precise definition. The

SPH smoothing kernel adapts to the local environment: from very localized and peaked for the Gaussian peak to very extended for the void-like structure. However, its shape does not adapt and always remains the same: the mass of a point is smeared out over a circular area. The VTFE and DTFE smoothing kernels also adapt to the local environment, but in comparison with the SPH smoothing kernel it is much more localized. The shape of the VTFE and DTFE kernels is also not rigid, but can be seen to adapt to the local environment. The SPH, VTFE and DTFE resolution varies over the reconstructed maps, with the resolution highest in high density regions and lowest in low density regions. From Fig. 3.19 we may conclude that the SPH resolution is always worse than the VTFE and DTFE resolution. The main difference between the VTFE and DTFE kernel is that the VTFE kernel is slightly more localized, while the DTFE kernel is continuous.

3.10 Processing of reconstructed fields

Once a density field has been reconstructed it may subsequently be processed. Processing may involve various operations. The most important ones are profile or image construction and, subsequently, filtering. Other relevant processing operations include various statistical analyses and feature detection. Here we shortly discuss each of these.

3.10.1 Profile and image construction

In practice it often occurs that one wants to obtain a one-dimensional section (profile) or a two-dimensional slice (image) at a regular grid of locations through a reconstructed field. Almost without exception these profiles or images are discretized and comprise field values at a finite set of regular intervals, locations or grid-locations. It is important to realize that this involves an either one- or two-dimensional display-grid which is different from the sampling-grid used in the grid-based methods. In general, the display-grid will be much finer than the sampling-grid. An illustration of a two-dimensional display-grid used to construct a two-dimensional image of a reconstructed density field is shown in Fig. 3.20 for the VTFE, the DTFE and a grid-based method. The display-grid is superimposed on the basic underlying structure used to reconstruct the density field.

Following the VTFE and DTFE reconstruction procedures described in sections 3.7 and 3.8, the image production involves a large number of interpolations onto a set of regularly spaced locations. Here a practical complication occurs. In order to perform the linear interpolation at the locations of the display-grid-points, it is necessary to determine inside which Voronoi cell (VTFE) or Delaunay tetrahedron (DTFE) each display-grid-point lies. In principle a brute force summation over all Voronoi cells or Delaunay tetrahedra is possible, but this becomes prohibitively expensive for a large number of display-grid-points. Fortunately a number of more efficient algorithms exist.

For the VTFE this is rather straightforward. On behalf of the definition of a Voronoi cell one identifies for each image location the nearest sampling point. A very efficient way of doing this is by making use of multi-dimensional binary or $k - d$ trees (Bentley 1975, Friedman, Bentley & Finkel 1977, van de Weygaert 1987). For the DTFE, however, knowing the nearest sampling point will not suffice in determining inside which Delaunay triangle a particular display-grid-point lies. This is because the nearest sampling point may not always be located in the same Delaunay triangle as the display-grid-point (see Appendix 3.A for a

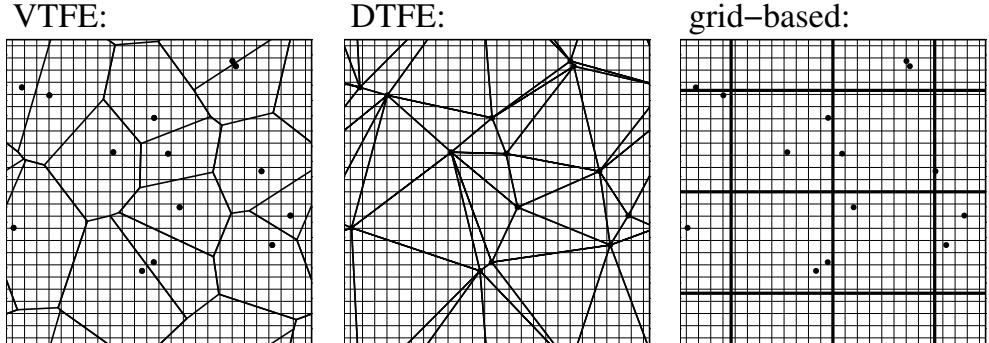


Figure 3.20 — Two-dimensional display-grids in the VTFE, DTFE and grid-based reconstruction methods. The grid is overlaid on top of the basic underlying structure used to reconstruct the density field. SPH-like methods are not shown, because of the inherent difficulty in visualizing their underlying structure, which does not consist of a subdivision of space in distinct non-overlapping structural elements, but of circles of different radius at each position in space.

demonstration of this point). This prompted us to use a more appropriate and efficient method, the *walking triangle algorithm*. For the case of two dimensions it was developed by Lawson (1977), while a generalization to three dimension has been implemented by Sambridge et al. (1995). In this algorithm one starts with an initial guess tetrahedron. The algorithm then checks if the point under consideration lies inside this triangle. If this is not the case, the algorithm jumps to an adjacent triangle in such a way that an almost linear path from the initial guess triangle to the final triangle is taken. For a more detailed description of this algorithm in two and three dimensions we refer the reader to Appendix 3.A.

As an example, in Fig. 3.21 a point distribution and a two-dimensional image through the three-dimensional DTFE reconstructed density field is shown. The depicted image is a direct DTFE reconstruction of the density field without the application of filtering or other processing procedures. The configuration shown is part of the outcome of a 256^3 particles GIF cosmological N -body simulation of cosmic structure formation (Kauffmann et al. 1999). Clearly visible in both the particle distribution and the reconstructed density field is the presence of hierarchical substructure: inside the main anisotropic filamentary structure high density clumps are present. Note the cellular structure of the low density regions surrounding the filament may be appreciated better in the density field than in the particle distribution. In Chapters 5 and 6 we further explore the application of the DTFE to the analysis of cosmological simulations.

3.10.2 Filtering

An often applied operation to reconstructed density fields is that of filtering. Filtering is often applied to identify features or structures at a particular scale or to compare the reconstructed field with the outcome of analytical calculations, which usually involve volume-weighted quantities.

In simple Fourier filtering the density field is convolved with some filtering function $W(\mathbf{x})$

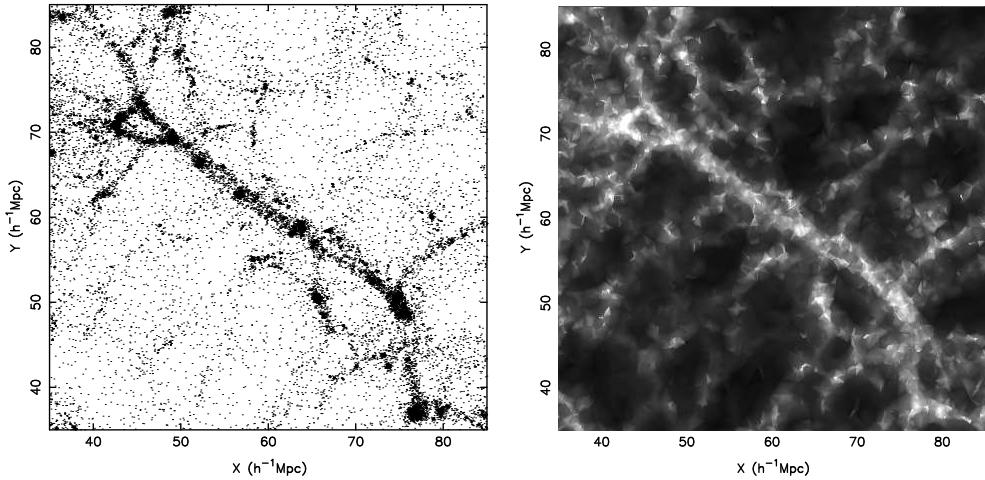


Figure 3.21 — Image of a characteristic filamentary region in the outcome of a cosmological N -body simulation. Left-hand frame: particle distribution in a thin slice through the simulation box. Right-hand frame: two-dimensional slice through the three-dimensional DTDE density field reconstruction.

conform Eqn. 3.12:

$$\widehat{f}_{\text{filtered}}(\mathbf{x}) = \frac{\int d\mathbf{y} \widehat{f}(\mathbf{y}) W(\mathbf{x} - \mathbf{y})}{\int d\mathbf{y} W(\mathbf{x} - \mathbf{y})}. \quad (3.52)$$

Common choices for the filter functions are Gaussian and top-hat filters, which in three dimensions are given by

$$\text{Gauss} \quad W = \frac{V}{(2\pi)^{3/2} R_G^3} e^{-r^2/2R_G^2}, \quad (3.53)$$

and

$$\text{top-hat} \quad W = \frac{3V}{4\pi R_T^3}. \quad (3.54)$$

Here R_G and R_T are the filtering radii, which define the resolution of the filtered fields.

As an example we have applied a Gaussian filter to the DTDE reconstruction of the filament shown in Fig. 3.21. In Fig. 3.22 the original DTDE density field is compared with the filtered fields for a number of increasing filtering radii. Only structures larger than the filtering radius are visible, while smaller structures are not resolved. This has the advantage that sampling noise is much less prominent. Also, by filtering one replaces the artificial triangular imprint of the DTDE kernel by a more natural smooth variation. However, such a smooth variation is enforced by the geometry of the smoothing kernel and is in this sense as artificial as the triangular imprint of the DTDE kernel. Clearly visible in the figure is that filtering causes anisotropic structures to appear more spherical. Also, the distinct cellular morphology of the low density regions surrounding the filament has due to the filtering operation changed into amorphous blobs with no clearly defined internal substructure.

More advanced filtering operations include Kalman filtering (Kalman 1960 and Kalman & Bucy 1961), of which the well-known Wiener filter is a special case. The Kalman filter

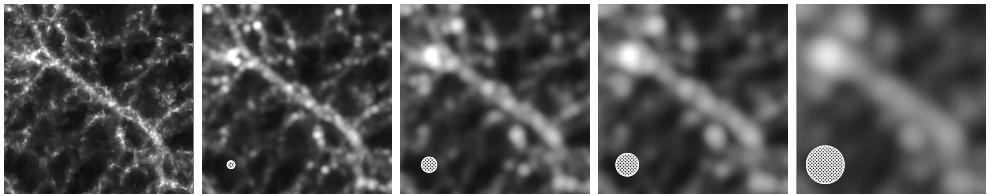


Figure 3.22 — Filtering of density fields. The left-hand frame depicts the original DTFE density field. The subsequent frames show filtered DTFE density fields. The FWHM of the Gaussian filter is indicated by the shaded circle in the lower left-hand corner of these frames.

is an efficient recursive filter for estimating the state of a dynamic system from a series of incomplete and noisy measurements. Similarly, the Wiener filter (Wiener 1949) filters out noise that has corrupted a signal by statistical means. In astronomy, the Wiener filter has found widespread use in a wide variety of applications such as the reconstruction of the cosmic web (e.g. Zaroubi et al. 1995, 1999, Erdogdu et al. 2004) and the analysis of the cosmic microwave background radiation (e.g. Bunn et al. 1996, Patanchon 2005).

3.10.3 Feature detection

A possible application of filtering is the identification of features or structures on a particular scale in the filtered field. Ramella et al. (2001), Kim et al. (2002), Marinoni et al. (2002) and Neyrick et al. (2005) have described cluster identification algorithms based on the Voronoi tessellation density estimator. They focussed on the problem of identifying clusters in imaging and redshift surveys.

Whereas the above described schemes are primarily limited to the identification of density peaks, a scheme based on the volume-covering DTFE should in principle be able to detect and discriminate between different kinds of structures (see Chapter 4). Recently, Aragón-Calvo et al. (2006) devised such a scheme, the *Multiscale Morphology Filter*, in which the input DTFE density field is smoothed over multiple scales. The resulting signal is subsequently processed through a morphology response filter. They showed that such a filter is able of identifying clusters, filaments and walls as geometrically distinct structures.

Another essential element of the large scale galaxy distribution are voids. They differ from clusters, filaments and walls by not being characterized by a certain configuration of galaxies, but instead by the absence of galaxies. Such regions are accurately described by the DTFE procedure, which is not the case for conventional density reconstruction schemes (see Chapters 5 and 6). Platen & van de Weygaert (2006) have used the DTFE density reconstruction as the input of a void identification scheme, the *Cosmic Watershed Algorithm*. In this scheme voids are defined as connected regions with a density smaller than a particular density threshold.

3.10.4 Statistical processing

The final processing operations we mention are various statistical analyses. These may include simple procedures, such as determining the probability distribution function, the two-point correlation function or the power spectrum of the density field. Examples of more advanced studies are wavelet decompositions of the density field (see e.g. Starck et al. 1998,

Donoho & Johnstone 1994, 1995) or topological analyses (see e.g. Shandarin et al. 2004). Clearly the quality of such analyses critically depends on the resolution of the input density field. As an example, in Chapter 4 we explicitly show that the DTFE is capable of tracking the two-point correlation function over much smaller scales than grid-based and SPH-like reconstruction methods.

3.11 Computational considerations

We have compared the computational demands of the DTFE with those of conventional methods. Below we have listed the sub-steps in which the characteristic application of the DTFE calculation of an image may be divided. We also have listed the scaling of each operation with the number of sampling points n .

1. Construction of the Delaunay tessellation: $O(n \log n)$;
2. Construction of the adjacency matrix: $O(n)$;
3. Calculation of the density values at each location: $O(n)$;
4. Calculation of the density gradient inside each Delaunay triangle: $O(n)$;
5. Interpolation of the density to an image grid: $O(ngrid^2 \cdot n^{1/2})$.

Depending on the implementation of step 1, step 2 may be automatically included in the construction of the Delaunay tessellation (see Appendix 3.A) and therefore omitted. The last step, the interpolation of the density to an image grid, is part of the post-processing operation and could be replaced by any other desired operation. Here it mainly depends on the number of grid-cells per dimension, denoted by $ngrid$.

The first step, the construction of the Delaunay tessellation corresponding to the sampling point distribution, is the most demanding in terms of computing speed and memory and dominates the total calculation time. The reason for the $O(n \log n)$ dependence is that the construction algorithm orders the points in a multidimensional binary tree (Bentley 1975, Friedman, Bentley & Finkel 1977, van de Weygaert 1987). The subsequent construction of the Delaunay tessellation itself is of order $O(n)$. The main problem which limits the number of points that can be used in the current implementation is however not the computing speed, but memory. We expect that by optimizing storage techniques, the amount of memory needed can be decreased with a factor of about two or three. Moreover, the memory issue could be completely circumvented by an algorithm which is currently under construction and which does not construct the complete Delaunay tessellation at the start of the image construction process, but instead only those triangles which are relevant around each grid-point on the fly.

Grid-based procedures are of order $O(n)$, while adaptive SPH-like reconstruction schemes have a similar demand on computing speed as the DTFE scheme. They are both of order $O(n \log n)$. The reason for this is that in SPH-like algorithms the points are also ordered in a multi-dimensional tree. In Table 3.1 we have listed for the examples of a set of 100 000 and of 200 000 points in two dimensions the amount of CPU time and memory needed for the calculation of an image on a Linux PC with a 2 Ghz processor and 1 Gb of internal memory. Clearly the TSC scheme performs best in terms of demands on both computing speed and memory,

Table 3.1 — Overview of the computational demands for several reconstruction procedures. Listed are the CPU time and memory needed to reconstruct the density field on a two-dimensional image.

# points	resource	TSC	SPH	DTFE
100 000	CPU time	1.4 s	6.0 s	23 s
	memory	5 Mb	15 Mb	97 Mb
200 000	CPU time	2.1 s	7.5 s	58 s
	memory	7.3 Mb	25 Mb	194 Mb

while the DTFE procedure performs the worst. It should be mentioned though that our current implementation of the two-dimensional Delaunay tessellation construction routine has not been optimized for computational speed. Currently there are algorithms available which are about a factor of 4-6 faster, making the speed comparable to our SPH implementation. We have already mentioned the possibilities for improving on the memory demands.

3.12 Summary, discussion and outlook

In this chapter we have described the fundamentals of the VTFE and DTFE. These are stochastic-geometrical methods to reconstruct density or intensity fields which are sampled by an irregularly distributed discrete point set. The core of these procedures is formed by the Voronoi (VTFE) and Delaunay (DTFE) tessellation of the discrete point set. Both methods conserve mass, yield volume-covering fields and are valid in any number of dimensions. The VTFE is a zeroth-order scheme, while the DTFE is a first-order linear scheme which produces continuous fields. Unlike conventional methods, such as fixed grid-based or adaptive SPH-like schemes, the VTFE and DTFE are completely natural methods in the sense that they do not invoke any user-specified parameters. In particular, they do not make use of a pre-specified smoothing kernel, which allows them to adapt to both the local density and geometry of the distribution of sampling points. Another important property of these methods is that they produce volume-weighted fields. This is in contrast with conventional methods, which tend to yield mass-weighted fields, complicating the comparison with analytical calculations.

We have compared the effective smoothing kernels of the VTFE and DTFE to the smoothing kernels of conventional procedures. We have shown that the VTFE and DTFE procedures attain the highest spatial resolution. In particular, since their kernels adapt to the local density of the sampling point distribution, they resolve both high and low density regions at the same time. The VTFE and DTFE kernels do not only adapt to the local density of the sampling point distribution, but also to its local geometry. This allows these methods to resolve anisotropic structures, which conventional reconstruction procedures tend to smear into a large volume, effectively making them more spherical.

We have discussed several post-processing procedures which may be applied to DTFE reconstructed density fields, such as profile and image construction and various statistical tests. Here we want to discuss in more detail the post-processing procedures which involve getting rid of the effects of sampling noise. For example, one may want to determine the radial mass profile of certain objects, or to determine their real underlying density profile (without a noise signal superposed). This may be done by ‘averaging’ over the noise by integrating the

reconstructed field over concentric shells or by the fitting of a model. The latter application is described in Chapter 6.

One of the most important post-processing procedures which increases the signal-to-noise ratio of a reconstructed density field is smoothing. Obviously, smoothing destroys some of the adaptive resolution and shape of the effective VTFE and DTFE kernels. Nevertheless, in section 3.4 we have argued that the combined use of the VTFE or DTFE reconstruction procedure and smoothing has an important conceptual advantage over conventional reconstruction procedures. These procedures involve a unclear mix of the two different mathematical procedures of reconstruction and filtering, which leads to interpretational difficulties with the danger of confusing mass-weighted and volume-weighted results. Beside the fuzzy nature of the resulting operation any subsequent smoothing represents therefore an additional smoothing on top of the primary smoothing which is hidden in the reconstruction procedure. As these two smoothing operations are not necessarily correlated, this complicates a subsequent analysis by a considerable amount.

In the VTFE and DTFE procedures the field reconstruction and any subsequent smoothing operation are clearly defined and distinct mathematical operations, making the interpretation of the results much more transparent. Examples of applications which involve smoothing include the normalization of cosmological numerical simulations through the rms variation of the density field when smoothed with a filter of a certain radius (e.g. Efstathiou et al. 1990, Saunders et al. 1991) or the comparison of simulated or observed cosmological quantities with analytical predictions. For example, Bernardeau & van de Weygaert (1996) discuss the determination of the volume-averaged expansion scalar (divergence of the peculiar velocity field) as well as the shear and vorticity of the velocity field. They concluded that the use of volume-weighted Voronoi and Delaunay interpolation techniques gives a significantly better agreement with theoretical predictions than analyses based on conventional, mass-weighted filters.

A very interesting possibility is the development of a natural smoothing procedure with a kernel function whose extent and shape adapts to the local density and geometry of the density field. In such a way the adaptive nature of the VTFE and DTFE reconstruction procedures would be preserved and such a filter would be a natural complementation to these procedures. A possible way of doing this is by making use of the properties of the tessellation beyond the contiguous Voronoi cell of each sampling point. Another interesting possibility is the use of multiscale morphology filters (Aragón & van de Weygaert 2006), in which the density field is smoothed over multiple scales.

Apart from the examples discussed here many other ways of post-processing VTFE and DTFE reconstructed fields in other to get rid of undesirable noise effects exist. A particularly promising example that we want to mention here is Kalman filtering (Kalman 1960, see also e.g. Maybeck 1979). This technique is an optimal recursive data processing algorithm, in which on the basis of statistical knowledge of the system noises and a priori knowledge of the system an optimal ‘noise-free’ estimate for the field comes out. VTFE and DTFE reconstructed fields could serve as the input for such an algorithm. An astronomical example of the use of Kalman filtering is the Wiener reconstruction of large-scale structure from observational surveys (e.g. Lahav et al. 1994, Zaroubi et al. 1999). Instead of devising smart post-processing algorithms an alternative approach may be to extend the VTFE and DTFE procedures by implementing one of the higher-order reconstruction schemes discussed in section 3.8.3.

The development of the DTFE was motivated by our interest in cosmological structure formation, in particular in describing the high density and anisotropic filamentary and wall-like regions, which conventional reconstruction methods fail to describe. In the rest of this thesis we present a number of examples of applications of the DTFE to such studies. In Chapter 4 we apply the DTFE to the analysis of cosmological simulations. The cosmological large scale structure distribution does not consist of isolated objects of a certain density or geometry, but instead represents a complex network of objects of different densities and scales, often dubbed the cosmic web or cosmic foam (e.g. Bond et al. 1996, van de Weygaert 2002). It is thus important to test whether the reconstruction method used is able of dealing with a density field with a hierarchy of spatial scales, geometries and densities. Therefore we first present a thorough assessment of the behavior of the DTFE in such circumstances in Chapter 4. In Chapter 5 we discuss the possibility of incorporating the DTFE in hydrodynamical simulation algorithms. Bernardeau & van de Weygaert (1996) and Schaap & van de Weygaert (2003) have shown that the Delaunay interpolation procedure is also very well suited for reconstructing the cosmological velocity field. In Chapter 6 we show how the DTFE enables a simultaneous study of both the structure and dynamics of the cosmic foam and apply the DTFE to the dynamical modelling of characteristic elements of the large scale galaxy distribution such as voids and filaments. In Chapter 7 we apply the DTFE to the analysis of the 2dFGRS redshift survey (Colless et al. 2001).

These applications are only first examples of what could be achieved by using the DTFE. Indeed, several authors have already described its use in the literature. Ramella et al. (2001), Kim et al. (2002) and Marinoni et al. (2002). A similar algorithm for the analysis of integral-field spectroscopic data has been implemented by Cappellari & Copin (2003). Bradac et al. (2004) have shown that the use of the DTFE allows for a study of the signature of CDM substructure on gravitational lensing. Shandarin et al. (2004) advocate the DTFE for systematic studies of the size, shape and topology of the cosmic web by means of Minkowski functionals. Neyrinck et al. (2005) have used the DTFE to identify halos in cosmological *N*-body simulations. Many other applications are thinkable, such as the reconstruction of intensity fields. Zeroth-order Voronoi techniques to identify clusters in imaging and redshift surveys have been used by Arad et al. (2004, 2005) advocated the use of the DTFE for a six-dimensional analysis of the phase space density. Aragón-Calvo & van de Weygaert (2006) have devised a multiscale morphology filter with the DTFE density field as the input in order to detect and discriminate between clusters, filaments and walls in the large scale matter distribution. Platen & van de Weygaert (2006) have used the DTFE to implement a void detection algorithm which is based on the watershed algorithm. In conclusion, the DTFE has started to establish itself as a valuable tool for a wide array astronomical research applications.

References

- Arad I., Dekel A., Klypin A., 2004, MNRAS, 353, 15
- Arad I., Johansson P.H., 2005, MNRAS, 362, 252
- Aragón-Calvo M., van de Weygaert R., Jones B.J.T., van der Hulst J.M., 2006, astro-ph/0610249
- Babu G.J., Feigelson E.D., 1996, J. Statis. Planing Inference, 50, 311
- Bentley J.L., 1975, Comm. ACM, 18, 509
- Bernardeau F., van de Weygaert R., 1996, MNRAS, 279, 693
- Bertschinger E., 1998, ARAA, 36, 599

- Boissonnat J.-D., Cazals F., 2002, Comp. Geometry Theory and Applications, 22, 185
- Bond J.R., Kofman L., Pogosyan D., 1996, Nature, 380, 603
- Bradac M. et al., 2004, A&A, 423, 13
- Braun J., Sambridge M., 1995, Nature, 376, 655
- Brown G.S., 1965, New Zealand Forestry Service Research Notes, 38, 1
- Bunn E.F., Hoffman Y., Silk J., 1996, ApJ, 464, 1
- Cappellari M., Copin Y., 2003, MNRAS, 342, 345
- Colberg J.M., 1998, PhD Thesis, University of Muenchen, Muenchen, Germany
- Colless M. et al., 2001, MNRAS, 328, 1039
- Cressie N., 1993, *Statistics for Spatial Data, revised edition*, John Wiley & Sons, New York, USA
- Delone B.N., 1934, Bull. Acad. Sci. USSR: Classe Sci. Mat., 7, 793
- Daley D.J., Vere-Jones D., 1988, *Introduction to the Theory of Point Processes*, Springer, New York, USA
- Diggle P., 1983, *Statistical Analysis of Point Processes*, Chapman & Hall, London, UK
- Dirichlet G.L., 1850, J. Reine Angew. Math., 40, 209
- Donoho D.L., Johnstone I.M., 1994, Biometrika, 81, 425
- Donoho D.L., Johnstone I.M., 1995, JASA, 90, 432
- Duckaerts C., Godefroy G., Hauw J.-J., 1994, Journal of Neuroscience Methods, 51, 47
- Ebeling H., Wiedenmann G., 1993, Phys. Rev. E, 47, 704
- Efstathiou G. et al., 1990, MNRAS, 247, 10
- Erdođdu P. et al., 2004, MNRAS, 352, 939
- Friedman J., Bentley J.L., Finkel R.A., 1977, ACM Trans. Math. Softw., 3, 209
- Gerald C.F., Wheatley P.O., 1999, *Applied Numerical Analysis*, Addison Wesley, Reading, MA, USA
- Hernquist L., Katz N., 1989, ApJS, 70, 419
- Hockney R., Eastwood J., 1988, *Computer Simulations Using Particles*, Taylor & Francis, Bristol, PA, USA
- Icke V., van de Weygaert R., 1987, A&A, 184, 16
- Jenkins A. et al., 1998, ApJ, 499, 20
- Kalman R.E., 1960, Journal of Basic Engineering, 82D, 35
- Kalman R.E., Bucy R.S., 1961, Journal of Basic Engineering, 83D, 95
- Karr A.F., 1986, *Point Processes and their Statistical Inference*, Marcel Dekker, New York, USA
- Kauffmann G. et al., 1999, MNRAS, 303, 188
- Kiang T., 1966, Z. Astrophys., 64, 433
- Kim R.S.J. et al., 2002, AJ, 123, 20
- Lahav O., Fisher K.B., Hoffman Y., Scharf C.A., Zaroubi S., 1994, ApJ, 423, 93
- Lawson C.L., 1977, in *Mathematical Software*, Vol. 3, Rice J. (ed.), Academic Press, New York, USA
- Lombardi M., Schneider P., 2001, A&A, 373, 359
- Marinoni C., Davis M., Newman J.A., Coil A.L., 2002, ApJ, 580, 122
- Martinez V.J., Saar E., 2002, *Statistics of the Galaxy Distribution*, Chapman & Hall/CRC press, Boca Raton, USA
- Matthes K., Kerstan J., Mecke J., 1978, *Infinitely Divisible Point Processes*, Wiley, New York, USA
- Maybeck P.S., 1979, *Stochastic Models, Estimation, and Control*, Vol. 1, Academic Press, New York, USA
- Monaghan J.J., 1992, ARAA, 30, 543
- Monaghan J.J., Latanzio J.C., 1985, Astr. Ap., 149, 135
- Neyrinck M.C., Hamilton A.J.S., Gnedin N.Y., 2005, MNRAS, 362, 337
- Okabe A., Boots B., Sugihara K., Nok Chiu S., 2000, *Spatial Tessellations, Concepts and Applications of Voronoi Diagrams*, 2nd edition, Wiley, Chichester, UK
- Ord J.K., 1978, Mathematical Scientist, 3, 23
- Owen J.M., Villumsen J.V., Shapiro P.R., Martel H., 1998, ApJS, 116, 155
- Patanchon G., Cardoso J.-F., Delabrouille J., Vielva P., 2005, MNRAS, 364, 1185

- Peacock J.A., 1983, MNRAS, 202, 615
- Platen E., van de Weygaert R., 2006, *private communication*
- Press W.H., Teukolsky S.A., Vetterling W.T., Flannery B.P., 1992, *Numerical Recipes*, Cambridge University Press, Cambridge, UK
- Ramella M., Boschin W., Fadda D., Nonino M., 2001, A&A, 368, 776
- Ripley B.D., 1981, *Spatial Statistics*, Wiley, New York, USA
- Ripley B.D., 1988, *Statistical Inference for Spatial Processes*, Cambridge University Press, Cambridge, UK
- Sambridge M., Braun J., McQueen H., 1995, Geophys. J. Int., 122, 837
- Saunders W. et al., 1991, Nature, 349, 32
- Schaap W.E., van de Weygaert R., 2003, in *Statistical challenges in modern astronomy III*, Babu G.J., Feigelson E.D. (eds.), Springer-Verlag, New York, USA
- Schoenberg I.J., 1946a, Quart. Appl. Math., 4, 45
- Schoenberg I.J., 1946b, Quart. Appl. Math., 4, 112
- Shandarin S.F., Sheth J.V., Sahni V., 2004, MNRAS, 353, 162
- Shapiro P.R., Martel H., Villumsen J.V., Owen J.M., 1996, ApJS, 103, 296
- Sibson R., 1981, in *Interpreting Multi-Variate Data*, Barnett V. (ed.), Wiley, Chichester, UK
- Sloan S.W., 1987, Adv. Eng. Software, 9, 34
- Starck J.-L., Murtagh F., Bijaoui A., 1998, *Image processing and data analysis. The multiscale approach*, Cambridge University Press, Cambridge, UK
- Stoyan D., Kendall W.S., Mecke J., 1995, *Stochastic Geometry and its Applications*, Wiley, Chichester, UK
- Sukumar N., 1998, PhD Thesis, Northwestern University, Evanston, IL, USA
- van de Weygaert R., 1987, *The Multidimensional Binary Tree and its Use in Astronomy*, report M.Sc. research project, Leiden University, Leiden, The Netherlands
- van de Weygaert R., Icke V., 1989, A&A, 213, 1
- van de Weygaert R., 1991, PhD Thesis, Leiden University, Leiden, The Netherlands
- van de Weygaert R., 1994, A&A, 283, 361
- van de Weygaert R., 2002, *Froth across the Universe*, in ‘2nd Hellenic Cosmology Meeting’, Plionis M., Cotsakis S. (eds.), Kluwer Academic Publishers, Dordrecht, NL
- Voronoi G., 1908, J. Reine Angew. Math., 134, 198
- Whitehurst R., 1995, MNRAS, 277, 655
- Wiener N., 1949, *Extrapolation, interpolation and smoothing of stationary time series*, Wiley New York, USA
- Zaroubi S., Hoffman Y., Fisher K.B., Lahav O., 1995, ApJ, 339, 446
- Zaroubi S., Hoffman Y., Dekel A., 1999, ApJ, 520, 413

3.A The walking triangle algorithm

To be able to interpolate the density field to an arbitrary point in space using the DTFE, one has to determine inside which triangle that point lies. Naively one might think that this triangle must be one of the triangles of which the nearest sampling point is a vertex. However, this is not necessarily the case. Consider a case in which the circumcenter of a Delaunay triangle is situated outside the triangle. In Fig. 3.23 it may be appreciated that in those cases there is a region of space inside the adjacent triangle which is closest to the sampling point but which does not belong to that triangle.

For this reason we have used another algorithm to determine inside which triangle a particular location lies. The ‘walking triangle algorithm’ is an efficient and conceptually simple algorithm for this purpose. This algorithm has been developed by Lawson (1977) for the case

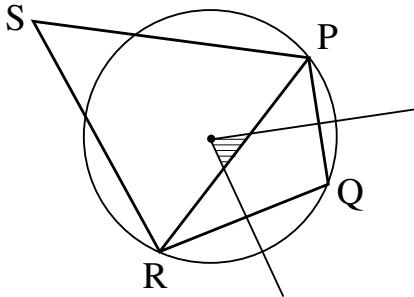


Figure 3.23 — Demonstration that the nearest sampling point to a particular point does not necessarily belong to the triangle inside which this point lies. Any point inside the shaded region lies inside triangle $\triangle PRS$, but is nearest to sampling point Q . This happens because the circumcenter of triangle $\triangle PQR$ does not lie inside $\triangle PQR$.

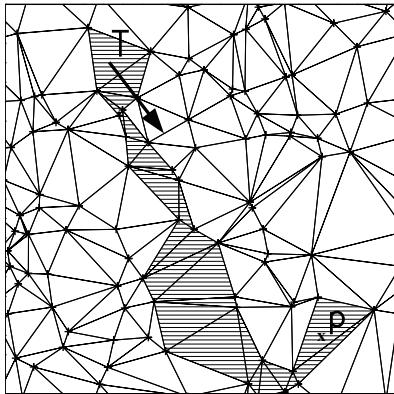


Figure 3.24 — The path taken by the walking triangle algorithm from the initial guess triangle T (upper left-hand corner) to the triangle containing point p (lower right-hand corner).

of two dimension (see also Sloan 1987). For this algorithm we need the *adjacency matrix* $\Omega(i, j)$, which specifies for each side j of each triangle i its adjacent triangle $\Omega(i, j)$. At the end of this appendix we discuss how one may calculate this matrix.

In the walking triangle algorithm, one starts with an initial guess triangle T (see Fig. 3.24). The sides should be defined in counterclockwise fashion (see Fig. 3.25). Then one tests if the point p lies inside T , by checking for each of its three sides whether p lies at its left-hand side (see Fig. 3.25). If this is not the case for a particular side i , then the algorithm immediately jumps to the adjacent triangle $\Omega(T, j)$ and starts testing the sides of this triangle in the same manner. This procedure continues until one arrives at the triangle inside which p lies. This is illustrated in Fig. 3.24.

The above procedure may be formulated in mathematical terms as follows. A point $p = (x, y)$ lies inside of a triangle if the following condition holds for each of its sides

$$(V_x^i - x)(V_y^j - y) > (V_x^j - x)(V_y^i - y). \quad (3.55)$$

Here $\mathbf{V}^i = (V_x^i, V_y^i)^T$ and $\mathbf{V}^j = (V_x^j, V_y^j)^T$ are the vertices at both ends of the side and the indices j and k are defined by cyclic rotation of the three indices (i, j, k) (see Fig. 3.25).

The walking triangle algorithm takes an almost direct path from the initial guess triangle to the final triangle as can be seen in Fig. 3.24. The speed of convergence is dependent on the linear distance between the initial guess triangle and the final triangle. The efficiency of the algorithm can be strongly improved by choosing an initial triangle close to the final triangle. In particular, if one calculates the density values at a regular grid of locations, the triangle

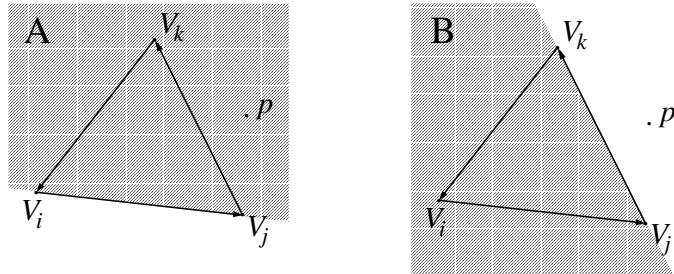


Figure 3.25 — A: point p lies on the left-hand side of side k . B: point p lies on the right-hand side of side i . A point lies inside a triangle if it lies on the left-hand side of all its three sides.

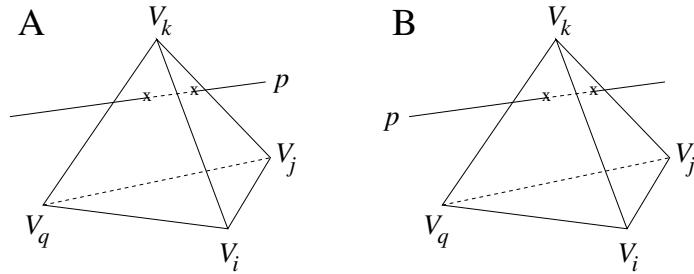


Figure 3.26 — A: point p lies on the inward side of wall j , but on the outward side of wall q . B: point p lies on the outwards side of wall j , but on the inward side of wall q . A point lies inside a tetrahedron if it lies on the inward side of all its four walls.

containing the previous point serves as a good initial guess triangle for the next grid-location.

In three dimensions a similar method can be used (Sambridge et al. 1995). In this case the concept of counterclockwise and left-hand and right-hand side do not make sense. Instead we check for each wall of a tetrahedron if the point $p = (x, y, z)$ lies inwards of the wall, which means at the same side of the wall as the nucleus of the tetrahedron not belonging to the wall (see Fig. 3.26). Mathematically, this condition may be formulated as

$$\begin{aligned} & [(\mathbf{V}^i - \mathbf{V}^j) \cdot [(\mathbf{V}^k - \mathbf{V}^j) \times (\mathbf{V}^l - \mathbf{V}^j)]] \\ & \cdot [(\mathbf{x} - \mathbf{V}^j) \cdot [(\mathbf{V}^k - \mathbf{V}^j) \times (\mathbf{V}^l - \mathbf{V}^j)]] \geq 0. \end{aligned} \quad (3.56)$$

Here \mathbf{V}^i , \mathbf{V}^j and \mathbf{V}^k are the three vertices of the four walls of the tetrahedron. Again the algorithm stops when the point lies inwards of all four walls of a particular tetrahedron and thus inside that tetrahedron.

3.A.1 The construction of the adjacency matrix

To be able to use the walking triangle algorithm, we need the adjacency matrix corresponding to the tessellation under consideration. The adjacency matrix $\Omega(i, j)$ specifies for each side j of each triangle i its adjacent triangle $\Omega(i, j)$. Ideally, this matrix would be determined while constructing the Delaunay tessellation. However, not all available routines also calculate the adjacency matrix.

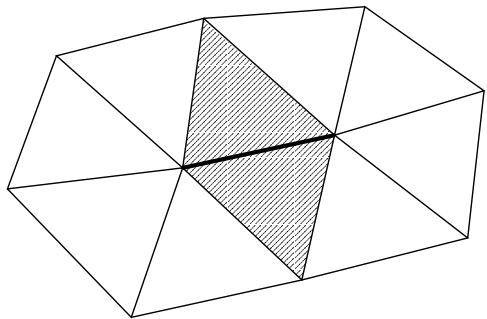


Figure 3.27 — The construction of the adjacency matrix. For each side of each triangles one checks which two triangles the two vertices of the side have in common. The thick line denotes the side under consideration. The triangles plotted are all the triangles with a vertex belonging to that side. The shaded triangles both have two vertices belonging to the side.

Of course, the matrix can be calculated by brute force. However, this becomes prohibitively expensive when the number of triangles and thus the number of points is large. Sambridge et al. (1995) describe an efficient alternative algorithm which scales linearly with the number of triangles. Given the triangles matrix $T(i, j)$, which specifies for each triangle the indices of its three vertices, one can construct a new matrix which specifies for each point the triangles of which it is a vertex. This can be done in a loop over the triangles matrix $T(i, j)$ by assigning triangle i to vertex j . The adjacency matrix can subsequently be constructed in a loop over all sides of all triangles. For each side one considers which two triangles the two vertices have in common. One is the triangle under consideration, the other its adjacent triangle. This process is illustrated in Fig. 3.27.

4

Resolving Complex Geometric Patterns and Hierarchical Substructure in the Cosmic Web

W.E. Schaap & R. van de Weygaert

ABSTRACT — We investigate the performance of the Delaunay Tessellation Field Estimator (DTFE) towards tracing two key aspects of the large scale cosmic matter distribution: the hierarchically structured matter distribution and the complex web-like pattern consisting of anisotropic filamentary and sheet-like features. The DTFE is a self-adaptive density estimation and multi-dimensional interpolation scheme which is based on the Delaunay tessellation of a given spatial point distribution. The DTFE does not make use of any artificial, user-specified filtering. The performance of the DTFE is compared with that of the conventional rigid grid-based TSC procedure and the scale-adaptive yet shape-insensitive SPH method. To test the ability to explore structural hierarchy, a set of three Soneira-Peebles self-similar point distributions is processed with the DTFE, SPH and TSC procedures. The scaling properties, indices and autocorrelation functions are assessed and compared to those of the generating point distribution. Three Voronoi clustering models – a wall, a filament and a cluster model – are used as templates to test the ability to trace anisotropic web-like patterns. The shape of flattened sheets and elongated filaments in the DTFE, SPH and TSC field reconstructions is confronted with the shape of the point distribution. Also the volume occupation of regions enclosed by iso-density contours is compared with that of other methods and the point distribution. The DTFE proves itself superior in both scale and shape resolution. It manages to reproduce the self-similar scaling of the Soneira-Peebles models to an impressive level, even successfully recovering the scaling indices. The study also shows its success in tracing web-like patterns. Not only the visual appearance of the DTFE reconstructions, but also the recovery of the shape and, to some lesser extent, the volume occupation of the web-like features demonstrates the potential of the DTFE as a key instrument in the study of the cosmic web. The DTFE is a natural and surprisingly accurate filtering and interpolation instrument for the study of the cosmic web.

4.1 Introduction

Three major characteristics of the large scale structure of the universe are the web-like spatial arrangement of galaxies and mass into elongated filaments, sheet-like walls and dense compact clusters, the existence of large near-empty void regions and the hierarchical nature of the mass distribution, marked by substructure over a wide range of scales and densities. According to the standard paradigm of cosmic structure formation, the theory of gravitational instability (Peebles 1980), this intricate spatial pattern has emerged as a result of the gravitational growth of tiny (Gaussian) density perturbations and the accompanying tiny velocity perturbations in the primordial universe.

Over the past decades this basic framework of cosmic structure formation has been shaped and confirmed by an almost revolutionary progress in observational cosmology, underpinned and supported by major advances in theoretical understanding guided by ever larger computer simulations of structure formation. Powerful new telescopes and detectors along a broad expanse of the electromagnetic spectrum have opened up our view of the infrastructure of our universe out to high redshifts. Of paramount significance for our understanding of cosmic structure formation is the information provided by large systematic galaxy redshift surveys, of which the 2dFGRS (Colless et al. 2001) and the Sloan Digital Sky Survey (Stoughton et al. 2002) are the most prominent examples. They have mapped the spatial distribution of hundreds of thousands to a million galaxies over huge swathes of the nearby universe. Since 1986, when the CfA redshift survey (de Lapparent, Geller & Huchra 1986) became available, we have learned that galaxies trace out an intricate spatial pattern throughout the observable universe. Since then, the LCRS, 2dFGRS and SDSS redshift surveys have confirmed and established the web-like spatial organization of cosmic matter on Megaparsec scales as a basic characteristic of our universe. These redshift surveys reveal a sponge-like arrangement, coined the cosmic web (Bond, Kofman & Pogosyan 1996, see also van de Weygaert 2002 and references therein), with galaxies aggregating in striking geometric patterns such as prominent filaments, vaguely detectable walls and dense compact clusters on the periphery of giant voids.

Perhaps as impressive as the advances in observations are those in theoretical and computer studies of the formation and evolution of large scale structure. Cosmological N -body and advanced statistical tests have acquired a central position within any viable cosmological study of the involved gravitational, hydrodynamical and radiative processes and in the interpretation of the multitude of complex cosmological observational results. Over the last decades N -body computer experiments have become considerably more complex and realistic (for an extensive review see e.g. Bertschinger 1998). The first cosmological simulations were performed in the seventies and consisted of only a couple of hundreds of particles, each representing huge dark matter concentrations (e.g. White 1976, Aarseth, Turner & Gott 1979). Awesome advances in computing power have lead to the current situation in which simulations of 512^3 dark matter particles are gradually becoming the norm, while state-of-the-art simulations supersede this number by orders of magnitude and in some cases have almost acquired the status of a true ‘universe in silicon’. The largest simulation carried out thus far, the millennium simulation (Springel et al. 2005), contains more than 10^{10} particles and has a dynamic range in excess of 10^5 along each dimension. Such simulations are able to simultaneously follow the development of large scale structure as well as the hierarchical build-up of galaxy clusters and galaxy haloes by the infall and merging of small dark matter haloes.

4.1.1 The cosmic web

A first important characteristic of the cosmic matter distribution is its web-like geometry, marked by highly elongated filamentary and flattened planar structures as well as dense compact clusters surrounding large near-empty void regions (see Fig. 4.1). As borne out by a large sequence of N -body computer experiments, web-like patterns in the overall cosmic matter distribution do represent a universal but possibly transient phase in the gravitationally driven emergence and evolution of cosmic structure. The N -body calculations have shown that web-like patterns defined by prominent anisotropic filamentary and planar features – and with characteristic large underdense void regions – are a natural manifestation of the gravitational cosmic structure formation process.

The existence of the cosmic web can be understood through the tendency of matter concentrations to contract and collapse gravitationally in an anisotropic manner. In a generic random density field the gravitational force field at any location will be anisotropic. For a particular structure the internal force field of the structure hangs together with the flattening of the feature itself. It induces an anisotropic collapse along the main axes of the structure. The more quiescent external ‘background’ force field, the integrated gravitational impact of all external density features in the universe, will also be anisotropic. It is this Megaparsec-scale tidal shear pattern which is the main agent for the contraction of matter into the filaments which trace out the cosmic web. Its link with the cluster distribution can be understood through the implied quadrupolar mass distribution, in which the clusters are to be found at the sites of the overdense patches (Bond, Kofman & Pogosyan 1996), van de Weygaert & Bertschinger 1996).

A second manifest feature of the large scale matter distribution is the marked and dominant presence of large roundish underdense regions, the voids. They form in and around density troughs in the primordial density field. These evolve in nearly empty void regions with sharply defined boundaries marked by filaments and walls. Their essential role in the organization of the cosmic matter distribution was recognized early after their discovery (Gregory & Thompson 1978, Einasto, Joeveer & Saar 1980, Kirshner et al. 1981, 1987, Icke 1984). Recently, their emergence and evolution has been explained within the context of hierarchical gravitational scenarios (Sheth & van de Weygaert 2004).

Perhaps the most significant and characteristic property is the hierarchical nature of the cosmic matter distribution. Over a wide range of spatial and mass scales objects and/or structures are embedded within structures of a larger dimension and a lower density. This is a manifestation of clustering in gravitational instability scenarios in which the power spectrum of primordial density fluctuations is a decreasing function of the spatial scale. The currently most viable candidate for the gravitationally dominant constituent of the cosmic inventory is a species of cold dark matter, whose fluctuations would indeed involve such a hierarchical spectrum. It leads to a scenario in which the first objects to form are small compact objects which subsequently merge with their surroundings as the larger scale density excess in which they are embedded condense out of the cosmic background (Press & Schechter 1974, Bond et al. 1991). As a result extended structures which may still be in the process of collapsing, or have not yet reached a perfect stage of virialization, often contain a large amount of smaller scale structures of higher density which have collapsed at an earlier epoch. As a result we find a large variety of structures and features in the universe, with a wide range of densities and each marked by different spatial and mass scales. The presence of a hierarchy of embedded

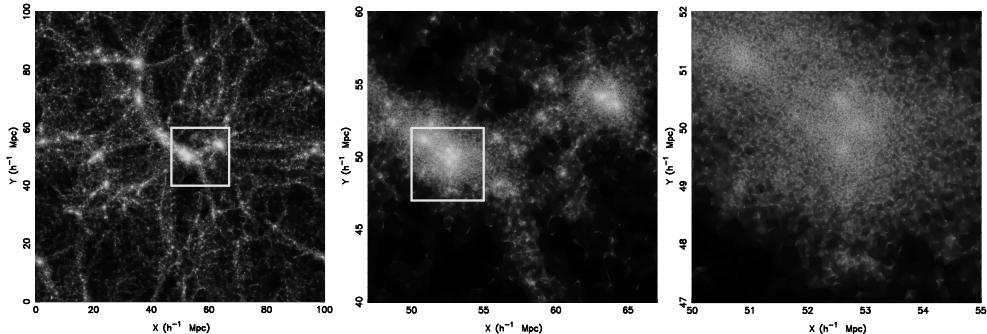


Figure 4.1 — Cosmic density field illustrating the large dynamic range which is present in the large scale matter distribution. In the left-hand frame the density field in a $10h^{-1}\text{Mpc}$ wide slice through a cosmological simulation is depicted. In the subsequent frames zoom-ins focusing on a particular structure are shown. On all depicted scales structures are present.

substructures on different scales is illustrated in Fig. 4.1. It shows a slice through a simulation of cosmological structure formation, together with a number of zoom-ins which focus in on a particular substructure. In the central zoom-in small clumps in the filamentary structure connecting the two main structures are clearly visible. The bottom frame zooms in on a huge concentration of matter which can be seen to display detailed substructure including at least three subclumps at its core. Note that the density fields were obtained by means of the DTDE procedure outlined in this study.

Observationally, we recognize traces of the hierarchical formation process in the galaxy distribution on Megaparsec scales. The large unrelaxed filamentary and wall-like superclusters contain various rich clusters of galaxies as well as a plethora of smaller galaxy groups, each of which has a higher density than the average supercluster density. Zooming in on even smaller scales, within groups large galaxies themselves are usually accompanied by a number of smaller satellites and dwarf galaxies. The imprint of hierarchical clustering may also be found in fully collapsed structures, such as clusters and even haloes of galaxies. When studying the hot X-ray emitting intracluster gas, more evenly distributed than the galaxies, the majority of clusters appears to display some measure of substructure (e.g. Schuecker et al. 2001). Even the Coma cluster appears to be marked by a heavy infalling group (Neumann et al. 2003). Also galaxies bear the marks of their hierarchical formation. The most visible manifestation concerns the presence of streams in their dark haloes, remnants of infalling dwarf galaxies (e.g. Helmi et al. 1999, Freeman & Bland-Hawthorn 2002).

4.1.2 Analyzing the cosmic web

Cosmological theories describe the development of structure in terms of continuous (dark matter) density and velocity fields. However, we have seen that to a large extent our knowledge stems from a discrete sampling of these fields. In the observational reality galaxies are the main tracers of the cosmic web and it is mainly through measuring of the redshift distribution of galaxies that we have been able to map its structure. Likewise, simulations of the evolving cosmic matter distribution are almost exclusively based on N -body particle computer calculations, involving a discrete representation of the features we seek to study. Both

the galaxy distribution as well as the particles in an N -body simulation are examples of spatial point processes in that they are

- discretely sampled;
- have an irregular spatial distribution.

The translation of discretely sampled and spatially irregularly distributed sampled objects into related continuous fields or sufficiently descriptive cosmological measures is not necessarily a trivial procedure. It involves issues of smoothing and spatial interpolation of the measured data over the sampling volume, which is a subject of considerable importance and interest in many different branches of science. Key references on the involved problems and solutions include those by Ripley (1981, 1989), Watson (1992) and Cressie (1993). While of considerable importance for astronomical purposes, many available methods escaped attention. A systematic treatment and discussion within the astronomical context is the study by Rybicki & Press (1992), who focussed on linear systems as they developed various statistical procedures related to linear prediction and optimal filtering, commonly known as Wiener filtering. An extensive, systematic and more general survey of available mathematical methods can be found in a set of publications by Lombardi & Schneider (2001, 2002 and 2003).

The main issue is the processing of the discretely sampled field into a continuous field which retains all required information. While demanding in itself, it is complicated by the highly inhomogeneous nature of the sampling point distribution. The number density of galaxies in redshift survey maps and N -body particles in computer simulations is supposed to be proportional to the underlying matter density. Because this involves a hierarchical and anisotropic matter distribution ideally one should define a technique which would not lose information against the backdrop of a highly inhomogeneous spatial resolution. As a result nearly all existing techniques for analyzing galaxy redshift surveys or numerical simulations of cosmic structure formation have important shortcomings with respect to how they treat the web-like geometry of the large scale matter distribution and the hierarchical character of the cosmic matter distribution. The limited available mathematical machinery has often been a major obstacle in exploiting the potentially large information content of the cosmic web.

4.1.2.1 Data compression: statistical measures

One strategy is to distill various statistical measures characterizing particular aspects of the large scale matter distribution (see Martínez & Saar 2002 for an extensive review). The clustering of galaxies and matter is most commonly described in terms of a hierarchy of correlation functions (Peebles 1980). The two-point correlation function – and the equivalent power spectrum, its Fourier transform (Peacock & Dodds 1994, Tegmark et al. 2004) – remains the mainstay of cosmological clustering analysis and has a solid physical basis. However, the non-trivial and non-linear patterns of the cosmic web are mostly a result of the phase correlations in the cosmic matter distribution (Ryden & Gramann 1991, Chiang & Coles 2000). While this information is contained in the moments of cell counts (Peebles 1980, de Lapparent, Geller & Huchra 1991, Gaztañaga 1992) and, more formally so, in the full hierarchy of M -point correlation functions ξ_M , except for the lowest orders their measurement has proven to be practically unfeasible (Peebles 1980, Szapudi & Szalay 1998, Martínez & Saar 2002). Nor

do these higher-order correlation functions readily translate into a characterization of identifiable features in the cosmic web. Many other attempts to describe and identify web-like features have been of a mainly heuristic nature. In particular the detection of elongated filamentary features has received some attention. The connectedness of elongated supercluster structures in the cosmic matter distribution was first probed by means of percolation analysis, introduced and emphasized by Zel'dovich and coworkers (Zel'dovich 1982), while a related graph-theoretical construct, the minimum spanning tree of the galaxy distribution, was extensively probed and analysed by Bhavsar and collaborators (Barrow, Bhavsar & Sonoda 1985, Graham, Clowes & Campusano 1995) in an attempt to develop an objective measure of filamentarity. Another heuristic test for tracing filaments is based on the correlation in the orientation of galaxy position angles (Pimbblet 2005), while Stoica et al. (2005) recently described a mathematically more rigorous and solid technique for filament finding, the Candy model. Topological characteristics of the cosmic web have been the subject of a few studies concentrating on the shape of the local matter distribution. A few studies determined the statistical properties of moments of inertia of the matter distribution, (Babul & Starkman 1992, Luo & Vishniac 1995, Basilakos, Plionis & Rowan-Robinson 2001), concepts which are closely affiliated to the full characterization of the topology of the matter distribution in terms of four Minkowski functionals (Mecke, Buchert & Wagner 1994, Schmalzing et al. 1999).

4.1.2.2 Field reconstruction

The strategy we follow is the translation of the full discrete inhomogeneous spatial point distribution into a continuous density field and/or any of other sampled fields such as the cosmic velocity field. Instead of characterizing the matter distribution in terms of a few statistical parameters we seek to define the local density field throughout the sampling volume. The processed field forms the basis for any further post-processing analysis. This may involve the determination of statistical measures (see above) but may also involve visualization procedures or procedures for identifying features such as voids, filaments, clusters and/or dark haloes.

Ideally, the field reconstruction should retain all available information. Conventional techniques for the purpose of deriving a density field from a discrete set of galaxy or simulation particle locations almost without exception involve some kind of user-defined filtering of the discrete data. This may involve a fitting procedure and/or a likelihood-based interpretation of the data within the context of a preconceived model of the cosmic matter distribution, incorporating a set of assumptions on the cosmological model and structure formation scenario. A particularly successful and well-defined example is that of Wiener filter-based reconstructions (see e.g. Rybicki & Press 1992, Erdoğdu et al. 2004, Zaroubi et al. 1995). Even outside the context of a user-defined (cosmological) model, conventional techniques for the purpose of deriving a density field from a set of galaxy or simulation particle locations without exception involve filtering and/or smoothing operations. Nearly always this is based on artificial smoothing/filtering kernels whose functional form and parameter values need to be specified by the user. This leads to a degree of arbitrariness and subjectivity which bears strongly on the derived results and reflects at least to some extent the prejudices of the designer. Due to their often rigid geometry or characteristic scale many known filtering kernels do tend to erase substructure on a scale smaller than the filter radius, diminish the flattened or elongated morphology of the spatial patterns, introduce artificial topological features in sparsely sampled

regions and amplify the contrast of particular patterns or objects. Even while sophisticated multi-scale wavelet techniques are being recruited towards the analysis of the cosmic web and do manage to improve the ability to resolve substructure, such methods do nonetheless tend to tune the result via the definition of the wavelet base (Martínez 2006, Jones 2006).

4.1.3 The Delaunay Tessellation Field Estimator

In this work we analyze the Delaunay Tessellation Field Estimator (DTFE). This method is based on a basic concept from the field of stochastic and computational geometry, the Delaunay tessellation of the point sample (see Okabe et al. 2000 and references therein). This spatial volume-covering division of space into mutually disjunct triangular (in two dimensions) or tetrahedral (in three dimensions) cells adapts to the local density and geometry of the point distribution. The DTFE method exploits this virtue and thus adapts automatically and in an entirely natural fashion to changes in the density and the geometry of the distribution of sampling points. The Delaunay tetrahedra are used both to obtain optimal local estimates of the spatial density and as multi-dimensional intervals for linear interpolation of the field values sampled or estimated at the location of the sampling points (see Okabe et al. 2000).

The DTFE involves an extension of the interpolation procedure described by Bernardreau & van de Weygaert (1996), who used Delaunay tessellations for the specific purpose of estimating the cosmic velocity divergence field, and showed the method's superior performance with respect to conventional interpolation procedures. They also proved that the obtained field estimates involved those of the proper volume-weighted quantities, instead of the usually implicit mass-weighted quantities, which corrected a few fundamental biases in estimates of higher-order velocity field moments. Subsequently, the DTFE procedure was developed by for the general purpose of rendering fully volume-covering and volume-weighted fields from a discrete set of sampled field values, including the density field as defined by the point sampling itself (see the previous chapters of this thesis). Instead of involving user-defined filters which are based on artificial smoothing kernels the resulting main virtue of the DTFE is that it is intrinsically self-adaptive. In essence it involves filtering kernels which are defined by the local density and geometry of the point process or object distribution. On the basis of its interpolation characteristics the DTFE is the first-order version of a wider class of tessellation-based multi-dimensional and entirely local interpolation procedures, commonly known as natural neighbour interpolation (Watson 1992, Sambridge, Braun & McQueen 1996, Sukumar 1998, Okabe 2000).

An essential step of the DTFE procedure concerns the determination of field values at the sampling points. Assuming that these are sampled proportionally to an underlying density field, the Delaunay tessellation is used to define optimal density estimates. Tessellation-based methods for estimating the density have been introduced by Brown et al. (1965) and Ord (1978). In astronomy Ebeling & Wiedenmann (1993) were the first to use tessellation-based density estimators for the specific purpose of devising source detection algorithms. This work has recently been applied to cluster detection algorithms by Ramella et al. (2001), Kim et al. (2002) and Marinoni et al. (2002). Along the same lines, Ascasibar & Binney (2005) suggested that the use of a multi-dimensional binary tree might offer a computationally more efficient alternative. However, these studies have been restricted to raw estimates of the local sampling density at the position of the sampling points and have not yet included the more elaborate interpolation machinery of the DTFE and natural neighbour interpolation methods.

The DTFE is described in detail in the previous chapters of this thesis, while a short description is presented in appendix 4.A.3.

4.1.4 The DTFE and cosmic patterns

A proper and profound understanding and quantification of the cosmic web is not possible without tools which are capable of tracing hierarchically structured and anisotropic spatial patterns in an entirely objective fashion. The various aspects characterizing the complex and non-trivial spatial structure of the cosmic web have proven to be notoriously difficult to quantify and describe. For the analysis of web-like patterns the toolbox of descriptive measures is still largely ad-hoc and is usually biased towards preconceived notions of their morphology and scale. None of the conventional, nor even specifically designed measures of the spatial matter distribution have succeeded in describing all relevant features of the cosmic web. Even while they may succeed in quantifying one particular key aspect, it usually excludes the ability to do so for other characteristics. The DTFE, on the other hand, is able to follow and accurately describe the patterns in the matter distribution in an entirely natural and automatic fashion. Its ability to reproduce at least three key characteristics of the web-like cosmic matter distribution makes it a potentially highly promising tool for studies of the large scale matter distribution.

While DTFE reconstructed (density) fields are optimal in the sense of defining a continuous and unbiased representation of the data while retaining all information available in the point sample, the method finds its fulfilment and purpose in forming the basis for a variety of analysis and processing tools seeking to describe and dissect the spatial (sub)structure and geometry of the cosmic matter distribution. A potentially interesting application would be its implementation in the SURFGEN machinery, which seeks to provide locally defined topological measures, *cq.* local Minkowski functionals, of the density field (Sahni, Sathyaprakash & Shandarin 1998, Shandarin, Sheth & Sahni 2004). A recent sophisticated technique for tracing the cosmic web is the skeleton formalism developed by Novikov, Colombi & Doré (2006), based on the framework of Morse theory (Colombi, Pogosyan & Souradeep 2000). The skeleton formalism seeks to identify filaments in the web by identifying ridges along which the gradient of the density field is extremal along its iso-contours (see also Sousbie et al. 2006). Such an analysis may profit from the unbiased web-like density fields defined by the DTFE. In fact, two major extensions of the DTFE already set out to the identification of key aspects of the cosmic web. The watershed algorithm of Platen & van de Weygaert (2006) is a void detection algorithm set to outline the hierarchical nature of the cosmic void population. The detection of web-like anisotropic patterns over a range of spatial scales, such as we may expect for a hierarchical matter distribution, is the purpose of the multiscale morphology filter (MMF), introduced by Aragón-Calvo et al. (2006).

In the meantime the DTFE has been applied in a number of studies of cosmic structure formation. These studies do indeed suggest a major improvement over the more conventional analysis tools. Evidently, even though density/intensity field analysis is one of the primary objectives of the DTFE formalism, one of its important features is its ability to extend its Delaunay tessellation-based spatial interpolation to any corresponding spatially sampled physical quantity. The most directly related example is that of the analysis of the cosmic velocity field, intimately coupled to that of the corresponding density field (Bernardeau & van de Weygaert 1996, Schaap & van de Weygaert 2003, Romano-Díaz 2004, Chapter 6 of this thesis).

Extrapolating this observation, and encouraged by the success of Voronoi-based methods in identifying dark haloes in N -body simulations (Neyrinck, Hamilton & Gnedin 2005), Arad, Dekel & Klypin (2004) used the DTFE to assess the six-dimensional phase-space density distribution of dark haloes in cosmological N -body simulations. While a fully six-dimensional analysis may be computationally cumbersome (Ascasibar & Binney 2005), the splitting of the problem into separate spatial and velocity-space three-dimensional tessellations may indeed hold promise for an innovative analysis of the dynamical evolution of dark haloes. The ability of the DTFE to trace sharp density contrasts impelled Bradac et al. (2004) to compute the surface density map for galaxies from the projection of the DTFE reconstructed three-dimensional density field. Bradac et al. (2004) used the obtained surface density map to compute the gravitational lensing pattern around the object, upon which Li et al. (2006) evaluated the method in its ability to trace higher-order singularities.

4.1.5 Plan and organization

The scope of the work presented in this chapter is to quantify and test the performance of the DTFE with respect to its ability to resolve hierarchical density distributions, to identify substructure and to recover, quantify and analyze anisotropic spatial patterns. For comparison, we process the same data sets with two other methods: a rigid grid-based method, TSC, which has no adaptive characteristics in terms of scale nor shape and an SPH-like smoothing algorithm which does have a kernel which adapts to scale but remains insensitive to shape.

The performance of the DTFE in dissecting hierarchically embedded density structures is tested on the basis of three different Soneira-Peebles model realizations. The self-similar scaling of this fractal-like and analytically tractable model forms a very useful template in that we can test the TSC, SPH and DTFE methods in their ability to recover the underlying parameters or scaling indices. For testing the ability of the DTFE reconstruction procedure to trace and recover anisotropic web-like features we invoke a set of Voronoi clustering model realizations (van de Weygaert 1991). These models use Voronoi tessellations as skeleton of the point/galaxy distribution. The used model templates are a regular Voronoi wall model, a Voronoi filament model and a Voronoi cluster model, each defining a point distribution around the walls, edges or vertices of a Voronoi tessellation.

The chapter is organized as follows. In the first three sections the ability of TSC, SPH and DTFE to recover hierarchically embedded substructure is assessed. Section 4.2 describes the set of Soneira-Peebles models which we have in our study. A qualitative discussion of the density field reconstructions of the Soneira-Peebles model realizations follows in section 4.3. The quantitative analysis of these reconstructions, in section 4.4, focusses in particular on the scaling characteristics. Sections 4.5, 4.7 and 4.8 focus on the ability of the TSC, SPH and DTFE procedures to produce density fields which reflect the morphological and topological characteristics of web-like anisotropic spatial patterns. Section 4.5 introduces the Voronoi clustering models which we will use as templates for web-like galaxy distributions. After the presentation of the TSC, SPH and DTFE density field reconstructions and a discussion of their qualitative aspects in section 4.6, section 4.7 will involve a quantitative shape and morphological analysis of the recovered density field while section 4.8 focusses on one particular topological aspect, the volume occupation. The final section, sect. 4.9, contains a summary and discussion of our results and of the implication for the application of the DTFE to more complex and intricate spatial patterns and point distributions.

4.2 Substructure and structural hierarchy: the Soneira-Peebles model

One of the key properties of the large scale galaxy distribution is the presence of a hierarchy of embedded substructures with a large dynamic range in terms of densities and spatial scales (see Fig. 4.1). As a result of the hierarchical progression of structure formation we currently recognize a large range and variety of different structures in the spatial galaxy distribution. These are marked by different spatial and mass scales and a wide spectrum of densities. To analyze, understand and properly interpret such matter distributions, we do need tools that are able to identify and properly quantify the full range of features and densities which we encounter in the cosmic matter distribution.

Conventional filters, based on user-defined kernels, are bound to fail in the identification of substructures smaller than the (local) kernel function. This may be so even while such condensations are present within the galaxy or particle distribution. Dependent on their sensitivity to the shape of objects they may also bias their selection towards a particular range of features while discarding others. On the basis of its definition the DTFE is expected to be optimally sensitive to any significant local density enhancement as well as any local geometrical anomaly.

In the following sections we test its corresponding characteristics on the basis of its performance in the case of an analytically clearly defined and fully understood ‘fractal-like’ point distribution, the class of Soneira-Peebles models (Soneira & Peebles 1978). In order to evaluate the virtues of the DTFE (appendix 4.A.3) we confront its performance with that of two different representative smoothing-interpolation schemes. The TSC procedure (appendix 4.A.1) is rigid with respect to its scales sensitivity as well as its shape sensitivity, while the SPH procedure does involve a strong scale dependence but lacks shape sensitivity (appendix 4.A.2).

4.2.1 The Soneira-Peebles model

In the late seventies Soneira and Peebles (1977, 1978) noticed that the observed angular galaxy distribution on the sky displays self-similar behavior. They succeeded in reproducing the main clustering statistics of this distribution with a simple analytic model with a few adjustable parameters, which determine the complexity and dynamic range of the resulting point distribution. Because these properties may be varied in a predictable fashion, we have used the Soneira-Peebles model to test the performance of the DTFE with respect to point distributions with a similar dynamic range as the observed galaxy distribution on the sky.

In essence the Soneira-Peebles model consists of a recipe to distribute a given number of points. The starting point is a level-0 sphere of radius R . In this sphere η level-1 spheres are placed with radius R/λ and $\lambda > 1$. The new spheres are placed at a random position inside the level-0 circle, such that their centers fall inside the original level-0 sphere. Within each of these η level-1 spheres, one places η level-2 spheres of radius R/λ^2 . This process is repeated until one ends up with in total η^L level- L spheres of radius R/λ^L . At the center of each of these level- L spheres a point is placed. One therefore ends up with in total η^L points, which in the Soneira-Peebles model represent galaxies. We will refer to this procedure up to this point as the *singular Soneira-Peebles model*. This procedure is illustrated in Fig. 4.2.

The Soneira-Peebles model is controlled through three parameters, η , L and λ . The effect of varying these parameters on the resulting point distribution is illustrated in Fig. 4.3. For

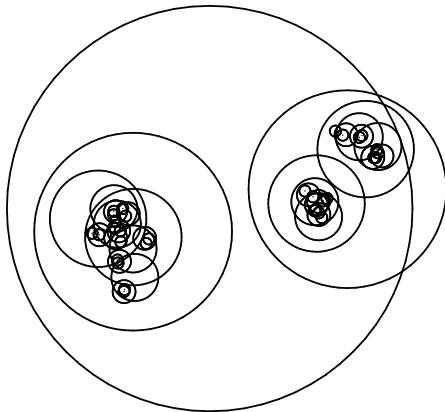


Figure 4.2 — The Soneira-Peebles model. Inside a level-0 sphere η level-1 circles are placed with a radius which is smaller by a fixed factor. This process is repeated until one ends up with η^L level- L circles. At the center of these level- L circles η^L points are placed, which form the resulting Soneira-Peebles point distribution.

a given number of points, η determines the dynamic range of the resulting point distribution. For a small value of η , many levels are needed to reach a fixed number of points, while a large value of η results in a smaller number of levels. A small value of η also results in a smaller filling fraction of space with spheres than a high value of η (top row in Fig. 4.3). L denotes the total number of levels and therefore determines the range of densities and scales in the resulting point distribution. For a fixed value of η , L also determines the total number of points (central row in Fig. 4.3). Finally, for given values of η and L , λ determines the range of spatial scales. A value of λ close to 1 means that subsequent spheres of higher levels are of comparable size. Values of λ much larger than one mean that each subsequent level consists of spheres which are significantly smaller than the spheres in the preceding level (bottom row in Fig. 4.3).

An important property of the Soneira-Peebles model is that it is one of the few analytic self-similar models of the galaxy distribution for which the two-point correlation function can be analytically evaluated. In M dimensions it is given by

$$\xi(r) \sim r^{-\gamma}, \quad (4.1)$$

with

$$\gamma = M - \left(\frac{\log \eta}{\log \lambda} \right) \quad \text{for} \quad \frac{R}{\lambda^{L-1}} < r < R. \quad (4.2)$$

The parameters η and λ may be chosen such that the two-point correlation function of the resulting point distribution matches the two-point correlation function of the observed galaxy distribution.

From Fig. 4.2 it may be appreciated that the Soneira-Peebles model involves a hierarchy of structures of varying densities and characteristic scales, with the higher level spheres corresponding to high density structures of small scale and the lower level spheres corresponding to low density structures of larger scale. As each of the spheres is constructed in the same way, the resulting point distribution is self-similar and forms a bounded fractal. The fractal geometry of a point set is often characterized by the fractal dimension D , which is defined as

$$D = \lim_{r \rightarrow 0} \frac{\log N(r)}{\log (1/r)}. \quad (4.3)$$

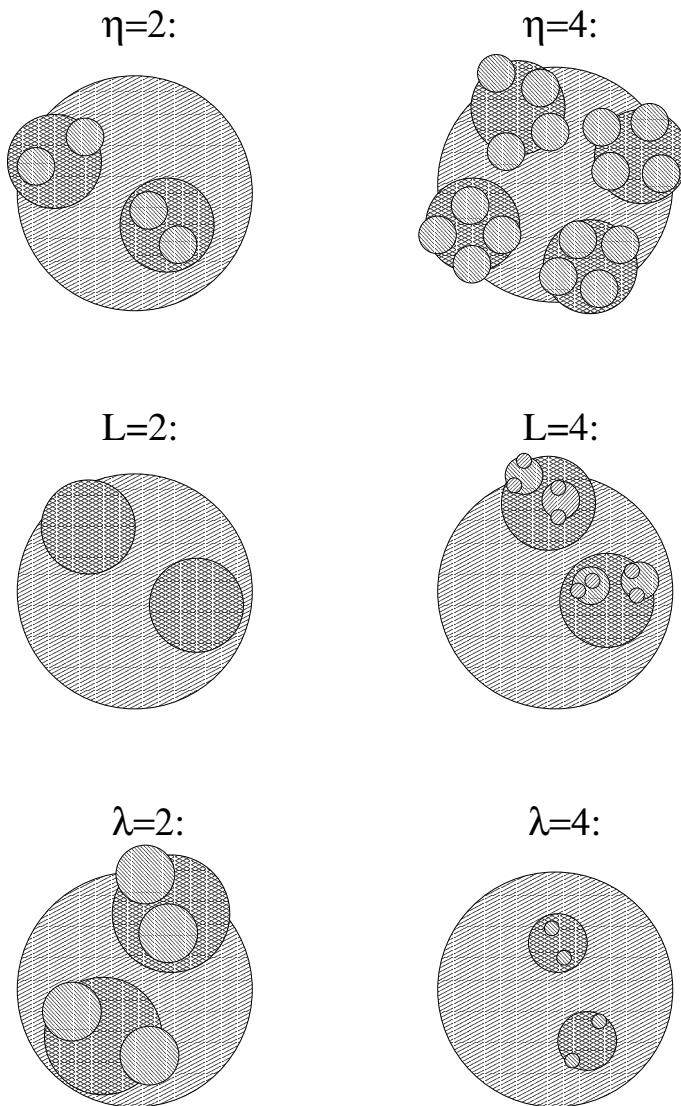


Figure 4.3 — The physical meaning of the three defining parameters η , L and λ of the Soneira-Peebles model. The upper row shows the effect of varying η , the number of circles which is placed in each circle. The central row shows the effect of varying L , the total number of levels. The bottom row shows the effect of varying λ , the ratio of the radius of each circle with the radius of subsequent circles of one level higher.

Here $N(r)$ is the number of non-empty cells in a partition of constant cell size r . If the Soneira-Peebles model would contain an infinite amount of levels, the resulting point distribution would have fractal dimension $D = (\log \eta)/(\log \lambda)$ (we refer the reader to Martínez et al. (1990) for an extensive description of the fractal-like properties of the Soneira-Peebles model).

The original Soneira-Peebles model, which here we will refer to as the *extended Soneira-Peebles model*, consists of the superposition of several singular Soneira-Peebles models with different values for η , λ and L . The ranges of values used for η , λ and L were chosen such that the resulting point distribution resembled the galaxy distribution in the Lick survey (Seldner et al. 1977) and simultaneously the observed two- and three-point correlation functions. Subsequently, absolute magnitudes were assigned to the points according to a particular luminosity distribution such that the apparent magnitude distribution of the projected point distribution resembled that of the projected apparent magnitude distribution of the galaxies in the Lick survey.

4.2.2 Realizations: singular Soneira-Peebles models

A number of different singular Soneira-Peebles model point distributions in two dimensions ($M = 2$) were constructed. As an illustration we describe two particular realizations. Realizations SPS1 is characterized by the parameter set $\eta = 4$, $\lambda = 1.9$ and $L = 8$. The total realization contains 65 536 points. The second set, SPS2, is specified by the parameters $\eta = 2$, $\lambda = 1.75$ and $L = 14$, yielding a distribution of in total 16 384 points. In both realizations the points have been placed in a box with periodic boundary conditions. The radius of the initial circle, in units of boxsize, is $R = 0.5$. In Table 4.1 we have listed some relevant quantities of these two realizations.

Table 4.1 — Overview of a number of relevant properties of the $\eta = 2$ and $\eta = 4$ singular Soneira-Peebles model realizations. R_L denotes the radius of the level- L circles. The filling factor is the fraction of space occupied by the level- L circles. $\langle d \rangle$ denotes the mean distance between nearest neighbors. The peak density is defined as the inverse of the area of the level- L circles.

	realization SPS1	realization SPS2
η	4	2
λ	1.90	1.75
L	8	14
#points	65 536	16 384
R_L	$2.9 \cdot 10^{-3}$	$2.0 \cdot 10^{-4}$
filling factor	0.41	$1.6 \cdot 10^{-3}$
$\langle d \rangle$	$1.2 \cdot 10^{-3}$	$2.5 \cdot 10^{-4}$
peak density	$3.7 \cdot 10^5$	$8.1 \cdot 10^7$
fractal dimension D	2.2	1.2

The resulting point distributions are shown in the top left-hand frames of Fig. 4.4 and 4.5. The subsequent frames in the top row zoom in on a particular smaller scale structure in the point distribution. The figures clearly show the differences between the realizations SPS1 and SPS2. The two models have a very different visual appearance. This can be mainly

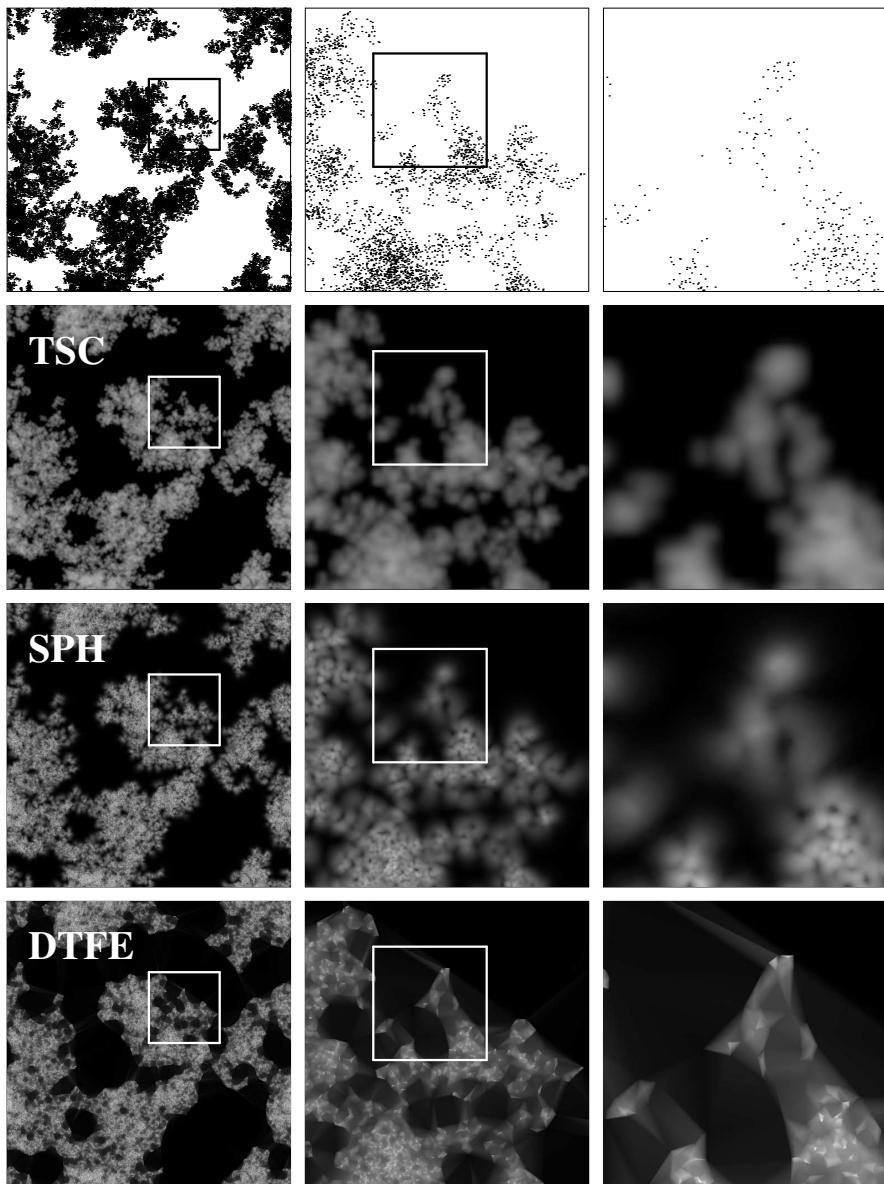


Figure 4.4 — Singular Soneira-Peebles model with $\eta = 4$, $\lambda = 1.9$ and $L = 8$. Top row: full point distribution (left-hand frame) and zoom-ins focusing on a particular structure (central and right-hand frames). Rows 2 to 4: corresponding density field reconstructions produced using the TSC, SPH and DTFE methods.

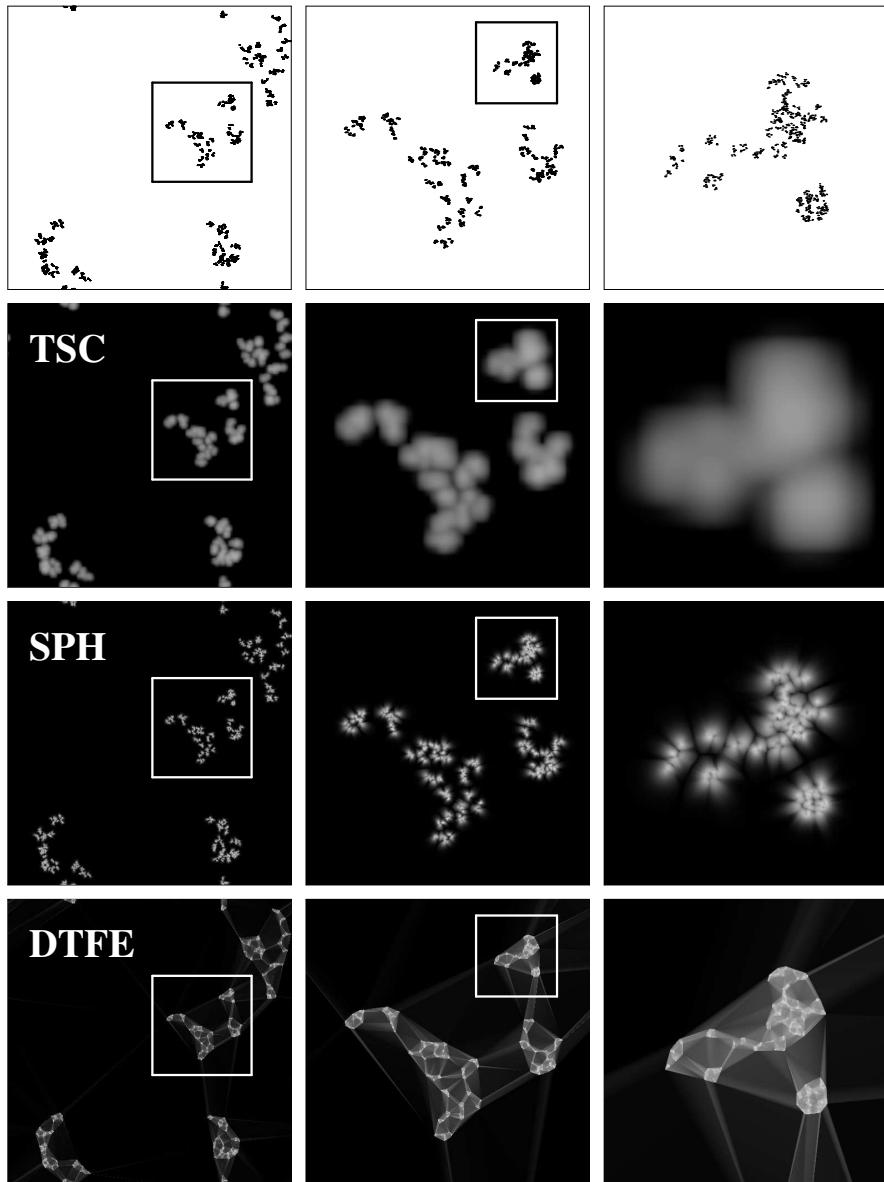


Figure 4.5 — Singular Soneira-Peebles model with $\eta = 2$, $\lambda = 1.75$ and $L = 14$. Top row: full point distribution (left-hand frame) and zoom-ins focusing on a particular structure (central and right-hand frames). Rows 2 to 4: corresponding density field reconstructions produced using the TSC, SPH and DTFE methods.

understood in terms of their different filling factors, the fraction of space occupied by the highest level circles. For the SPS1 $\eta = 4$ model this is more than two orders of magnitude larger than that of the SPS2 $\eta = 2$ model (see Table 4.1).

The two realizations also have a very different fractal dimension, $D = 2.2$ for the SPS1 $\eta = 4$ model and $D = 1.2$ for the SPS2 $\eta = 2$ model. It is indicative of the more extreme fractal distribution represented by the latter. Even though this distribution contains only a fourth of the number of points in the $\eta = 4$ SPS1 realization (see Table 4.1) it contains structures of a much larger dynamic range, both in density and in size, than the SPS1 model. The highest level circle has radius $R_L = 2.0 \cdot 10^{-4}$, resulting in a peak density (defined as the inverse area of this circle) of $8.1 \cdot 10^7$ points per volume unit, more than two orders of magnitude larger than in the $\eta = 4$ model. As an example of this in the third column we zoom in on a patch in the SPS2 $\eta = 2$ model with many more points than the equivalent region in the SPS $\eta = 4$ model (1024 vs. 297). It is a result of the clustering of points in such a way that prominent structures and features can be discerned over a larger range of scales and levels. The contrast in the spatial point distribution of the SPS1 $\eta = 4$ realization is clearly of a considerably lower level. On the other hand, the spatial distribution of the latter is characterized by a larger range in scale from level to level due to the higher value of λ .

4.2.3 Realizations: extended Soneira-Peebles models

In the extended Soneira-Peebles model several singular Soneira-Peebles model point distributions are superposed on top of each other. Soneira and Peebles (1977, 1978) have shown that for certain choices of the parameters η , λ and L for the composing singular Soneira-Peebles models both the angular galaxy distribution as observed in the Lick survey and the observed angular two- and three-point correlation functions may be reproduced.

Superposing various different singular Soneira-Peebles realizations, each with a different self-similar behavior and dynamic range, produces a highly complex spatial point distribution containing structures and features over a large range of scales and with highly varying densities. We have constructed a point distribution SPE1 following this recipe. We have added 30 singular Soneira-Peebles realizations with $\eta = 2$ and $\lambda = 1.75$ and with L varying between 7 and 16. The resulting point distribution consists of 111 936 points and is shown in the top left-hand frame of Fig. 4.6.

4.3 Soneira Peebles model: density field reconstructions

We have reconstructed the density fields which correspond to the $\eta = 4$ and $\eta = 2$ singular Soneira Peebles model point distributions shown in the top column of Figs. 4.4 and 4.5, as well as to the extended Soneira-Peebles model point distribution shown in the top column of Fig. 4.6 by means of a fixed, grid-based TSC reconstruction procedure, an adaptive SPH-like reconstruction prescription and the DTFE procedure. Greyscale images of the corresponding density field reconstructions are also shown in Figs. 4.4, 4.5 and 4.6: the 2nd column depicts the TSC density field reconstruction, the 3rd column the SPH density field reconstruction and the 4th column the DTFE field reconstruction. In each case the left-hand column corresponds to the density distribution in the full box, while the next two columns depict the density field in successive zoom-ins. These have been selected according to whether they contain interesting features.

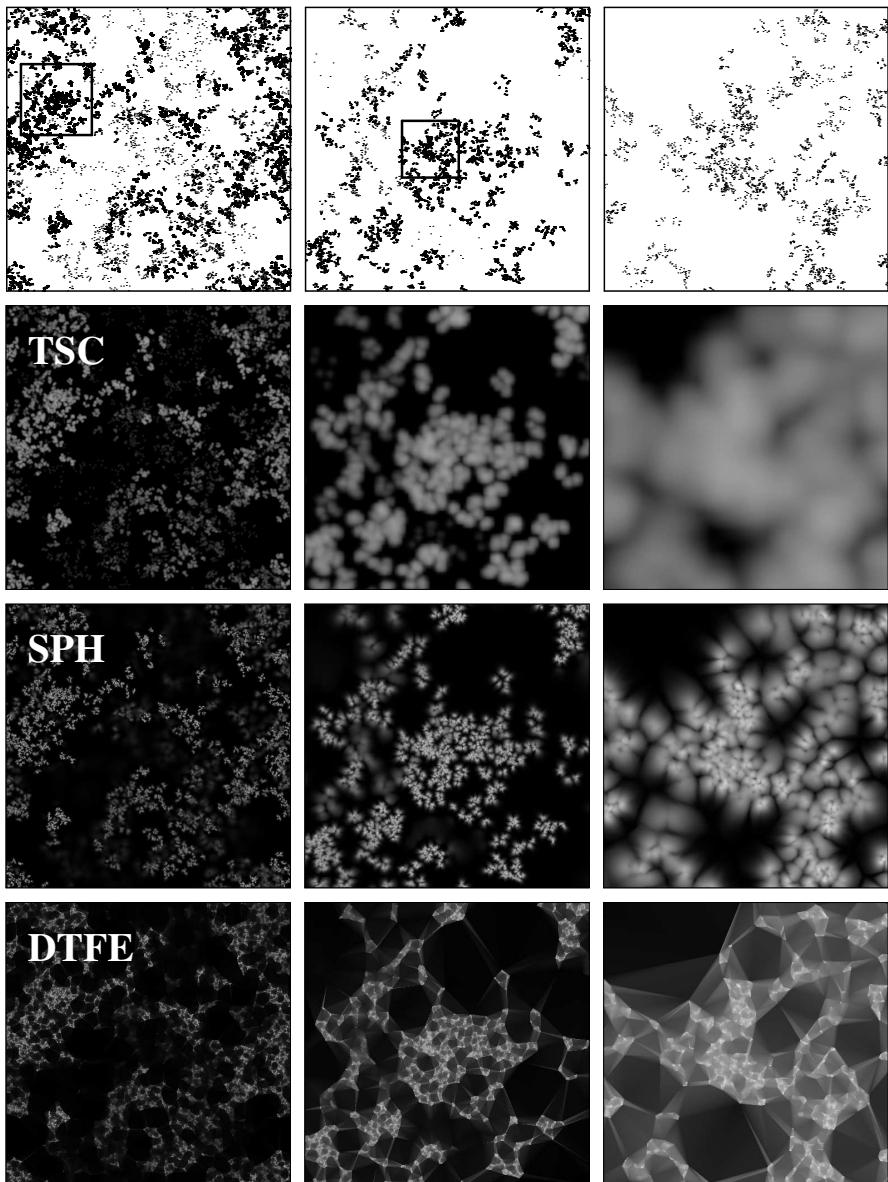


Figure 4.6 — Extended Soneira-Peebles model. Top row: full point distribution (left-hand frame) and zoom-ins focusing on a feature with interesting substructure (central and right-hand frame). Rows 2 to 4: corresponding density field reconstructions produced using the TSC, SPH and DTFE methods.

4.3.1 The $\eta = 4$ singular model

First consider the less extreme case of the $\eta = 4$ SPS1 model, shown in Fig. 4.4. The highest level circle has radius $R_L = 2.9 \cdot 10^{-3}$, resulting in a peak density (defined as the inverse area of this circle) of $3.7 \cdot 10^5$ points per volume unit, more than two orders of magnitude lower than in the $\eta = 2$ SPS2 model. When inspecting the complete box (left-hand column) the reconstructions of the different methods appear to be qualitatively similar. The TSC reconstruction seems somewhat more blurry than the SPH and DTFE reconstructions: the TSC density field shows less contrast between low and high density structures. Also, structures appear to be somewhat more extended. On this scale the difference between the SPH and DTFE reconstructions is very small. The only noticeable difference is that the DTFE reconstruction have a somewhat crispier appearance.

The differences between the various methods become more apparent when zooming in on particular patches in the density field reconstructions, as can be clearly appreciated from the central and right-hand columns of Fig. 4.4. Going from TSC to SPH to DTFE structures are better resolved and are characterized by a higher contrast. Structures in the TSC density field reconstructions are more blurry than their counterparts in the SPH reconstructions, while the latter again are somewhat blurrier than the corresponding features in the DTFE reconstructions.

These differences in visual appearance can be directly related to the effective smoothing kernels corresponding to the reconstruction methods (see Chapter 3). The TSC procedure is based on the use of a fixed and rigid grid. The method is therefore unable to trace local variations in the number density of the point distribution on scales smaller than the grid scale. This means that structures smaller than the size of a grid cell are unresolved and smeared out over the extent of at least one grid cell. The SPH kernel adjusts itself to variations in the sampling density, with a kernel size defined such that it always contains a certain amount of sample points. This number is usually in the range of a few dozen points. However, the SPH kernel is less sensitive to the configuration of the local point distribution than the DTFE kernel. Its shape is user-defined and fixed, in most cases spherical, and does not adjust to the local point distribution. As a result the SPH kernel perform less than optimal when it encounters typical anisotropic features or sharp transition regions between distinct morphological features. The effective DTFE kernel not only adjusts itself to the local number density of sample points, with on average the least number of neighbour points, but also to the local variations in the geometry of the point distributions: it manages to trace elongated filamentary features, flattened features, large low density regions or extremely compact high density features automatically and without the need to specify a priori their shape.

In all reconstructions some artefacts related to the characteristic geometry of the effective smoothing kernels are visible. In the case of the TSC procedure the rigid geometry of the TSC grid produces maxima which lie at the centers of grid cells. The SPH kernel, on the other hand, produces artefacts at the transition regions between different structures as it tends to preferentially sample from only one of the surrounding structures, usually the densest one, and subsequently interpolates according to the user-defined kernel shape instead of one defined by the local geometry. As a result it produces artificial wings which fall off smoothly towards the background, rendering a correction for the artificial wings very difficult. Also the DTFE reconstructions contain artificial wings at the outskirts of structures. These wings have a distinct triangular shape. They are particularly prominent at sites where the density drops

very quickly to (almost) zero, yielding Delaunay triangles which suddenly become very extended. While these artificial wings are usually much more extended than the ones in the SPH reconstructions, be it that the corresponding density levels have a much lower amplitude than the SPH reconstructions. It is therefore much easier for the DTFE to separate the wings from genuine structures.

4.3.2 The $\eta = 2$ singular model

Fig. 4.5 presents the resulting density reconstructions for the more extreme point distribution of the $\eta = 2$ singular Soneira-Peebles model SPS2. The reconstructions all clearly reveal the density and size characteristics of the $\eta = 2$ SPS2 model. The differences between the different reconstructions are more pronounced than in the case of the $\eta = 4$ SPS1 reconstructions. Evidently, the TSC method is incapable of reproducing the small scale structures in the density field. This is particularly evident from the right-hand panel of Fig. 4.5. This finest zoom-in onto the $\eta = 2$ conglomerate contains a large number of points distributed over a range of substructures. TSC produces three featureless clouds over which the corresponding sampling points have been distributed. The three clouds occupy a much larger area than that of the original point clouds. More interesting are the differences between the SPH and DTFE reconstructions. Their typical artefacts are clearly visible in the depicted frames in the second and third row of Fig. 4.5. In the SPH reconstruction groups in the point distribution smear out gently into their surroundings, clearly showing the imprint of the (artificial) spherical filter. The DTFE reconstruction, on the other hand, can be seen to connect distinct groups of points with artificial low-level wings. This is a result of its inability to deal with regions of zero density. Of particular relevance is the comparison between the SPH and DTFE reconstructions in the finest zoom-ins (right-hand column). The DTFE reconstruction is considerably more successful in reproducing the substructure present in the point distribution. Unlike the SPH reconstruction it manages to detect the small sub-clumps in the point distribution.

4.3.3 The extended Soneira-Peebles model

A visual comparison of the different reconstructed density fields shown in Fig. 4.6 provides the same impressions as that obtained from the figures of singular Soneira-Peebles models in figs 4.4 and 4.5. A major difference is that the extended Soneira Peebles model has a much more pronounced appearance than the singular Soneira Peebles models. This holds true for all three methods, the TSC, SPH and DTFE density field reconstructions.

While the SPH and DTFE density field reconstructions on the scale of the complete box (left-hand column), do appear to be relatively small we can already discern the more blurry nature of the TSC reconstruction. This impression is confirmed as we zoom in on the patch in the left-hand corner (central column of panels). The TSC procedure is unable to accurately describe the point distribution at scales smaller than the grid size. The rich and complex structures present in the point distribution on this scale have been smeared out into featureless blobs. The differences between the SPH and the DTFE reconstruction also start to become apparent. In the DTFE reconstruction the artificial low-level wings connecting the different structures begin to show up, while in the SPH reconstructions the smooth and broad wings at the outside of structures start to become visible. In the finest zoom (right-hand panels) the differences between the reconstruction methods start to dominate the appearance of the

images. The SPH method smears out the point distribution into circular blobs with no internal distribution. The larger features get smeared out into their surrounding environment as is most apparent through the artificial wings which smoothly fall off into the background. The DTFE method does manage to recover all structures visible in the point distribution. At this level, however, also the artificial low density wings become quite prominent and form clearly visible bridges between the main structures.

Although a visual comparison between the different reconstructions shows that there are strong differences between the reconstruction procedures, it is not obvious which reconstruction is the best representation of the Soneira-Peebles point distribution. What is clear that TSC does not provide a good representation as it fails to resolve small-scale structures. The differences between SPH and DTFE are more subtle and a more quantitative measure is needed to assess the performance of the different reconstruction procedures.

4.4 Soneira-Peebles model reconstructions: quantitative analysis

In addition to a visual inspection of the density field reconstructions produced by the different reconstruction procedures, a proper systematic and quantitative study is crucial. By means of a quantitative analysis of some statistical characteristics of the Soneira-Peebles model reconstructions we assess which reconstruction method results in density fields whose properties are closest to those of the original point distribution.

The first quantitative analysis of the obtained density field reconstructions concerns the density frequency distribution, ie. the one-point distribution function, the PDF. The fractality of the singular Soneira-Peebles models finds its expression in power-law PDFs which have a very characteristic scaling behaviour. On the basis of the parameters of the models we can predict the slope of the power-law PDF, or the scaling of the (higher order) correlation functions, which forms a magnificent test for the viability and quality of the density field reconstructions of the various methods.

The extended Soneira-Peebles model, on the other hand, is not strictly self-similar and in general analytical predictions of the PDF are not available. The quality of the reconstructions of the extended Soneira-Peebles models are therefore investigated in a different fashion, by means of a two-dimensional Kolmogorov test of the reconstruction in five different areas of the point distribution. In addition, we compare the autocorrelation function of the reconstructed density fields with that of the pure two-point correlation function of the point distribution.

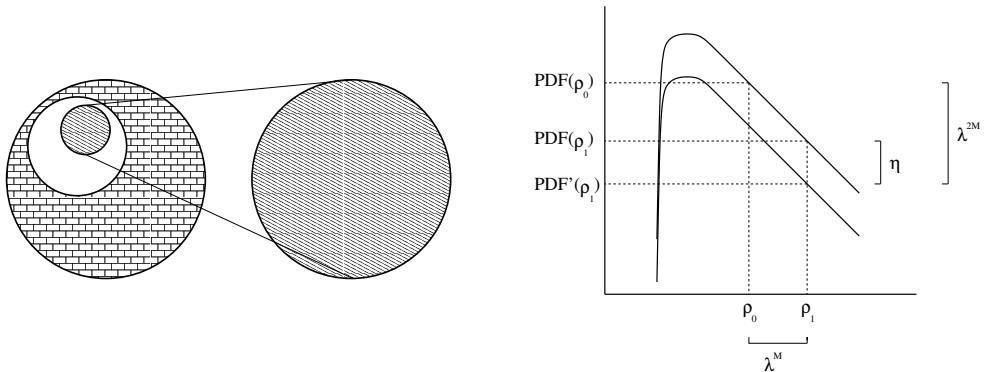
4.4.1 Singular models: one-point density distribution function

The one-point distribution function of the density field (PDF=probability distribution function) is the most direct and simple statistic of a density field. The PDF is the simple unconditional distribution function for the value of the density field ρ , defined as

$$\text{PDF}(\rho) = \frac{d}{d\rho} \frac{\int_S d\mathbf{x} \rho'(\mathbf{x})}{\int d\mathbf{x} \rho'(\mathbf{x})}, \quad (4.4)$$

where S denotes the region of space for which $\rho' < \rho$.

An essential property of the Soneira-Peebles model is its self-similarity. Concretely this means that if a circle of a certain level m is scaled to the size of a lower level n and therefore larger circle the resulting point distribution inside the magnified circle should have the same



(a) After scaling a higher level circle (small shaded circle on the left) to the size of a lower level circle (large circle with bricks), the properties of the resulting point distribution inside the magnified circle (large shaded circle on the right) should be statistically equal to those of the point distribution inside the lower level circle (large circle with bricks).

(b) Calculation of the slope of the power-law region of the PDF of a Soneira-Peebles density field. The lower curve denotes the contribution PDF' to the PDF of a circle of a given level by a circle of one level higher.

Figure 4.7 — The self-similar scaling characteristics of singular Soneira-Peebles models.

statistical properties as the point distribution inside an n -level circle (see Fig. 4.7). For a reconstruction procedure to be appropriate we will require that the same property holds for reconstructed density fields.

Mathematically, the condition of self-similarity implies that the PDF corresponding to a density field $\rho(\mathbf{x})$ inside an n -level circle of radius R/λ^n should be equal to the PDF inside the reference circle of radius R , after the density field in the n -level circle has been scaled according to

$$\rho(\mathbf{x}) \rightarrow \rho_n(\mathbf{x}) = \rho(\mathbf{x})/\lambda^{Mn}, \quad (4.5)$$

in which M is the dimension of space. The self-similarity of the resulting PDFs finds its expression in a power-law behaviour of the PDFs. The slope α of the PDF can be found by derived as follows.

Consider the PDF at the high density values. High density regions correspond to areas of high-level circles. Since there are η^n n -level circles, each n -level circle contributes a factor $\text{PDF}(\rho)/\eta^n$ to the PDF at density ρ (see Fig. 4.7). If this circle would be scaled to the size of a lower-level circle, self-similarity would imply that the resulting PDF is equal to the PDF of the lower-level circle. This scaling results in a density which has decreased by a factor λ^{Mn} , with M the dimension of space. Also, we have to scale the PDF with a factor λ^{Mn} to correct for the increase in area. Yet another multiplication factor of λ^{Mn} has to be included to properly normalize the PDF (per density unit). In total this results in an increase by a factor λ^{2Mn} .

We then find, as may be followed from Fig. 4.7, that the slope α is given by

$$\alpha = \frac{\log \text{PDF}(\rho_1) - \log \text{PDF}(\rho_0)}{\log \rho_1 - \log \rho_0} = \frac{\log(\lambda^{2Mn}/\eta^n)}{\log(1/\lambda^{Mn})} = \frac{D}{M} - 2, \quad (4.6)$$

in which D is the fractal dimension of the singular Soneira-Peebles mode. In turn, the value of the fractal dimension D of the Soneira-Peebles model is fully determined by the parameters that specify the model, η , L and λ .

We have calculated the PDF of the scaled density field inside higher level circles and compared it with the PDF of the density field inside the 0-level circle, the reference circle. At each level the density field was computed at a regular two-dimensional grid, within a circle of the appropriate size. Subsequently, the obtained values were scaled by applying Eqn. 4.5, followed by Eqn. 4.4. In order to ascertain independence of the adopted gridsize, we followed an iterative procedure in which we repeated the computation of the PDF until the results did not change anymore as the resolution got increased even further.

The PDF at a particular resolution level was obtained by averaging the PDFs determined within the various individual circles at that level. A slight complication occurs when circles of a particular level are overlapping or when circles partly lie outside their parent circle (ie. the containing circle one level higher). In such cases self-similarity breaks down. We have therefore restricted the Soneira-Peebles construction process to circles which do not overlap and which are completely located inside their parent circle.

4.4.2 Singular models: PDFs

The resulting average PDFs are plotted in Fig. 4.8 for the $\eta = 4$ and Fig. 4.9 for the $\eta = 2$ model. For the $\eta = 4$ model we have adopted $\lambda = 3$, which makes it less cumbersome to construct a non-overlapping distribution. The top left, top right and bottom left frames in Fig. 4.8 and Fig. 4.9 show the scaled density field PDFs for the TSC, SPH and DTFE density field reconstructions. In the lower right-hand frames we have plotted the ratio between the scaled PDFs of the DTFE and SPH density field reconstructions. In the three PDF frames the solid black lines depict the scaled density field PDF for a range of levels. Whether it concerns the TSC, the SPH or the DTFE PDFs, Fig. 4.8 and Fig. 4.9 show significant systematic differences in the scaled PDFs at the different levels. The PDF reaching out to the highest densities corresponds to the density field at the ‘root’ level (0-level). While all level PDFs are of comparable amplitude at low densities, there are marked differences at the high density tail. All level PDFs have a more or less steep drop towards zero at high densities, be it that the corresponding density maximum is smaller as the level increases: the range of density values shrinks as the level of the circles increases. This can be understood from the fact that the self-similarity of our point distributions is limited by the finite number of points in the realizations. The various level PDFs can only be expected to be equivalent for perfectly self-similar point distributions. The finite number of points implies model representations with a finite range of densities and spatial scales. Upon scaling the densities at a particular level this translates into a decreasing maximum density as the level is higher.

Even though the PDFs are not exactly self-similar it is still very instructive to compare the PDFs corresponding to the density fields as reconstructed by the different methods. Even taking into account the limited range of self-similarity, comparison between the TSC, SPH

and DTFE PDFs for both the $\eta = 2$ and $\eta = 4$ realizations does produce two important observations.

4.4.2.1 Self-similarity and scaling range: comparison

The first observation is that the root level TSC PDF drops to zero at much lower densities than the root level SPH and DTFE PDFs, with the root level DTFE PDF extending out to an order of magnitude higher density than the SPH PDF. This is a reflection of the considerably larger spatial extent of the TSC kernel. It results in structures which are smeared out over larger areas and which have significantly lower densities. Also the SPH smoothing kernel is (somewhat) larger than the effective DTFE smoothing level. This holds equally true for each of the levels present in the model realizations. Evidently, reflected by the superior range of the DTFE PDFs, the DTFE is the method that manages to cover the largest dynamic range in density values.

The second difference is related to the self-similar behavior of each of the model reconstruction PDFs. Even though none of the three methods produces scaled PDFs which are exactly equal at all levels there are substantial differences in the way they deviate from self-similarity. The PDFs of the TSC reconstructions do not match at any density range. Even while taking into account the generic behaviour of higher level PDFs falling to zero at lower densities than the lower level PDFs, we find that for both the $\eta = 2$ and $\eta = 4$ realizations the midrange shoulders of the scaled PDFs are larger as the level is higher. This is also a reflection of the TSC kernel's properties: compact structures of high density are smeared out over large areas of lower density. The scaling of the density fields in higher level circles to lower density values (Eqn. 4.5) projects these values to even lower densities, producing the shoulders visible in Fig. 4.8. To a lesser degree the same phenomenon is visible in the SPH reconstructions, mainly visible in the case of the highest level circles. The proper scaling at most levels, however, is a reflection of the fact that the SPH smoothing kernel adjusts its scale to the local density of the point distribution. The shoulders in the SPH PDFs for the highest level circles is a result of the remaining smoothing of the SPH kernel (visible in the range of $0 < \rho < 5$ for $\eta = 2$ and $0 < \rho < 100$ for $\eta = 4$). Most telling is the finding that for the DTFE PDFs there are no shoulders visible! Thus, of the three methods, the DTFE is the one method which manages to recover the presence of self-similar scaling behaviour in its density field reconstruction.

4.4.2.2 DTFE versus SPH

While the rigidity and inflexibility of the TSC grid reconstructions make them evidently inferior to the adaptive SPH and DTFE methods, an evaluation of the two latter methods should be based on a somewhat more subtle assessment. The differences between the SPH and DTFE PDFs are elevated by assessing their ratio, depicted in the lower right-hand frame of Fig. 4.8 and Fig. 4.9.

For the lowest density bin ($\rho = 0.1$) the SPH PDF has a much larger value than that of the DTFE PDF, for both the $\eta = 2$ and $\eta = 4$ realizations. It is a consequence of the considerably more rapid exponential fall-off of the SPH kernel compared to the linear fall-off of the DTFE kernel. Reversely, at the highest density values we see that the ratio of DTFE to SPH tends towards infinity as the SPH PDFs fall off to zero much more quickly than the DTFE PDF: in

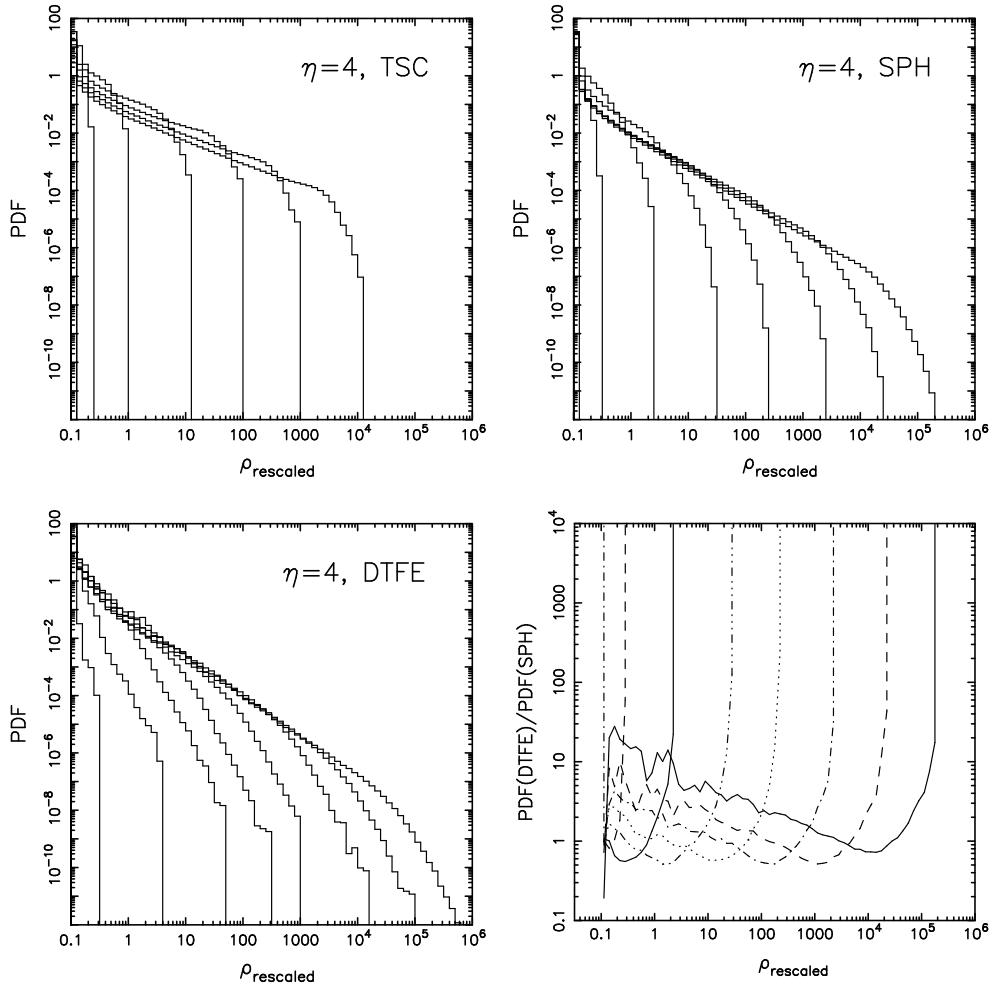


Figure 4.8 — Average PDFs of the density field in circles of different level (see text for a description) for the different reconstruction methods. In the bottom right-hand frame the ratio of the SPH and DTFE PDF is plotted. Model with $\eta = 4$, $\lambda = 3$ and $L = 8$.

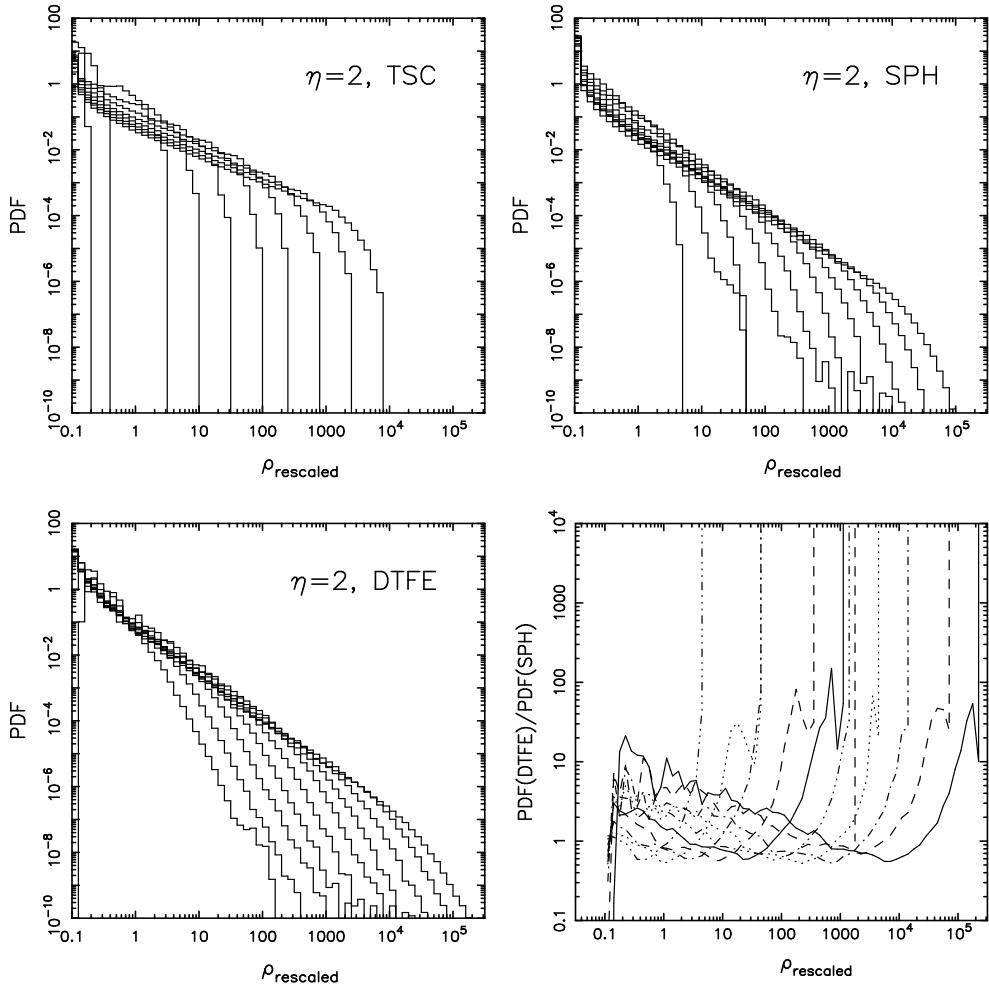


Figure 4.9 — Average PDFs of the density field in circles of different level (see text for a description) for the different reconstruction methods. In the bottom right-hand frame the ratio of the SPH and DTFE PDF is plotted. Model with $\eta = 2$, $\lambda = 1.75$ and $L = 14$.

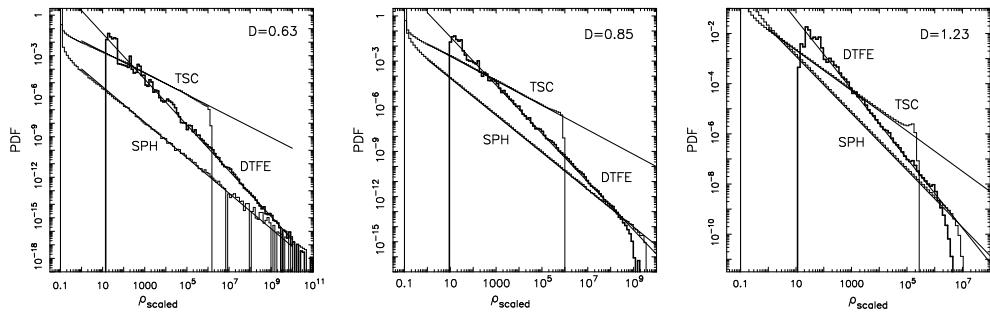


Figure 4.10 — Scaled PDFs of Soneira-Peebles density field reconstructions. Each frame corresponds to a Soneira-Peebles realization of a different fractal dimension, denoted in the upper right-hand corner. In each frame the TSC, SPH and DTFE reconstructed PDFs are shown.

the DTFE field one finds sufficient locations of very high density while they are absent in the SPH field. Meanwhile in the medium range of densities, we find the DTFE PDF to be higher than the SPH PDF, with a ratio smoothly falling off as a function of density until it attains a value lower than unity. Note that the extent of the medium range and the density value at which the DTFE/SPH ratio diverges towards infinity is a function of the level at which the density field has been determined.

On the basis of this ratio evaluation we see that the DTFE is able to resolve structures at a finer resolution and over a larger density range. By contrast, on average SPH smears out the finest structures over a larger area, resulting in more moderate density values.

4.4.2.3 PDF power-law scaling

As we have seen above the density field reconstructions of all three methods have a range over which the PDF has a power-law behaviour. The root level PDFs have the most extended range over which they attain a power-law character, with apparently only the DTFE reconstructions mapping each level in a fully self-similar fashion onto the basic 0-level PDF. The value of the PDF slope α can be deduced from the self-similar properties of the singular Soneira-Peebles model (see Eqn. 4.6). Fig. 4.10 reveals that the value of the slope of the power-law PDFs are not the same for the different reconstruction procedures.

We have measured the PDFs corresponding to three different Soneira-Peebles realizations, each with a different fractal dimension. For each of the three models we show the resulting PDFs for the TSC, SPH and DTFE density field reconstructions in Fig. 4.10. Going from left to right, the frames in this figure correspond to Soneira-Peebles realizations with fractal dimensions of $D = 0.63$, $D = 0.85$ and $D = 1.23$.

In determining the scaled PDFs in the log-log diagrams of Fig. 4.10 we have not corrected the density field inside the scaled circle for the smaller amount of points with respect to a 0-level circle. The reason is that a scaled circle does not represent a homogeneous sampling of the density field inside the 0-level circle at a lower sampling density. On the contrary, the sampling density is the same, with the exception of the high density cores which are sampled by a smaller amount of points. These high density regions have been removed from the scaled circle (clearly illustrated by Fig. 4.8).

In order to compute the various power-law slopes α simple straight lines have been fitted

Table 4.2 — Slopes of the power-law region of the PDF of a Soneira-Peebles density field as reconstructed by the TSC, SPH and DTFE procedures. The theoretical value (Eqn. 4.6) is also listed. Values are listed for three different Soneira-Peebles realizations, each with a different fractal dimension D .

D	$\alpha(\text{theory})$	$\alpha(\text{TSC})$	$\alpha(\text{SPH})$	$\alpha(\text{DTFE})$
0.63	-1.69	-0.81	-1.32	-1.70
0.86	-1.57	-0.82	-1.24	-1.60
1.23	-1.39	-0.79	-1.13	-1.38

to the log-log diagrams of the power-law sections of the derived PDFs. These straight lines have also been inserted in Fig. 4.10. The derived values α are listed in Table 4.2, along with the expected theoretical value for the three different Soneira-Peebles models (Eqn. 4.6). If anything, table 4.2 forms an surprisingly strong testimony of the success of the DTFE reconstructions to probe the full hierarchical substructure of a model! While the TSC and SPH reconstruction procedures fail to reproduce the correct expected slope, for all three models the DTFE slopes are fully consistent with the theoretically expected values. The minor differences between the DTFE and theoretical slope are mere statistical noise, to be ascribed to the finite number of points and the errors in the (linear) DTFE reconstruction (see Chapter 8).

4.4.2.4 Singular model PDFs: summary

Amongst the three density field reconstruction methods – TSC, SPH and DTFE – the DTFE method has been shown to be by far the most succesful in reproducing not only the qualitative impression of the density field structure but also the quantitative characteristics of the one-point density field distribution function. The DTFE method is superior in

- the dynamic range of (scaled) density field reconstructions;
- its ability to uncover the self-similar scaling in a density field;
- recovering the power-law PDF of a reconstructed density field, including the theoretically expected quantitative values of a characteristic parameter like the slope of the power-law.

With the singular Soneira-Peebles model as test template, the failure of the TSC and SPH reconstructions and the success of the DTFE to accurately recover the self-similar properties of the corresponding point distributions is a strong argument for using a tessellation-based method like the DTFE for properly assessing hierarchical cosmological mass distributions.

4.4.3 Extended models: Kolmogorov-Smirnov tests

Straightforward self-similar power-law scaling as observed for the singular Soneira-Peebles models does not exist for the extended Soneira-Peebles models. It renders the differences between particularly the SPH and DTFE reconstructions more subtle. To assess the resulting reconstructions we first study the corresponding one-point distribution functions. In the next

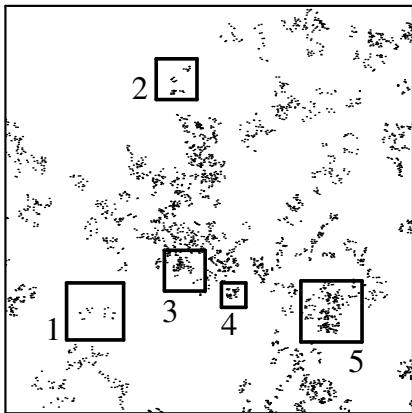


Figure 4.11 — The five regions to which we have applied a Kolmogorov-Smirnov test. The complete box corresponds to the finest zoom-in of the extended Soneira-Peebles model shown in the top right-hand frame of Fig. 4.6.

subsection this is followed by a study of the second-order moment of the density field, ie. of the two-point correlation function.

Because we cannot test for power-law scaling and compare with a theoretically inferred value of power-law slope of the PDF, we choose to follow an alternative quantitative statistical test for assessing the quality of the density field reconstructions. The two-dimensional Kolmogorov-Smirnov (K-S) tests (Peacock 1983, Fasano & Franceschini 1987, see also Press et al. 1992) allow us to qualitatively assess the quality of the different reconstructions by comparing the statistics for the density field reconstructions with the same statistics computed directly from the Soneira-Peebles point distribution.

In the two-dimensional K-S test the integrated probability around each data point (x_i, y_i) is calculated for each of four quadrants around this point. The same is done for the reconstructed density field $\widehat{\rho}(\mathbf{x})$. The two-dimensional K-S statistic D is then taken to be the maximum difference (ranging both over data points and quadrants) of these two integrated probabilities. Fasano & Franceschini (1987) have derived the probability that D is larger than observed for a binomial random point process of N points distributed over an intensity field $\lambda(\mathbf{x}) = \widehat{\rho}(\mathbf{x})$. Fasano & Franceschini (1987) demonstrated that their equations for the K-S probability give accurate results only for probabilities smaller than around 0.20. For larger probabilities the precise values may differ from the calculated values.

Using the results of Fasano & Franceschini we have calculated the K-S probability for the extended Soneira-Peebles point distribution shown in Fig. 4.6 and the corresponding reconstructed fields. We measured the K-S probability in 5 selected regions, indicated in Fig. 4.11. The reason for this is twofold. One is that the substructures reach down to very small scales, imposing practical difficulties in computing the integrated probabilities for the reconstructed fields. In addition, we are particularly interested in how well individual small-scale substructures are reconstructed. These regions all lie within the finest zoom-in shown in Fig. 4.6, i.e. within the region shown its top right-hand frame.

The resulting K-S probabilities for the 5 regions are listed in Table 4.3. We may conclude that for SPH and DTFE the data and models are not significantly different and thus represent reasonable reconstructions. Only the TSC reconstructions of the 5 regions differ significantly from the point distribution. It is interesting to note that the calculated K-S probability is always larger for the DTFE reconstruction than for the SPH reconstruction. It indicates that the

Table 4.3 — Two-dimensional Kolmogorov-Smirnov tests. Listed is the probability that the two-dimensional K-S statistic is larger than measured in the TSC, SPH and DTFE reconstructions for 5 selected regions in the extended Soneira-Peebles model realization (see Fig. 4.11).

region	# points	TSC	SPH	DTFE
1	16	$2.8 \cdot 10^{-3}$	0.43	0.50
2	32	$3.5 \cdot 10^{-3}$	0.46	0.60
3	94	$1.7 \cdot 10^{-8}$	0.73	0.97
4	32	$3.2 \cdot 10^{-4}$	0.82	0.86
5	225	$1.8 \cdot 10^{-6}$	1.0	1.0

DTFE reconstruction is in better agreement with the point distribution than the SPH reconstruction.

For both the SPH and DTFE reconstructions the K-S probability is larger for regions containing a larger number of points. This indicates that both reconstruction methods perform better on larger scales. This is not the case for the TSC method for which the probability is smallest for regions 3 and 5, the ones containing most points. This is a result of the rigid spatial resolution scale of the TSC method, centering around an optimal scale, in the order of the size of one gridcell. TSC cannot resolve structures smaller than this scale.

Nonetheless, while the Kolmogorov-Smirnov test offers a reliable statistical measure of how well a certain reconstructed field describes a given point distribution, its outcome is a single number which does not provide physical information on how and where the reconstructed field deviates from the point distribution.

4.4.4 Extended models: two-point correlation function

To obtain information on the performance of the field reconstructions as a function of scale we have studied the second-order autocorrelation function of the (continuous) reconstructed density fields.

4.4.4.1 Field autocorrelations and point correlations

The continuous field second-order autocorrelation function is defined by

$$\xi(r) = \langle \delta(\mathbf{x}) \delta(\mathbf{x} + \mathbf{r}) \rangle, \quad (4.7)$$

in which $\delta(\mathbf{x})$ is the density excess or deficit with respect to the mean density ρ_b of the field,

$$\delta(\mathbf{x}) = \frac{\rho(\mathbf{x}) - \rho_b}{\rho_b}. \quad (4.8)$$

The field autocorrelation function is the first order characteristic of spatial correlations in a density field. In practice, it is usually probed on the basis of the galaxy distribution, through its two-point correlation function $\xi_{gg}(r)$. It describes the degree of clustering as a function of the distance between the sample points (galaxies).

Mathematically, the two-point correlation function ξ_{pp} is defined by the excess probability of finding a neighbour a distance r from a given point. By regarding this as the probability of

finding a pair with one object in each of the volume elements dV_1 and dV_2 , $\xi_{pp}(r)$ is defined as

$$dP = n_0^2 [1 + \xi_{pp}(\mathbf{r})] dV_1 dV_2. \quad (4.9)$$

Here n_0 is the average number density of points in the sample. The two-point correlation function is closely affiliated to the autocorrelation function of the continuous field it is supposed to sample, with the two definitions being equivalent when the discrete point process fairly samples the underlying density field (see Bertschinger 1992). Within the cosmological context standard lore assumes an isotropic universe, implying that $\xi(\mathbf{r}) = \xi(r)$.

4.4.4.2 Significance

In cosmology $\xi(r)$ plays a pivotal role (see e.g. Peebles 1980). Not only is it the first order measure of deviations from homogeneity – and as such has been the one clustering characteristic which almost without exception has been determined for any large scale galaxy redshift survey – but in itself has also a strong physical significance as the Fourier transform of the power spectrum. In the early linear phases of cosmic structure evolution it thus embodies a complete specification of (Gaussian) primordial fluctuations while in later phases it retains its key role as link between the non-linear matter distribution and the associated cosmic matter flows.

For a full specification of the (non-linear) matter distribution one would need to determine the full range of higher-order correlation functions. In practice this is unfeasible. A possible and alternative approach would be to accomplish this via the continuous equivalents of the point correlation functions. Here we investigate which continuous field reconstruction method does indeed yield reliable and sensible estimates of the autocorrelation function. If any of these – TSC, SPH or DTFF – does indeed prove to be measure $\xi(r)$ correctly, it will be worthwhile to investigate its higher order correlation aspects further. As yet, $\xi(r)$ serves as an indicator of how well a reconstructed field describes a point distribution at a particular scale.

4.4.4.3 Autocorrelations: TSC, SPH and DTFF reconstructions

We determined the two-point correlation function (Eqn. 4.9) ξ_{pp} of the extended Soneira-Peebles point distribution (Fig. 4.6), as well as the second order autocorrelation function (Eqn. 4.7) $\xi(r)$ of the corresponding TSC, SPH and DTFF density field reconstructions. The thick solid line in Fig. 4.4.4.3 depicts the two-point correlation function of the point distribution, while the solid thin lines correspond to the autocorrelation functions of the various continuous fields (tagged according to the method).

Note that at very small distances all three autocorrelation function differs from the two-point correlation function. Note that the wiggles of the two-point correlation function are an artefact resulting from the fact that only a very few point pairs are involved: below a scale of $r \sim 2 \cdot 10^{-6}$ no point pairs are present. This is not the case for the reconstructed fields, by construction they yield continuous correlation values all the way down to $r = 0$.

At sufficiently large scales all reconstructed fields recover the same correlation function as the point distribution. At smaller scales, on the other hand, we find deviations between the various autocorrelation functions and the two-point correlation function. All three methods display a shoulder, its location and amplitude determined by the effective kernel scale. The TSC method fits the point distribution beyond a scale of $r \gtrsim 0.002$ (in units of boxlength), the

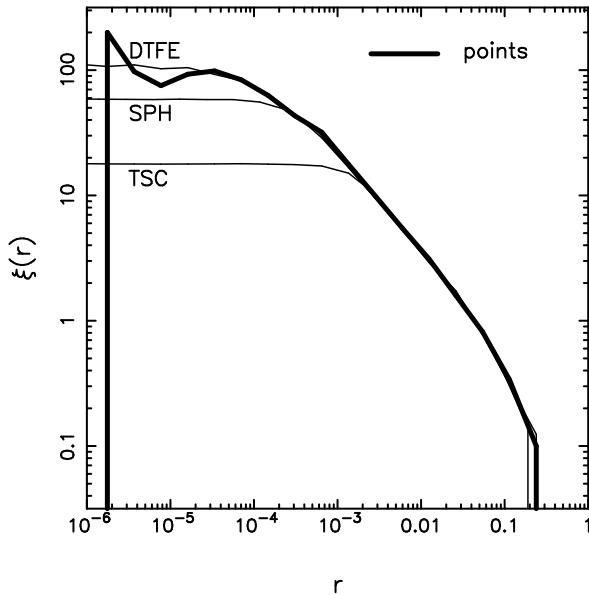


Figure 4.12 — Two-point correlation function $\xi(r)$ of the extended Soneira-Peebles model as a function of distance r . The thick solid line indicates the two-point correlation function of the point distribution. The thin solid lines correspond to the autocorrelation functions of the reconstructed fields, tagged accordingly with the name of the reconstruction method. The scale r is in units of boxlength.

SPH method beyond a scale of $r \gtrsim 2 \cdot 10^{-4}$ and the DTFE method beyond a scale of $r \gtrsim 2 \cdot 10^{-5}$. While at sufficiently large scales all three field reconstructions do reproduce the correct spatial correlation characteristics at smaller scales we do see major differences. The SPH and TSC procedure correlation functions break away from the two-point function at a considerably larger radius than the DTFE reconstruction: because of their larger effective kernel scale they are not able to follow the field correlations at a smaller range. Only the DTFE procedure seems to be able of giving a reasonably accurate description of the correlation function at the smallest scales ($r \lesssim 10^{-4}$),

In summary, the DTFE procedure manages to reproduce the correlation characteristics of the point distribution over many orders of magnitude. It yields confidence in its ability to be very suitable for studying not only second order correlations but also higher order clustering characteristics of the cosmic matter distribution.

4.5 Web-like spatial patterns: Voronoi clustering models

In the following sections we investigate the performance of the DTFE with respect to its ability to trace anisotropic web-like patterns. We do this on the basis of a class of heuristic models of cellular matter distributions, *Voronoi clustering models*. These models offer flexible templates for cellular patterns, they are easy to tune towards a particular spatial cellular morphology. The experience and knowledge obtained from the analysis of these test models will provide a necessary benchmark serving as reference and calibration point for more complex and intricate realistic circumstances. The latter, in the form of N -body simulations of

cosmic structure formation in various cosmologies is described in Chapter 6.

In the eighties Icke & van de Weygaert (1987) pointed out that the basic cellular geometry of the cosmic web may be successfully modelled by Voronoi tessellations. In the Voronoi clustering model, the geometric cellular structure of Voronoi tessellations is used as the skeleton of the cosmic matter distribution, defining the structural frame around which matter will gradually assemble in the course of the development of cosmic structure. The spatial patterns of the Megaparsec scale cosmic web are both geometric and stochastic in character. It is precisely these characteristics which are shared by Voronoi tessellations.

One might take the resulting Voronoi clustering models as pure ad-hoc toymodels, the result of a pure void-driven description of cosmic structure formation, in which a cellular pattern gradually unfolds as matter is set to migrate away from the primordial location of void centres towards the high density features in the cosmic foam (Icke & van de Weygaert 1987, van de Weygaert & Icke 1989). On the other hand, it may be argued that they form the asymptotic approximations for regular hierarchical cosmic structure formation scenarios, in which dominant voids at any one cosmic epoch have formed as a result of void merging process (Sheth & van de Weygaert 2004).

The great virtue of the Voronoi models is that they represent a conceptually simple model for a cellular or foam-like distribution of galaxies, providing a realistic rendering and representation of the spatial distribution of walls and filaments. This, in combination with their flexibility and versatility and ease of construction makes them ideal tools for statistical studies of spatial clustering. Also, because they are model distributions which are far less restricted in resolution and number of particles than conventional N -body experiments, cellular structure can be generated over a part of space beyond the reach of most N -body experiments. Voronoi models are therefore also useful for studying the properties of galaxy clustering in cellular structures on very large scales, for example in very deep pencil beam surveys, as well as for studying the clustering of clusters in these models. For an extensive review of Voronoi clustering models see van de Weygaert (2001) (see also Icke & van de Weygaert 1987, van de Weygaert 1991, van de Weygaert 2002).

4.5.1 The asymptotic cosmic web: Voronoi clustering models

Voronoi clustering models form an idealized and asymptotic description of the outcome of the cosmic structure formation process within gravitational instability scenarios. These models are based on the notion that voids play a key organizational role in the development of structure in the universe. In other words, they are based on the notion of the universe resembling a soapsuds of expanding bubbles. In its original form they were defined by Van de Weygaert & Icke (1989). The models assume that cosmic structure is the result of a distribution of expanding voids. Each void forms around a dip in the primordial density field. The involved density deficit corresponds to an effective repulsive peculiar gravity. As a result matter streams out of gradually expanding void-like regions, which would gradually evolve to a more spherical shape (Icke 1984), and ultimately collides with the surrounding matter distribution.

According to the prevailing cosmological view the primordial density field is a spatial Gaussian random field, with peaks and dips over a large range of scales. The power spectrum of the fluctuations is hierarchical, with small-scale structures emanating before they merge into the larger structures. Also voids evolve hierarchically, small voids merging into ever larger void-like regions. For an evolving void hierarchy there is an additional crucial process,

that of the disappearance of small voids embedded within large scale overdensities as the small voids collapse along with the latter. A detailed assessment of evolving void hierarchies, by Sheth & van de Weygaert (2004), involving an extension of the excursion set formalism (Bond et al. 1991), demonstrated that this leads to a peaked distribution of the sizes of voids around a characteristic value. In other words, within hierarchical scenarios of structure formation one would expect that at any one cosmic epoch most voids would have comparable sizes and excess expansion rates, which would involve some interesting implications for the resulting spatial matter distribution.

A geometrically interesting situation arises when the findings of Sheth & van de Weygaert (2004) are extrapolated to the asymptotic extreme in which the ‘peaked’ void distribution degenerates into a ‘spiked’ distribution marked by only one characteristic void size. Taking the voids as the dominant structure-shaping component of the universe the large scale structure can be seen as a close packing of spherically expanding regions. It will yield a cosmic matter distribution organized by a population of equally sized, spherical voids, all expanding at the same rate, akin to the scenario suggested by Icke (1984). It would involve a situation in which the matter distribution in the large scale universe is set up by matter being swept up in the bisecting interstices between spheres of equal expansion rate. In this idealization, the walls and filaments would be found precisely at the midplanes between expanding voids. This asymptotic description of the cosmic clustering process would lead to a geometrical configuration which is precisely that of a Voronoi tessellation.

4.5.2 Voronoi tessellations: the skeleton

An illustration of a set of three-dimensional Voronoi polyhedra is shown in Fig. 4.13. Around the central Voronoi cell, with shaded walls, the neighbouring Voronoi cells are depicted as wireframe objects. By means of this geometrical image we can easily appreciate the four characteristic morphological elements in three-dimensional tessellations. The central shaded polyhedron is a typical specimen of a Voronoi cell. Its boundary consists of a set of polygonal Voronoi walls. Each polygonal wall is shared with one of the neighbouring cells. The wireframe representation of these neighbouring cells shows that each Voronoi wall is outlined by the Voronoi edges, each forming the intersection of three Voronoi walls and three Voronoi cells. Finally, at the tip of each Voronoi edge we do find a Voronoi vertex (indicated by the dots) at the intersection of four Voronoi edges, six Voronoi walls and four Voronoi cells.

Within the cellular framework of the Voronoi model the interior of each Voronoi cell is considered to be a void region. The planes forming the surfaces of the cells are identified with the walls in the galaxy distribution. The edges delineating the rim of each wall are to be identified with the filaments in the galaxy distribution. In general, what is usually denoted as a flattened ‘supercluster’ or cosmic ‘wall’ will comprise an assembly of various connecting walls in the Voronoi foam, as the elongated ‘superclusters’ or ‘filaments’ will usually consist of a few coupled edges. Finally, the most outstanding structural elements are the vertices, tracing the surface of each wall, outlining the polygonal structure of each wall and limiting the ends of each edge. They correspond to the very dense compact nodes within the cosmic network, amongst which rich virialised Abell clusters form the most massive representatives.

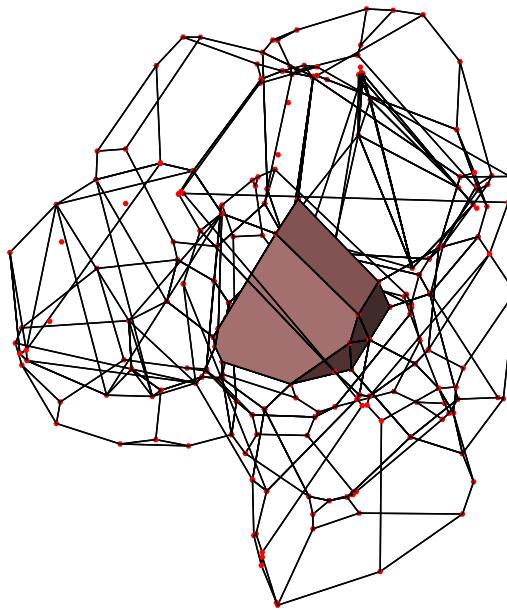


Figure 4.13 — Wireframe illustration of interrelation between various Voronoi tessellation elements. The central ‘Voronoi cell’ is surrounded by its wire-frame depicted ‘contiguous’ Voronoi neighbours. The boundaries of the cells are the polygonal ‘Voronoi walls’. The wire edges represent the Voronoi edges. The ‘Voronoi vertices’, indicated by red dots, are located at each of the two tips of a Voronoi edge, each of them located at the centre of the circumsphere of a corresponding set of four nuclei. Courtesy: Jacco Dankers.

4.5.3 Geometric clustering

A mere qualitative assessment of three-dimensional Voronoi geometries yields the interesting observation that the non-Poissonian distribution of the Voronoi walls, edges and vertices is a stochastic process characterized by strong spatial correlations. This is readily apparent from Fig. 4.13. The non-trivial morphology of spatially clustered geometrical elements not only determines the overall clustering properties of its galaxy population but also forms a stark contrast to less realistic stochastic toy models as e.g. the double Poisson process. The important repercussion is that the geometric Voronoi components themselves are grouping into coherent ‘super’structures, inducing intrinsic spatial correlations over scales substantially superseding the basic cell scale (van de Weygaert 2002). The implied superclustering has been strikingly illustrated in the analysis of the clustering properties of vertices in Voronoi tessellations by van de Weygaert (2001, 2002). The remarkable success of Voronoi models to reproduce the clustering of galaxies and clusters of galaxies should therefore be seen in the light of the stochastic, non-Poissonian and geometric nature of the spatial distribution of walls, filaments and clusters framing the cosmic web. It demonstrates the viability of the Voronoi models as flexible templates for studying realistic galaxy distributions around geometrical features represented by the corresponding components in the Voronoi tessellations.

4.5.4 Voronoi clustering models: formalism

In practice the Voronoi clustering models are based on the distribution of galaxies within the various elements of this geometric structure. The obvious shortcoming of these models is the fact that they do not and cannot address the galaxy distribution on small scales, i.e. the distribution within the various components of the cosmic skeleton. This will involve the complicated details of highly non-linear small-scale interactions of the gravitating matter.

For our purpose we take the route of complementing the large-scale cellular distribution induced by Voronoi patterns by a user-specified small-scale distribution of galaxies. We distinguish two different yet complementary approaches (see van de Weygaert 2002). One is the fully heuristic approach of *Voronoi element models*, genuine tools for the systematic investigation of specific individual details of the full cellular structure. They are particularly apt for studying systematic properties of spatial galaxy distributions confined to one or more structural elements of non-trivial geometric spatial patterns. The second, supplementary, approach is that of the *Voronoi evolution models*, which attempts to ‘simulate’ foam-like galaxy distributions on the basis of simplified models of the evolution of the Megaparsec scale distribution. One suggestive model would relate to explosion models (see e.g. Ostriker 1988), once considered a viable alternative for gravitational instability scenarios but in the meantime rendered obsolete by a large array of cosmological observations. One particular noteworthy class of Voronoi models are the *Voronoi kinematic models*. These represent an asymptotic description of clustering in gravitational instability scenarios of cosmic structure formation.

Based on the generic formalism to generate distributions of galaxies within Voronoi tessellations, described in detail in appendix 4.B, the details of the various parameters of the formalism define the different classes of Voronoi cluster models. In all cases, the different models are based on the displacement of a sample of N ‘model galaxies’. The initial spatial distribution of these N galaxies within the sample volume V is purely random, their initial locations $\mathbf{x}_n(t_0)$ ($n = 1, \dots, N$) defined by a homogeneous Poisson process. A set of M nuclei or expansion centres within the volume V corresponds to the cell centres, or expansion centres driving the evolving matter distribution. The nuclei have locations \mathbf{y}_m ($m = 1, \dots, M$). The first step of the formalism is to determine for each galaxy n the Voronoi cell \mathcal{V}_α inside which it is initially located.

The path $\mathbf{x}_n(t)$ along which a galaxy moves within the Voronoi skeleton depends on how far it has moved away from its initial location $\mathbf{x}_n(t_0)$. For a specific galaxy n this path, illustrated in Fig. 4.14, may consist of the following sequence,

- cell displacement $\mathbf{s}_\alpha(t)$ radially directed away from the expansion centre j_α ;
- wall displacement $\mathbf{s}_{\alpha\beta}(t)$ within the Voronoi wall $\Sigma_{\alpha\beta}$,
 $\Sigma_{\alpha\beta}$ defined by the nucleus j_α and its natural neighbour j_β ;
- edge displacement $\mathbf{s}_{\alpha\beta\gamma}(t)$ along the Voronoi edge $\Lambda_{\alpha\beta\gamma}$,
 $\Lambda_{\alpha\beta\gamma}$ defined by j_α and its natural neighbours j_β and j_γ .

This path is encapsulated in the equation

$$\begin{aligned}
 \mathbf{x}_n(t) &= \mathbf{y}_\alpha + \mathbf{s}_{n\alpha}(t) + \mathbf{s}_{n\alpha\beta}(t) + \mathbf{s}_{n\alpha\beta\gamma}(t) \\
 &= \mathbf{y}_\alpha + s_{n\alpha}(t)\hat{\mathbf{e}}_{n\alpha} + s_{n\alpha\beta}(t)\hat{\mathbf{e}}_{n\alpha\beta} + s_{n\alpha\beta\gamma}(t)\hat{\mathbf{e}}_{n\alpha\beta\gamma}.
 \end{aligned} \tag{4.10}$$

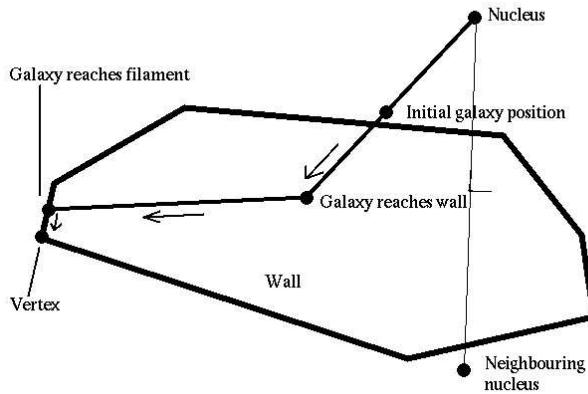


Figure 4.14 — Schematic illustration of the Voronoi kinematic model. Courtesy: Jacco Dankers.

The details of the complete formalism for generating these spatial distributions can be found in appendix 4.B.

The underlying cosmology may enter via two factors. The spatial distribution of the M nuclei may be identified with the void minima in the cosmic density field. If so, their locations are dictated by the primordial random density and/or potential field. Secondly, the time dependence of the global void expansion rate $R(t)$ with which the particles move out of the voids is directly coupled to the cosmological structure growth factor.

4.5.4.1 Voronoi element models

In this study we use Voronoi element models, tailor-made heuristic ‘galaxy’ distributions in and around specific elements of a Voronoi tessellation. Pure Voronoi element models are distributions in which all galaxies reside in either 1) the walls, 2) the edges/filaments or 3) the vertices/clusters. Unlike the Voronoi kinematic model or the Voronoi explosion model, all model galaxies are immediately projected onto wall, edge or vertex following the path depicted in the Fig. 4.14. They are generated with the following set of cell, wall and edge path factors (see appendix 4.B, Eqn. 4.38)

$$\text{Walls : } (s_{n\alpha}, s_{n\alpha\beta}, s_{n\alpha\beta\gamma}) = (v_n, 0, 0);$$

$$\text{Filaments : } (s_{n\alpha}, s_{n\alpha\beta}, s_{n\alpha\beta\gamma}) = (v_n, \sigma_n, 0); \quad (4.11)$$

$$\text{Clusters : } (s_{n\alpha}, s_{n\alpha\beta}, s_{n\alpha\beta\gamma}) = (v_n, \sigma_n, \lambda_n).$$

where the values of the parameters v_n , σ_n and λ_n characterize the crossing of the galaxies’ path with the wall, edge or vertex towards which it moves. *Singular Voronoi element models* place all model galaxies in either walls, edges or vertices. The versatility of the model also allows combinations of element models, in which field (cell), wall, filament and vertex distributions

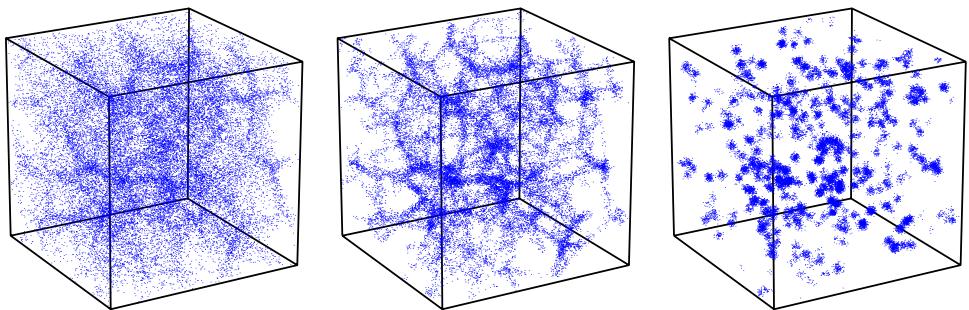


Figure 4.15 — Three different patterns of Voronoi element galaxy distributions, shown in a three-dimensional cubic setting. The depicted spatial distributions correspond to a wall-dominated Voronoi universe (left-hand frame), a filamentary Voronoi universe (central frame) and a cluster-dominated Voronoi universe (bottom frame).

are superimposed. The characteristics of the patterns and spatial distribution in these *mixed Voronoi element models* can be varied and tuned according to the fractions of wall galaxies, filament galaxies, vertex and field galaxies.

Characteristic examples of the resulting (singular element) galaxy distributions are represented in boxes in Fig. 4.15. The depicted distributions concern a wall-dominated Voronoi universe (left-hand frame), a filamentary Voronoi universe (centre) and a cluster-dominated Voronoi universe (right-hand frame). Note that a finite thickness is assigned to all elements. This is achieved by adding a random displacement according to a given density profile. In the depicted Voronoi wall model the galaxies are displaced according to a Gaussian density profile in the direction perpendicular to the wall to which they belong. In the Voronoi filament model the same was done for the directions perpendicular to the filaments, while in the Voronoi cluster model the galaxies are displaced such that the clusters represent three-dimensional Gaussian peaks. The scale of each of the Gaussian distributions define the width of the corresponding elements.

4.5.5 Voronoi clustering models: realizations

Three different three-dimensional Voronoi model galaxy distributions were generated. Each model belongs to the class of Voronoi element models (see Fig. 4.15). One model realization corresponds to the model in which galaxies are exclusively located inside walls, a second one where these are concentrated in and around filaments and a third one restricted to galaxies located within clusters.

Each realization is generated within a cubic box with periodic boundary conditions. In each box 512 nuclei are placed at random locations. The Voronoi tessellation of these 512 nuclei is the skeleton of the matter distribution. In total 262 144 particles, the model galaxies, are distributed within the corresponding Voronoi walls (the *wall model*), Voronoi edges (the *filament model*) or the Voronoi vertices (the *vertex model*) according to the formalism presented above. For all three models the structural features are given a Gaussian density profile in the direction(s) perpendicular to the relevant structures, with a width R_s amounting to 1/400 in units of the boxsize. Scaling the model distributions by assigning a characteristic size of about

$25h^{-1}$ Mpc (e.g. Hoyle & Vogeley 2002, Plionis & Basilakos 2002, Arbabi-Bidgoli & Müller 2002) for voids, the box is identified with a cosmological box of size $200h^{-1}$ Mpc. For a Gaussian width of 1/400th of the boxsize it implies a physical size of $R_w = R_f = R_c = R_s = 0.5h^{-1}$ Mpc for each wall, filament and vertex. The three-dimensional images in Fig. 4.15 provide an idea of the resulting spatial distributions. The distinct differences in morphology of the patterns delineated by the three Voronoi element distributions are emphasized by the (effectively) two-dimensional galaxy distribution within thin slices. The top row in Fig. 4.17 provides a telling illustration.

4.6 Voronoi clustering models: density field reconstructions

To assess the performance of the DTRE procedure in its ability to outline and recover spatial features in the density field corresponding with the generated galaxy distributions we analyze and compare the DTRE density field reconstructions for the three Voronoi element models with those produced by the TSC and SPH procedure. The TSC procedure (appendix 4.A.1) is rigid with respect to its scales sensitivity as well as its shape sensitivity, while the SPH procedure does involve a strong scale dependence but lacks shape sensitivity (appendix 4.A.2).

We applied the TSC, SPH and DTRE procedure to obtain the three-dimensional density fields of all three Voronoi model distributions. The panels in three subsequent rows of Fig. 4.17 depict the density field reconstructions by the TSC, SPH and DTRE technique for the wall model (left-hand column), the filament model (central column) and the cluster model (right-hand column) within the corresponding thin slice through the center of the box. These greyscale density field images reveal substantial qualitative differences between the TSC, SPH and DTRE reconstructions.

4.6.1 Filaments

For an impression of the quality of the density reconstructions images of the full three-dimensional structure and morphology of the patterns in the resulting density field offer the most direct and objective probe. It is particularly worthwhile to visualize the structures in a filamentary distribution. In Fig. 4.16 the central part of the sample box of the Voronoi filament model realization is shown. The size of the depicted box is 30% of the complete one.

The full three-dimensional galaxy distribution is shown in the top left-hand frame. The resulting TSC, SPH and DTRE density fields are represented in the top right-hand (TSC), bottom left-hand (SPH) and bottom right-hand (DTRE) frames. The density fields are represented by iso-density contour levels. The levels were chosen such that 65% of the mass is enclosed within regions of density equal to or higher than the contour level. The galaxy distribution in the upper left-hand frame, on the other hand, contains all galaxies within the region. Evidently, they have distributed themselves over a large range of densities and thus occupy a larger fraction of space.

The appearances of the TSC, SPH and DTRE patterns do differ substantially. Part of this is due to a different effective scale of the filter kernel. This can be directly appreciated from the fact that the 65% mass contour corresponds to a density contour $\rho = 0.55$ in the TSC field, $\rho = 1.4$ in the SPH reconstruction and $\rho = 2.0$ in the DTRE reconstruction (ρ in units of the average density).

The fine filamentary maze seen in the galaxy distribution is hardly reflected in the TSC

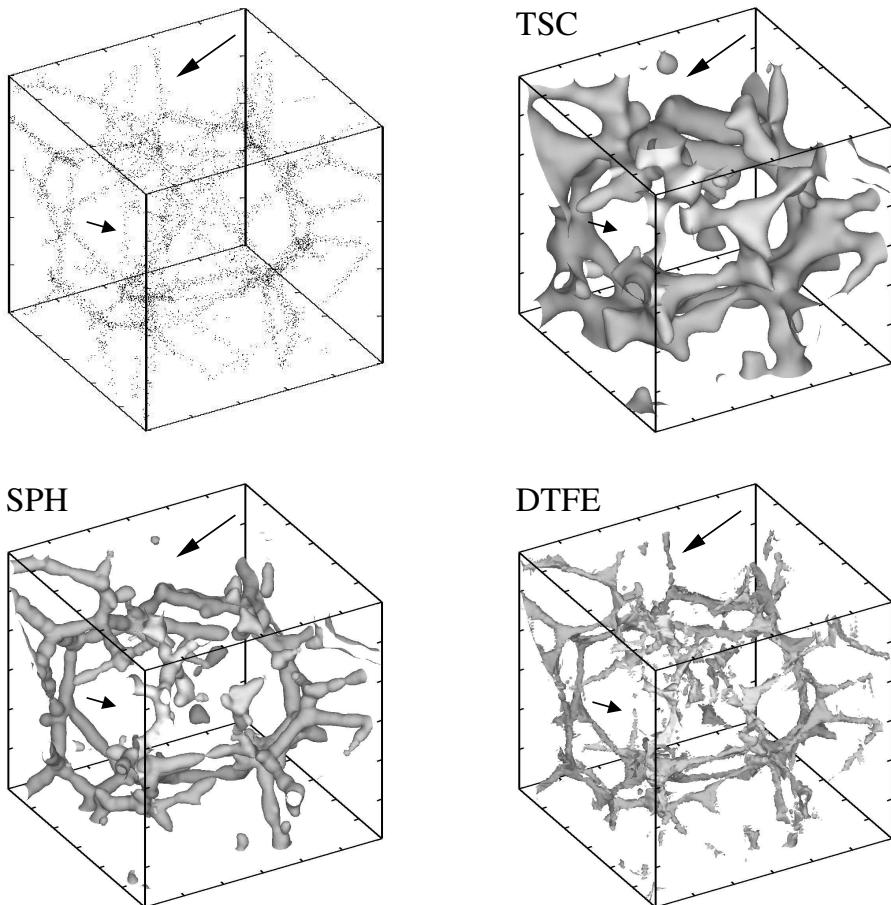


Figure 4.16 — Three-dimensional visualization of the Voronoi filament model and the corresponding TSC, SPH and DTFE density field reconstructions. The density contours have been chosen such that 65% of the mass is enclosed. The arrows indicate two structures which are visible in both the galaxy distribution and the DTFE reconstruction, but not in the TSC and SPH reconstructions.

grid based reconstruction. While the global structures, such as the almost ringlike arrangement of filaments halfway in the box, are also found in the TSC density field, we may identify a range of morphological features dependent on the density of the filaments and the cell-size of the TSC grid. The latter determines the minimum scale of the features in the TSC field, and tends to smear out individual structures over a significantly larger volume. In general we find that if still identifiable as an individual feature, they do not resemble their counterparts in the galaxy distribution. Long, thin and dense filaments in the galaxy distribution result into elongated features in the density field, be it with a substantially larger width. We can also observe the merging of less dense features with surrounding structures into flattened patches in the density field. Also it is almost unfeasible to recognize any individual filaments or nodes in regions with a large concentration of filaments.

The SPH density field fares considerably better at reproducing the web-like arrangement of filaments in the galaxy distribution. The basic configuration of the filamentary web is clearly outlined. Nonetheless, a comparison with the galaxy distribution shows that the filaments in the SPH density field are not as finely outlined as observed in the galaxy distribution: their width is determined by the scale of the SPH kernel which apparently tends to smooth out the features into a pattern of tubes. Also we see that the density field at various locations resembles a chain of spherical blobs, a result of the spherical character of the SPH kernel. These blobs are artefacts caused by changes in the galaxy number density. Bridging substantial density gradients is problematic for SPH reconstructions, for a major part due to the fact that the SPH kernel does not include any directional information. Because its scale is set by the closest point, at locations with large density gradients it may fail to reflect the proper scale in various directions.

It is the DTFE reconstruction (bottom right-hand frame of Fig. 4.16) which yields the most outstanding reproduction of the filamentary web-like character of the galaxy distribution. A detailed comparison between the galaxy distribution and the density surfaces show that it manages to trace the most minute details in the cosmic web. Note that the density contours do enclose only 65% of the mass, and thus relates to a smaller volume than suggested by the features in the galaxy distribution itself. The success of the DTFE method is underlined by identifying a few features in the galaxy distribution which were identified by the DTFE but not by SPH and TSC. The arrows in Fig. 4.16 point at two tenuous filamentary features visible in the galaxy distribution as well as in the DTFE field, yet entirely absent from the TSC and SPH fields. In comparison to the inflated contours of the SPH and TSC reconstructions, the structure outlined by the DTFE density field has a more intricated, even somewhat tenuous, appearance marked by a substantial richness in structural detail and contrast. At several locations structures in the galaxy distribution may be recognized which are also visible in the DTFE reconstruction, but not in the TSC and SPH reconstructions. Two such examples are indicated by the arrows in Fig. 4.16. Both these examples concern tenuous filamentary extensions. Finally, some artefacts of the DTFE method are also visible: in particular near intersections of filaments we tend to find triangular features which can not be identified with similar structures in the galaxy distribution. Nearby filaments are connected by relatively small tetrahedra, translating into high density features of such shape.

Some of the characteristics and differences emphasized in the discussion above can be recognized even more clearly when comparing the structure in a thin two-dimensional slice through the density field. The greyscale plots in the 2nd to 4th row in the central column

of Fig. 4.17 depict these two-dimensional sections through the filamentary cosmic web of the *Voronoi filament model*. The top frame shows the generating galaxy distribution in a section of finite thickness centered on the central slice of the density plots. Filaments tend to show up as mere points or short edges in the three-dimensional slice's galaxy distribution, marking the location where a three-dimensional filament pierces through the depicted slice. Only occasionally is the orientation of the filaments such that they form salient elongated edges within the slice. We find stark differences between the TSC and SPH greyscale plots on the one hand and the intricately traced pattern in the DTFE slice on the other hand. The low resolution of the TSC map is expected, but the large contrast of the SPH map versus the DTFE map is less so. The blobby character of the SPH reconstruction is quite obvious, as is the occasional related inability to trace various features. One such feature is the filament in the extreme lower left-hand corner. SPH is quite successful when it gets to tracing out more compact and less elongated concentrations, the dense concentration in the lower right-hand corner is a telling example. By contrast, it is the DTFE map which manages to outline the elongated features in most detail, reproducing a pattern of thinly outlined and dense edges intersecting at high density nodes. Occasionally one can even find features in the DTFE density map that cannot be traced in the particle distribution. Sometimes this is due to the fact that we are dealing with an interpolation algorithm that may find features in regions of low density, sometimes it is an artefact produced by the triangular DTFE kernels.

The imprint of the TSC, SPH and DTFE kernels on the resulting density field reconstructions can be appreciated in a more quantitative way, from the one-dimensional probes through the corresponding density field. The central panel of Fig. 4.18 shows such a probe through the Voronoi filament model realization. The location of the probe, indicated in the insert, is chosen such that it passes through four filamentary structures. In the case of the 2nd and 3rd filament the probe passes almost head-on, which results in the DTFE producing a thin and high peak. The dotted and dashed profiles are the same filaments according to the SPH and TSC reconstructions. In both situations the imprint of the SPH and TSC kernels is obvious: the broadly winged peaks are the product of the a sharply defined features with the corresponding kernels. The width of the latter leads to a considerable and artificial level of smoothing. This is particularly true when the width of the filaments is defined by only a few galaxies, as is often true for tenuous filaments (or walls). It is reassuring to see that the DTFE reconstruction does also manage to follow the natural scale of a feature. This can be observed when looking at the DTFE performance in the case of the first and fourth filament along the path of the probe. The latter passes the first and fourth filament obliquely, reflected in a more extended peak profile along the probe. While the kernel imprint is still visible in the SPH and TSC profiles of the first peak, we see a reasonable agreement between the TSC, SPH and DTFE reconstructions of the extended fourth peak. The differences between the different reconstructions concern the resolution of substructure within the structures. The DTFE finds even the finest substructures, in line with the earlier findings on the ability of the DTFE to trace hierarchical patterns. If anything, the DTFE peaks occasionally appear to be too sharp and thin. This is a consequence of its strongly local definition: the local linear DTFE interpolation may occasionally yield artefacts, the related higher-order natural neighbour interpolation technique would ameliorate the occasional sharp density gradients.

In all, we can conclude that the DTFE is highly successful at outlining the intricate web-like and filamentary distribution of the galaxies. Its ‘crispy’ character is a result of its kernel

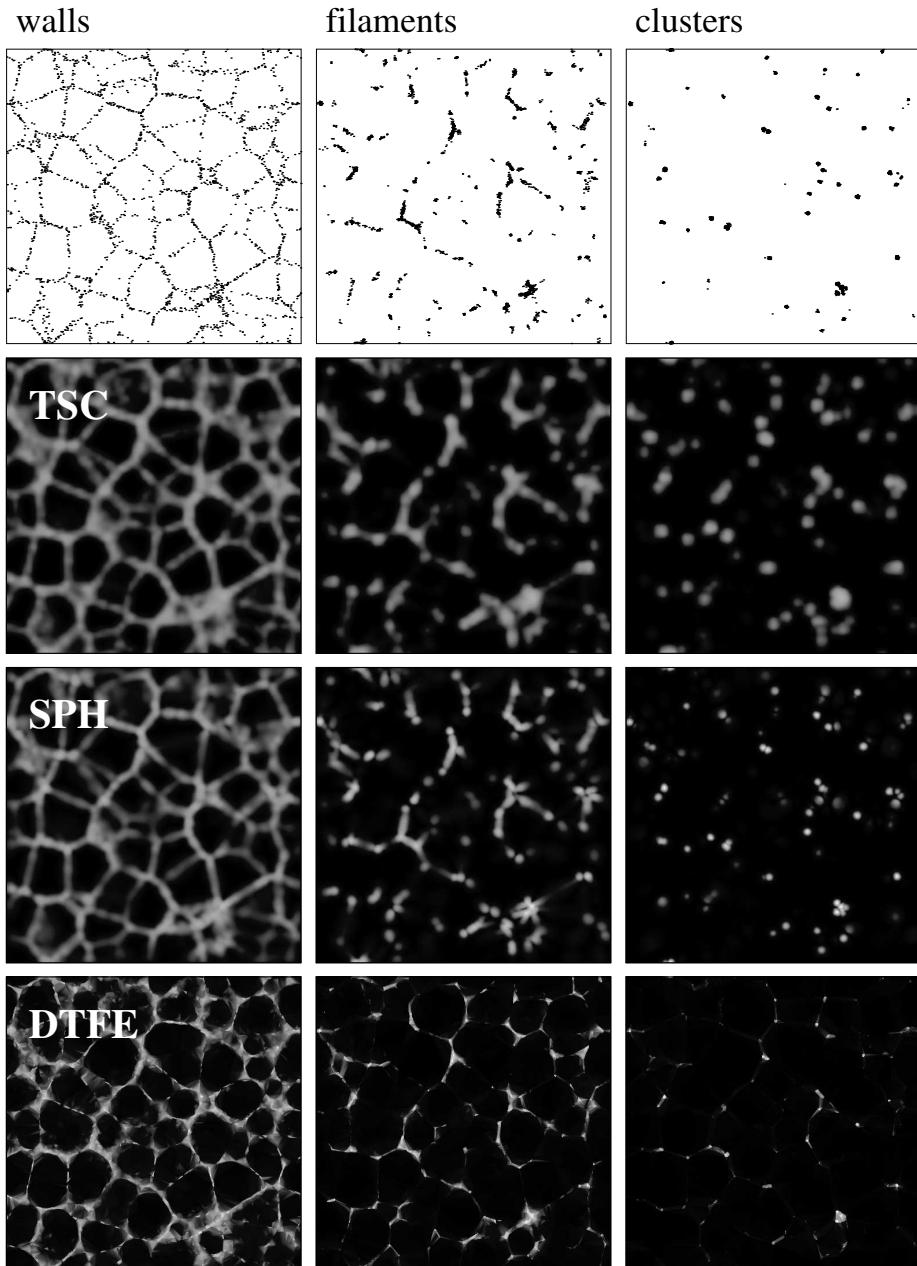


Figure 4.17 — Voronoi clustering models with galaxies located exclusively in either walls (left-hand column), filaments (central column) or clusters (right-hand column). Top row: thin slice through the corresponding galaxy distributions. Rows 2 to 4: two-dimensional slices through the three-dimensional density field reconstructions produced using the TSC, SPH and DTFE methods. Also compare the spatial galaxy distribution in Fig. 4.15.

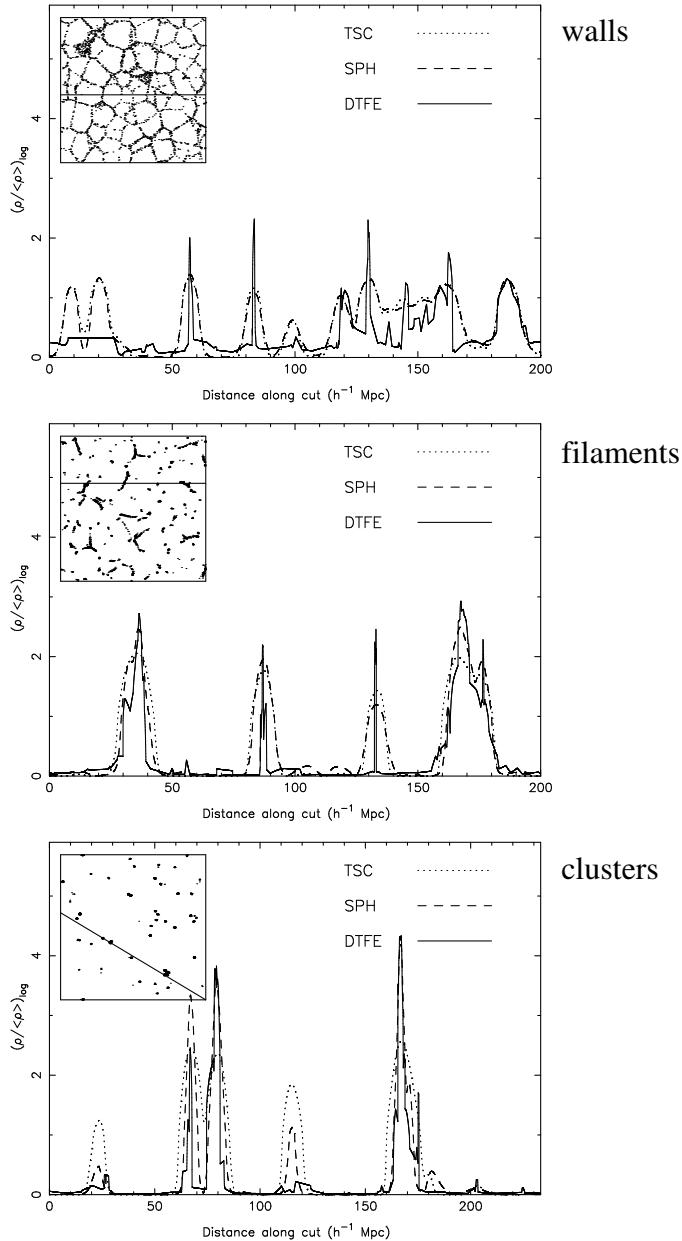


Figure 4.18 — Density profiles along one-dimensional sections through three-dimensional density field reconstructions of Voronoi ‘element’ models. Top: a one-dimensional section through a wall-like galaxy distribution. Centre: profile along a one-dimensional section through a filamentary galaxy distribution. Bottom: profile along a one-dimensional section through a cluster dominated galaxy distribution. The inserts in the upper left-hand corner of each frame show a two-dimensional slice through the simulation box with the solid line indicating the one-dimensional section along which the depicted density profiles have been computed. In each frame the profiles for the corresponding TSC (dotted), SPH (dashed) and DTFE (solid) density field reconstructions are shown.

being adaptive in both scale as well as in direction. The lesser performance of TSC, and particularly so SPH, is to be attributed to the nature of their kernel. While the rigid grid-based kernel of TSC prevents any adaptation to the local density, the adaptive nature of the SPH kernel enables SPH to largely adapt to the density. However, it fails to follow the morphology of the local density distribution.

4.6.2 Walls and clusters

The performance of the DTFE in the case of a filamentary web-like structure has been amply discussed above. For a complete census of its ability to probe web-like patterns we also need to understand its ability to probe either wall-like or cluster-dominated patterns. In principle it would have been possible to show and analyze three-dimensional iso-density contours analogous to those in Fig. 4.16 for the Voronoi wall and vertex models. However, due to their extent and low density the walls occupy a much larger fraction of space and tend to block each other from view. This will render it difficult to get a transparent view of the structure in the various density field reconstructions. Also for the Voronoi cluster model configuration we do not present a three-dimensional image for assessing its main characteristics. The clusters/vertices define a mainly isolated population of objects without strong and coherent structural connections. Therefore we restrict ourselves to a discussion of the greyscale maps of a planar section through the density field reconstructions, in combination with a comparison of the density profiles along a one-dimensional probe through the density field.

The left-hand column of Fig. 4.17 contains a set of two-dimensional sections through the TSC, SPH and DTFE density field reconstructions of the Voronoi wall model. Walls have a substantially lower density than their filamentary peers, they are large moderately dense two-dimensional slabs. A planar section through such a distribution will produce a pattern in which the walls are recognizable as edgelike features, in the case of the Voronoi model defining a tessellation pattern of two-dimensional cells. Only rarely the orientation of a wall will be such that it lies within or closely parallel to that of the sectional plane.

The results for the Voronoi wall model are comparable to that of the filament model. The distinctly anisotropic planar geometry of the wall models, characterized by thin walls, results in more puffy planar features for the TSC and SPH method while sharp structures are the product of the DTFE technique. The differences between the TSC and SPH reconstructions and DTFE reconstructions can be appreciated most directly from the one-dimensional probes in Fig 4.18. As for the filamentary distributions the intersections with the walls are narrow and strongly peaked spikes, while the TSC and SPH sections are considerably smoother and broader.

The Voronoi cluster model involves a spatial distribution of galaxies confined to compact clusters located at the vertices of a Voronoi tessellation. An example of such a realization is shown in the right-hand column of Fig. 4.17. Also here we recognize the poor spatial resolution of the TSC reconstructions. As a result of its large smoothing length the vertices can be recognized as large smooth blobs. Occasionally, such as for the cluster concentration in the lower right-hand corner, this leads to a featureless blob resulting from the merging of several clusters. In general, TSC does manage to reproduce the spatial distribution of clusters on scale superseding the kernel size, be it that it may have problems in highly concentrated regions. In general, it would be quite difficult to determine reliable peak density values or cluster masses in such supercluster regions. On the basis of a qualitative assesment it seems

that the SPH reconstruction performs better than the DTFE reconstruction. The clusters also look like spherical blobs, because of the adaptive SPH scale of the SPH kernel each individual cluster is clearly identified. The mass or density of each cluster is easily recognizable. The high number density of galaxies contained within spherically symmetric concentrations is beneficial for successful SPH reconstruction of clusters. However, when clusters are defined by less points than enclosed by the SPH kernel (~ 40), SPH encounters problems. While the DTFE is also successful in identifying nearly every cluster, the compact and peaked nature of the reconstruction is reflected in the almost pointlike objects. While the spherical nature of the SPH kernel is ideal for these cluster configurations, the DTFE formalism and the tetrahedral nature of the DTFE kernel yield several artefacts in the regions between the clusters: one can recognize typical triangular (low density) wings and various thin and tenuous artificial connections between the clusters. A faint imprint of web-like features can be seen in the two-dimensional section of Fig. 4.17, a reflection of the web-like spatial distribution of Voronoi vertices. Interestingly, the one-dimensional section passing through or near some five clusters do show a rather realistic rendering by the DTFE of the one-dimensional density profiles of three clusters, while the dominant imprint of their kernel mark both TSC and SPH reconstructions. The location of the section has been chosen such that it passed through or close to two clusters (located at about $24h^{-1}$ and $114h^{-1}$ Mpc along the probe), while passing straight through three other clusters (located at about $68h^{-1}$, $80h^{-1}$ and $166h^{-1}$ Mpc along the probe). Also notice that the DTFE hardly notices the presence of the first and fourth cluster due to the corresponding thin profiles, while TSC and SPH do detect a significant peak as the probe passes within a distance smaller than the kernel scale. Both the TSC and SPH recover a significant peak at places where the probe passes close to a cluster, while the DTFE signal is much smaller there (roughly two orders of magnitude). These differences may again be understood in terms of differences in the smoothing kernels employed by the reconstruction techniques.

4.7 Voronoi clustering models: shape and morphology analysis

The DTFE reconstructions of Voronoi clustering models, as well as those by the competing TSC and SPH methods, are evaluated on the basis of a qualitative visual assessment as well as by means of a few quantitative tests. The visual comparison of the TSC, SPH and DTFE density field reconstructions of the Voronoi wall model, Voronoi filament model and Voronoi cluster model do reveal various systematic and essential differences. Two aspects stand out:

- *Morphology*: the *local appearance* of the individual characteristic building blocks of the models (walls, filaments and clusters);
- *Topology*: the *global appearance* and connectivity of the reconstructed network. An important topological characteristic is the volume filling fraction.

While there are many conceivable measures for various morphological and topological aspects of the matter distribution, the quantitative tests presented here will address a few specific but highly characteristic properties which highlight the ability of the DTFE to correctly trace the morphology and topology of web-like patterns. The determination of the shape/anisotropy of the various features in our models forms a test for the DTFE's morphological sensitivity

while we compare the volume filling fraction of the reconstructed mass distributions in order to assess the topological virtues of the DTFE.

4.7.1 Shape and morphology

An important measure of the local density distribution concerns the shape of the density contour levels. Instead of a global assessment we identified various representative features in the three Voronoi element models and studied the shape in and around that feature on a range of spatial scales. In each model we measured the shape/local anisotropy in terms of the axis ratios of the local density distribution. The latter are computed from the eigenvalues of the mass inertia tensor.

Placing ourselves at the center of the feature we determine the shape of the selected Voronoi structures, i.e. walls, filaments or clusters. First the mass inertia tensor $I_{ij}(R)$ is determined inside a number of concentric spheres of radius R . The range of radii runs from one to five times the width of walls, filaments and clusters in our model realizations.

The shape of the particle distribution is taken as the reference point against which we judge the quality of the shape in the corresponding continuous field representations of the same feature. To ascertain proper mass/particle resolution we proceed by identifying a particular feature and zoom in on its position by placing it at the center of a surrounding three-dimensional subvolume. The surface density/density of walls and filaments in the Voronoi clustering models is uniform. The mass resolution of the walls, filaments and/or clusters can therefore be simply increased by them in with a proportionally larger amount of particles (resampling).

For the discrete particle/galaxy distribution the 3×3 mass inertia matrix $I_{ij}(R)$ is computed by summing over all particles \mathbf{x}_k within a distance R from the central position \mathbf{r}_e ,

$$I_{ij}(R) = \sum_k m_k (x_{k,i} - r_{e,i})(x_{k,j} - r_{e,j}). \quad (4.12)$$

For convenience, we have set the mass of each particle equal to unity, $m_k = 1$.

For the equivalent situation for the continuous density fields of the TSC, SPH and DTFE reconstructions we compute the inertial tensor $I_{ij}^{\text{TSC}}(R)$, $I_{ij}^{\text{SPH}}(R)$ and $I_{ij}^{\text{DTFE}}(R)$ from the integral over the same region,

$$I_{ij}(R) = \int_{(|\mathbf{x}-\mathbf{r}_e| < R)} d\mathbf{x} \rho(\mathbf{x}) (x_i - r_{e,i})(x_j - r_{e,j}). \quad (4.13)$$

The axis ratio of the structure is determined from the eigenvalues c_1 , c_2 and c_3 of the inertia tensor. After sorting these in descending order, the axis ratio a_1/a_3 of the longest over the smallest axis and a_1/a_2 of longest over medium axis follows from

$$\frac{a_1}{a_3} = \sqrt{\frac{c_1}{c_3}}; \quad \frac{a_1}{a_2} = \sqrt{\frac{c_1}{c_2}}. \quad (4.14)$$

The results of our shape analysis are shown in Fig. 4.19. From left to right, the three frames present the axis ratio of the longest over the smallest axis, a_a/a_3 , for walls, filaments and clusters. The open circles represent the shape of the particle distribution, the triangles

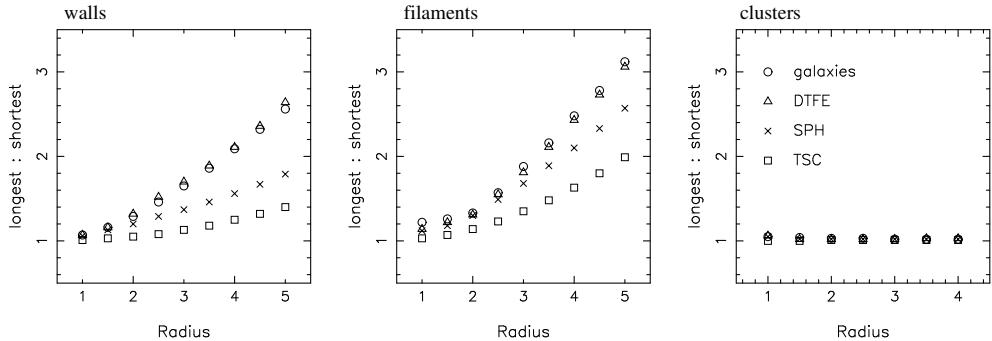


Figure 4.19 — Anisotropy measurements for the Voronoi models. Plotted is the longest-to-shortest axis ratio of the inertia tensor inside a series of concentric spheres centered on a characteristic structure as a function of the radius of the sphere. The radius is given in units of the standard deviation (σ) of the corresponding Gaussian density profiles. The left-hand frame corresponds to the Voronoi wall model, the central frame to the Voronoi filament model and the right-hand frame to the Voronoi cluster model. In each frame the results are shown for the TSC, SPH and DTFE reconstructions, as well as for the galaxy distribution. The meaning of the symbols is depicted in the right-hand frame.

the shape found in the equivalent DTFE density field, while crosses and squares stand for the findings of SPH and TSC.

For the case of the Voronoi cluster models we find a consistent result of near sphericity, $a_1/a_3 = 1$, for both particle distribution as well as for the all three density field reconstructions we find the consistent result of near sphericity, $a_1/a_3 = 1$, in the case of Voronoi cluster models (right-hand frame). In particular the clusters in the DTFE and SPH density fields appear to agree almost perfectly with those in the underlying particle distribution, with the TSC ones deviating somewhat in shape, in particular at small radii.

In the left-hand and central frames of Fig. 4.19 we see that the intrinsic shapes of the walls and filaments become more pronounced as the radius R increases. The uniform increase of the axis ratio a_1/a_3 with R is a reflection of the influence of the intrinsic width of the walls and filaments on the measured shape. For small radii the mass distribution around the center of one of these features is largely confined to the interior of the wall or filament and thus near-isotropic. As the radius R increases in value, the intrinsic shape of these features comes to the fore, resulting in a revealing function of shape as function of R .

The findings of our analysis are remarkably strong and unequivocal: over the complete range of radii we find a striking agreement between the DTFE and the corresponding particle distribution. SPH reveals systematic and substantial differences in that they are tend to be more spherical than the particle distribution, in particular for the strongly anisotropic distributions of the walls and filaments. In turn, the SPH shapes are substantially better than those obtained from the TSC reconstructions. The rigidity of the gridbased TSC density field reconstructions renders them the worst descriptions of the anisotropy of the local matter distribution. For the SPH density fields we still do find a reasonable agreement with the shape of the particle distribution for radii $R_2 R_c$. These differences rapidly increase as radius R increases. The spherical SPH kernel apparently promotes a good shape reproduction in the near-isotropic interior of both walls and filaments but brings about a tendency to sphericalize

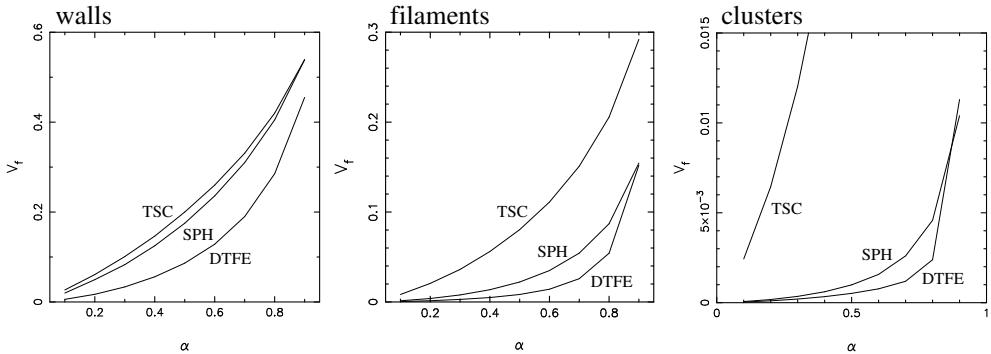


Figure 4.20 — Volume filling fraction as a function of enclosed mass fraction. Plotted is the volume fraction V_f of space with a density higher than a threshold density. This threshold density is varied such that the enclosed region of space contains a mass fraction α . The left-hand frame corresponds to the Voronoi wall model, the central frame to the Voronoi filament model and the right-hand frame to the Voronoi cluster model. In each frame the results are shown for the TSC, SPH and DTFE reconstructions.

the true shape once it gets to radii where the intrinsic shape reveals itself. Also notice that the discrepancies of the SPH shapes with the shape of the particle distribution appear to be considerably smaller in the case of the filamentary model (central frame) than that for the wall model.

These results show that the DTFE is indeed capable of an impressively accurate description of the shape of walls, filaments and clusters. The test discussed in this session appears to represent a striking confirmation of the visual impressions discussed before in the context of Fig. 4.16 and Fig. 4.17.

4.8 Voronoi clustering models: topology and volume occupation

The density field reconstructions of the Voronoi models (Figs. 4.16, 4.17 and 4.18) do not only differ in their local but also in their global character.

One particular manifestation of this aspect concerns the volume occupied by features in the spatial mass distribution. The same features are expected to occupy different volumes within the different field reconstructions. A particular filament occurs as a substantially wider feature in the TSC reconstruction than in the SPH reconstruction, and the same for the latter with respect to the DTFE reconstruction (see Fig. 4.17).

We compute and evaluate the volume filling fraction of features in the reconstructed TSC, SPH and DTFE density fields in order to address and quantify their ability to trace the topology of the cosmic mass distribution. Instead of identifying specific individual features within the web-like matter distribution, we chose to define structures by means of enclosing density contours. The corresponding density threshold ρ_c is parameterized by the fraction α of the total mass enclosed within these density contours, i.e. the volume of space for which $\rho > \rho_c$.

A reasonable test for the pattern tracing ability of the TSC, SPH and DTFE methods is that of the dependence of the volume fraction $V_f(\alpha)$ on the corresponding mass fraction α . We have determined the volume occupation $V_f(\alpha)$ for the Voronoi wall model, Voronoi filament

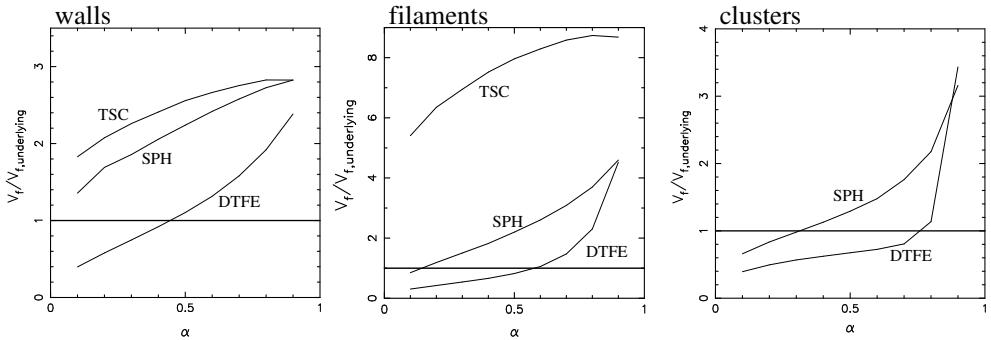


Figure 4.21 — Same as Fig. 4.20, but here the volume filling fraction has been normalized by the volume filling fraction of the underlying distribution, $V_{f,underlying}$ ($V_{f,gal}$ in the text). For guidance the solid horizontal line $V_f/V_{f,gal} = 1$ indicates the location of a density field which would perfectly trace the galaxy distribution.

model and the Voronoi cluster model. The results are plotted in Fig. 4.20.

In order to obtain an objective measure to calibrate the quality of the recovered volumes we need to compare the obtained $V_f(\alpha)$ curves with that of the corresponding volume fractions for the generating Voronoi galaxy distribution itself. In order to compute the corresponding ‘raw’ density field we need to know the (surface) densities for each of the walls, filaments and clusters, as well as the corresponding density profiles. In our Voronoi element models the walls, filaments and vertices have a Gaussian density profile. In the case of a Voronoi wall k , for instance, the density $\rho(\mathbf{x})$ around the wall runs as

$$\rho_k(\mathbf{x}) = \frac{\sigma_k}{R_w \sqrt{2\pi}} e^{-z^2/2R_w^2}. \quad (4.15)$$

In this equation σ_k is the surface density of the wall. Within each wall (and filament) the surface density is uniform, while the value of the surface density varies from wall to wall (and similarly for filaments) and is largely dependent on the size of the Voronoi cell to which it belongs. Appendix 4.C describes the calculation of the volume fraction $V_{f,gal}$ for the Voronoi wall model.

Fig. 4.21 depicts the same relation as in the corresponding top row, be it for the normalized volume fraction $V_f/V_{f,gal}$. For an ideally perfect density reconstruction, $V_f/V_{f,gal} = 1$.

4.8.1 Density field volume occupation: results

The $V_f(\alpha)$ in Fig. 4.20 confirms the visual impression of e.g. Fig. 4.16 that the DTFE reconstruction narrowly traces the density field features, possibly even too tight, an immediate consequence of the high resolution and shape adaptivity of the DTFE kernel. Over the whole α range the SPH density field occupies more volume than its equivalents in the DTFE field. We do see a trend of SPH more closely approaching the volume occupation of the DTFE density field in the case of clusters and filaments, while it appears to deviate more substantially for the Voronoi wall models.

Neither the DTFE, SPH or TSC procedures yield volume filling fractions which reproduce that of the underlying galaxy distribution over the whole range of enclosed mass fractions.

Instead, we notice that for all methods the (relative) volume filling fraction is increasing with the mass fraction α . In other words, as we incorporate features of lower density the density field reconstructions occupy an ever larger than the corresponding share of the cosmic volume.

An interesting observation is that for low mass fraction α , ie. for the highest density features, the DTFE density field occupies less volume than that of the corresponding galaxy distribution (lower frames Fig. 4.20). As the density threshold lowers and the enclosed mass fraction α rises we observe that the DTFE reconstruction occupies not only an expected larger fraction of space but also a fraction which grows relatively to the space occupied by the underlying galaxy population. It reaches the same volume occupation for $\alpha \sim 0.5$ for the Voronoi wall model, $\alpha \sim 0.6$ in the Voronoi filament model and $\alpha \sim 0.8$ in the Voronoi cluster model. Towards very high values of α we notice a steep rise of the DTFE volume occupation. This is to a large extent an artificial effect and reflects the presence the for DTFE characteristic extended low density triangular/tetrahedral wings in low density regions (see e.g. Figs. 4.5 and 4.4 for examples). The observed trend of a rising relative DTFE volume occupation, starting from less than representative for small α towards a volume larger than representative for high α , can be understood on the basis of intrinsic Voronoi tessellation properties, in particular the distribution function of Voronoi cell volumes in combination with the fact that the underlying density field is sampled by a Poisson point distribution. In appendix 4.D we provide some details with respect to a homogeneous point distribution. In a homogeneous point distribution, the 10% smallest cells correspond to a volume fraction $V_{10\%} \sim 4\%$, only 40% of the expected volume. For a Voronoi sampled density field we thus should expect the smallest cells, with the highest density estimates, to correspond to a smaller than representative volume fraction. Evidently, the reverse is true for the largest Voronoi cells. An inhomogeneous particle distribution offers a more challenging situation, yet will follow a similar behaviour.

The SPH density field, on the other hand, occupies more space than the galaxy distribution over the whole range of α . This reflects the fact that the scale of the SPH kernel, defined by the 40th nearest neighbour, tends to produce somewhat inflated volumes scaling proportionally to the local number density of galaxies. The latter explains the rising trend of $V_f/V_{f,gal}(\alpha)$.

For the wall, filament as well as the cluster models the TSC reconstruction involves the largest volume occupation. The TSC volume volume occupation differs substantially from that in the SPH and DTFE density field reconstructions. The relative deviation is considerably larger for the cluster and filament models than for the wall models. This is emphasized by the normalized volume fractions $V_f/V_{f,\text{underlying}}$ in the bottom row of fig 4.20. Proceeding from a wall-dominated pattern to a filament-dominated pattern and finally to a cluster-dominated distribution we see that TSC volume filling fraction is roughly 2 – 3 times that of the underlying galaxy distribution in the case of the wall models and no less than 25 – 40 times larger in the case of the cluster model (well beyond the scale of the figure). This is a telling manifestation of the rigidity of TSC, with the scale and shape of the TSC grid kernels totally insensitive to the size and anisotropy of the underlying density features. Walls on a scale comparable to that of the TSC grid resolution will end up with densities which may be somewhat lower than in reality. A considerably worse TSC representation is obtained when the dimension of a feature is substantially smaller than the gridsize in two or three dimensions. Filaments and compact objects such as clusters are illustrations of the latter.

4.9 Summary and discussion

The DTFE – Delaunay Tessellation Field Estimator – procedure is a self-adaptive filtering and interpolation scheme which does not make use of any artificial, user-specified filtering. The DTFE interpolation method was introduced for rendering fully volume-covering and volume-weighted physical fields from a discrete set of sampled field values. The DTFE is based on the use of the Delaunay tessellation of a given spatial point distribution to form the basis of a fully self-adaptive filter for discretely sampled fields. It is a natural and local technique which reconstructs and interpolates density, and other physical fields such as the corresponding velocity field, from a discrete point distribution to a fully volume-covering continuous field. A qualitative assessment (see e.g. Fig. 4.1) demonstrates the ability of the DTFE to successfully reproduce those aspects of an underlying density and velocity field which have been mentioned as essential aspects of the non-linear web-like cosmic matter distribution of the Megaparsec universe:

- The hierarchically structured matter distribution is resolved to the smallest possible resolution scale set by the particle number density;
- The DTFE retains the morphology cq. shape of the features and patterns in the matter distribution. The characteristic anisotropic filamentary and planar features of the cosmic web are fully reproduced in the continuous DTFE density field;
- The near-empty voids in the spatial matter distribution are reproduced optimally. Both their flat internal density distribution as well as their sharp outline and boundary are recovered in detail through the interpolation characteristics of the DTFE algorithm as well as by its tendency to suppress shot-noise.

The DTFE exploits two particular properties of Delaunay tessellations. It is straightforward to appreciate that on the basis of its definitions the Delaunay tessellation fully adapts to the local point distribution. This adaptive nature translates into volumes of Delaunay tetrahedra scaling inversely with the local density. The DTFE technique uses this property by producing a zeroth-order estimate of the local density at a sample point on the basis of the normalized inverse volume of the contiguous Voronoi cell of that point, which is the union of its surrounding Delaunay tetrahedra. Subsequently, it exploits the adaptive and minimum triangulation properties of Delaunay tessellation by using the Delaunay tetrahedra as adaptive multi-dimensional spatial interpolation intervals. In its interpolation characteristics, the DTFE represents a first-order version of the natural neighbour method, a smooth and local spatial interpolation technique. Natural neighbor interpolation is the most general and robust method of interpolation available to date. The resulting function is continuous everywhere within the convex hull of the data, and has a continuous slope everywhere except at the locations of the data themselves. The theoretical basis of the natural neighbour method was developed by experts in the field of computational geometry (Sibson 1981, Watson 1992). For three-dimensional samples with large number of points, akin to those found in large cosmological computer simulations, the more complex geometric operations involved in the pure nn-neighbour interpolation still represent a computationally challenging task. The first-order DTFE technique represents a viable and applicable alternative exploiting the same geometrical and adaptive properties of the higher order nn-neighbour methods while allowing the analysis of truly large data sets.

In this study we have tested the performance of the DTFE in the case of two different sets of galaxy distribution models, each forming a template with respect to one of two key properties of the large scale galaxy distribution:

- three self-similar Soneira-Peebles models to test distributions with a large range in densities and spatial scales;
- three Voronoi clustering models (wall, filament and cluster) to test complex cellular and/or web-like geometry of galaxy distribution.

The study consists of a comparison and confrontation with the performance of the more conventional TSC and SPH smoothing-interpolation procedures for the same model distributions. The TSC procedure is rigid with respect to the spatial scale as well as the shape of the mass distribution. SPH kernels are adaptive with respect to the local density of points but lack sensitivity to the geometry of the mass distribution.

Of the three methods the DTFE is the one method producing density field maps of the Soneira-Peebles models with the highest spatial resolution. Particularly impressive is its ability to resolve the self-similar scaling of the fractal-like Soneira-Peebles models. Moreover, only the DTFE manages to recover the scaling indices of the Soneira-Peebles density distribution. Both the TSC and SPH fields fail completely in reproducing the proper scaling properties. While TSC does not produce any scaling at all, SPH does manage to reproduce scaling of the density field over a wide range of spatial scales. However, it fails fully in recovering the proper scaling indices. The DTFE is also the only procedure whose density field has an autocorrelation function which agrees completely, down to the smalles scales, with the two-point correlation function of the Soneira-Peebles point distribution.

The limitations of the rigid grid-based TSC interpolations are particularly strong, it fails completely in following the structure of the Soneira-Peebles models on scales smaller than the TSC grid-size. SPH performs considerably better and over a much wider spatial range, instigated by its adaptivity to the spatial resolution of the model distributions. Only at the highest levels of resolution the imprint of the spherical SPH kernel reveals itself as a limiting factor in the form of artificial spherical wings around isolated point concentrations. Also, SPH does not manage to resolve structures consisting of less points than that used for the definition of the scale of the SPH kernel. In comparison, the DTFE needs less neighbours to define its kernel scale. Even the DTFE does reveal its artefacts and limitations in the case of the extreme geometry of the fractal-like pure Soneira-Peebles model. As a result of the volume-filling nature of Delaunay tessellations and the interpolation definition of the DTFE, the method always seeks to interpolate through empty regions. It cannot recover regions of zero density and instead will reveal artificial triangular wings of (very) low density.

A Voronoi wall model, a Voronoi filament model and a Voronoi cluster model are invoked to test the ability of the DTFE to recover the complex web-like geometry of the large scale galaxy and mass distribution. In these models the Voronoi tessellation of a particular nucleus distribution is used as the skeleton of the modelled galaxy distribution: galaxies are distributed around the walls, edges or vertices of the tessellation by means of a projection of a random distribution of points onto these geometrical Voronoi elements. For all three models the spatial web-like pattern outlined by the galaxies can be recognized in the DTFE density fields as clearly and thinly defined features. The higher spatial and shape resolution of the DTFE is

particularly superior in tracing the filamentary features. While it is also better at outlining the sheets in the galaxy distribution of the Voronoi wall model the contrast with respect to the TSC and SPH reconstructions is less pronounced, due to the extended nature of the walls. The limited spatial resolution and rigid gridlike nature of TSC renders it rather unsuitable for an analysis of the web-like features of the galaxy distribution: it does not manage to recover the anisotropy of the matter distribution on scales smaller than the grid resolution. Overall, to a considerable extent the DTFE and SPH algorithms do agree in the web-like pattern they trace. The differences between the DTFE and SPH do in particularly reflect the SPH kernel: the somewhat lower resolution tends to produce iso-density contours which are somewhat puffed up with respect to the equivalent DTFE contours while involving a more roundish morphology as a result of the spherical kernel. The spherical shape of clusters near the vertices of the Voronoi web proves to be particularly apt to the SPH scheme, while the sensitivity of the DTFE kernel to the local point density occasionally tends to exaggerate in reproducing overly compact high density peaks.

Amongst many possible quantitative tests, two were singled out as representative for assessing the ability of TSC, SPH and DTFE in recovering the morphological and topological properties of web-like patterns. The results have been compared with that of the raw particle distribution. The shape of individual Voronoi elements has been evaluated on the basis of the inertia tensor at particular locations. Also, the volume filling factor of iso-density contours in the TSC, SPH and DTFE density fields has been measured and studied as a function of the equivalent enclosed mass. The DTFE is outstanding in recovering the correct shapes of walls and filaments, considerably better than TSC and SPH. SPH does not manage to recover the appropriate shape of walls and filaments, but agrees with the DTFE and the particle distribution itself in recovering the spherical shape of clusters. The quality of SPH improves going from walls to filaments to clusters. The impression provided by the volume filling factor is less unequivocal. The visual impression of the DTFE iso-contours enclosing significantly more pronounced and compact regions is confirmed. Interestingly, at the highest density levels the DTFE contours appear to occupy less space than the original point distribution, while the occupied volume gradually expands with respect to the point distribution as more mass is enclosed. The TSC iso-density contours always enclose the largest volume, a situation which is relatively worst for the Voronoi cluster model. Although the SPH iso-density contours are always larger than the comparable DTFE ones, their behaviour is more alike than the TSC contours in the case of the filament and cluster models.

The limitations of each of the procedures that the rigid and inflexible grid of the TSC method is completely inappropriate for recovering web-like features in the cosmic matter distribution. The SPH procedures fares considerably better than TSC, be it its spherical smoothing kernel does introduce some artefacts and deficiencies when it gets to resolving the finest features. Determined on the basis of a fixed number of neighbours, SPH tends to smear out the smalles structures, particularly for anisotropically shaped ones, and results in features which occupy a significantly larger volume than the corresponding galaxy distribution. Also the DTFE patterns do contain some artefacts, of which the triangular imprint of its smoothing kernel is the most pronounced one.

One key aspect of the Megaparsec matter distribution has not been adressed in this study: the presence of voids and empty regions. Truely empty regions are a major obstacle for the DTFE. Its interpolation based formalism will assure that any empty regions gets reproduced

as a low density region. In reality, however, voids in the cosmic matter distribution will not be completely empty. The artefacts of SPH and the DTFE will therefore be less important. A forthcoming publication (Platen & van de Weygaert 2006), will describe and introduce a DTFE based procedure to detect voids and dissect their substructure (see Sheth & van de Weygaert 2004).

On the basis of the results presented in this study we may conclude that the DTFE is a particularly promising and versatile method for analyzing patterns in a spatial point distribution. In Chapters 6 and 7 we will apply the DTFE to the outcome of large-scale N -body simulations of cosmic structure formation as well as to real observationally obtained patterns such as those in the 2dFGRS and SDSS galaxy distributions. While here we have focussed on the density/intensity field of which a particle distribution is supposed to be reflection, we have also expanded our analysis to cosmic velocity fields (see Schaap & van de Weygaert 2003, Romano-Díaz 2004, Chapter 6). This, in turn, enables the study of the dynamics of Megaparsec features. Applications to any other discrete and inhomogeneously sampled field are self-evident.

While the applications and extensions of the DTFE are numerous we wish to emphasize two important ones. The continuous density, velocity and other fields form the basis for further more specific ‘post-processing’ analysis procedures. One application involves the identification of features in a density field. These may involve the detection of clusters, filaments or voids in the distribution of clusters. The watershed void detection algorithm (Platen & van de Weygaert 2006) has been mentioned before. The multiscale morphology filter (Aragón-Calvo et al. 2006) is tuned towards the identification of individual web components such as sheets and filaments. They have used this filter towards settling issues of dark halo properties as a function of cosmic environment.

A second class of applications seeks to exploit the tremendous potential of Delaunay and Voronoi tessellations as adaptive multi-dimensional grids in the context of system evolution simulations such as N -body simulations or hydrodynamical codes. One particular application is the use of the DTFE kernel as smoothing kernel. Usually N -body or hydrodynamical simulation routines invoke conventional density estimation routines, such as grid- or SPH-based kernels. In Chapter 5 we demonstrate the improvement embodied by the DTFE. The excellent continuity properties of natural neighbour fields may be utilized towards developing an highly desirable moving grid fluid dynamics code. Combining the virtues of Eulerian and Lagrangian formalisms, such a hydrocode would embody a true revolution in dynamical studies of cosmic structure formation. The large range of spatial scales and timescales of the structures and processes of relevance in the formation of galaxies and stars may ultimately demand such an approach. Indeed, attempts towards such implementations have already been introduced in the context of a few, specific, mainly two-dimensional applications (Whitehurst 1995, Braun & Sambridge 1995, Sukumar 1998). Alternative attempts towards the development of moving grid codes, in an astrophysical context, have shown their potential (Gnedin 1995, Pen 1998). Because in such dynamical studies we are dealing with a time sequence of e.g. a gradually evolving particle distribution the success of tessellation-based algorithms will depend on the ability to dynamically upgrade the tessellations without having to update a tessellation completely. Once this issue gets solved, a watershed will be reached in that the related tools will become much more efficient to use in cosmological and other dynamical systems.

References

- Aarseth S.J., Turner E.L., Gott J.R., 1979, ApJ, 228, 664
- Arad I., Dekel A., Klypin A., 2004, MNRAS, 353, 15
- Aragón-Calvo M., van de Weygaert R., Jones B.J.T., van der Hulst J.M., 2006, astro-ph/0610249
- Arbabi-Bidgoli S., Müller V., 2002, MNRAS, 332, 205
- Ascasibar Y., Binney J., 2005, MNRAS, 356, 872
- Babul A., Starkman G.D., 1992, ApJ, 401, 28
- Barrow J.D., Bhavsar S.P., Sonoda D.H., 1985, MNRAS, 216, 17
- Basilakos S., Plionis M., Rowan-Robinson M., 2001, MNRAS, 323, 47
- Bernardeau F., van de Weygaert R., 1996, MNRAS, 279, 693
- Bertschinger E., 1992, in *New Insights into the Universe*, Martinez V.J., Portilla M., Sáez D. (eds.), Springer, Berlin, Germany
- Bertschinger E., 1998, ARAA, 36, 599
- Bond J.R., Cole S., Efstathiou G., Kaiser N., 1991, ApJ, 379, 440
- Bond J.R., Kofman L., Pogosyan D., 1996, Nature, 380, 603
- Bradac M. et al., 2004, A&A, 423, 13
- Braun J., Cambridge M., 1995, Nature, 376, 655
- Brown G.S., 1965, New Zealand Forestry Service Research Notes, 38, 1
- Chiang L.-Y., Coles P., 2000, MNRAS, 311, 809
- Colless M. et al., 2001, MNRAS, 328, 1039
- Colombi S., Pogosyan D., Souradeep T., 2000, Phys. Rev. Lett., 85, 5515
- Cressie N., 1993, *Statistics for Spatial Data, revised edition*, John Wiley & Sons, New York, USA
- Ebeling H., Wiedenmann G., 1993, Phys. Rev. E, 47, 704
- Einasto K., Joeveer M., Saar E., 1980, Nature, 283, 47
- Erdogdu P. et al., 2004, MNRAS, 352, 939
- Fasano G., Franceschini A., 1987, MNRAS, 225, 155
- Freeman K., Bland-Hawthorn J., 2002, ARAA, 40, 487
- Gaztañaga E., 1992, ApJ, 398, 17
- Gnedin N.Y., 1995, ApJS, 97, 231
- Graham M.J., Clowes R.G., Campusano L.E., 1995, MNRAS, 275, 790
- Gregory S.A., Thompson L.A., 1978, ApJ, 222, 784
- Helmi A., White S.D.M., de Zeeuw P.T., Zhao H., 1999, Nature, 402, 53
- Hockney R., Eastwood J., 1988, *Computer Simulations Using Particles*, Taylor & Francis, Bristol, PA, USA
- Hoyle F., Vogeley M., 2002, ApJ, 566, 641
- Icke V., 1984, MNRAS, 206, 1
- Icke V., van de Weygaert R., 1987, A&A, 184, 16
- Jones B.J.T., 2006, *private communication*
- Kiang T., 1966, Z. Astrophys., 64, 433
- Kim R.S.J. et al., 2002, AJ, 123, 20
- Kirshner R.P., Oemler A., Schechter P.L., Shectman S.A., 1981, ApJ, 248, L57
- Kirshner R.P., Oemler A., Schechter P.L., Shectman S.A., 1987, ApJ, 314, 493
- de Lapparent V., Geller M.J., Huchra J.P., 1986, ApJ, 302, L1
- de Lapparent V., Geller M.J., Huchra J.P., 1991, ApJ, 369, 273
- Li G.-L. et al., 2006, astroph/0603557
- Lombardi M., Schneider P., 2001, A&A, 373, 359
- Lombardi M., Schneider P., 2002, A&A, 392, 1153
- Lombardi M., Schneider P., 2003, A&A, 407, 385
- Luo S., Vishniac E., 1995, ApJS, 96, 429
- Marinoni C., Davis M., Newman J.A., Coil A.L., 2002, ApJ, 580, 122

- Martínez V., Jones B.J.T., Dominguez-Tenreiro R., van de Weygaert R., 1990, ApJ, 357, 50
 Martinez V.J., Saar E., 2002, *Statistics of the Galaxy Distribution*, Chapman & Hall/CRC press,
 Boca Raton, USA
 Martínez V.J., 2006, *private communication*
 Mecke K.R., Buchert T., Wagner H., 1994, A&A, 288, 697
 Monaghan J.J., 1992, ARAA, 30, 543
 Monaghan J.J., Latanzio J.C., 1985, Astr. Ap., 149, 135
 Neumann D.M., Lumb D.H., Pratt G.W., Briel U.G., 2003, A&A, 400, 811
 Neyrinck M.C., Hamilton A.J.S., Gnedin N.Y., 2005, MNRAS, 362, 337
 Novikov D., Colombi S., Doré O., 2006, MNRAS, 366, 1201
 Okabe A., Boots B., Sugihara K., Nok Chiu S., 2000, *Spatial Tessellations, Concepts and
 Applications of Voronoi Diagrams*, 2nd edition, Wiley, Chichester, UK
 Ord J.K., 1978, Mathematical Scientist, 3, 23
 Ostriker J.P., 1988, in *Large Scale Structures of the Universe*, Audouze J. et al. (eds.),
 Kluwer, Dordrecht, The Netherlands
 Peacock J.A., 1983, MNRAS, 202, 615
 Peacock J.A., Dodds S.J., 1994, MNRAS, 267, 1020
 Peebles P.J.E., 1980, *The Large-Scale Structure of the Universe*, Princeton University Press,
 Princeton, USA
 Pen U., 1998, ApJS, 115, 19
 Pimbblet K.A., 1995, MNRAS, 358, 256
 Platen E., van de Weygaert R., 2006, *private communication*
 Plionis M., Basilakos S., 2002, MNRAS, 330, 399
 Press W.H., Schechter P., 1974, ApJ, 193, 437
 Press W.H., Teukolsky S.A., Vetterling W.T., Flannery B.P., 1992, *Numerical Recipes*,
 Cambridge University Press, Cambridge, UK
 Ramella M., Boschin W., Fadda D., Nonino M., 2001, A&A, 368, 776
 Ripley B.D., 1981, *Spatial Statistics*, Wiley, New York, USA
 Ripley B.D., 1988, *Statistical Inference for Spatial Processes*, Cambridge University Press,
 Cambridge, UK
 Romano-Díaz, 2004, PhD Thesis, University of Groningen, Groningen, The Netherlands
 Rybicki G.B., Press W.H., 1992, ApJ, 398, 169
 Ryden B.S., Gramann M., 1991, ApJ, 383, 33
 Sahni V., Sathyaprakash B.S., Shandarin S.F., 1989, ApJ, 495, 5
 Sambridge M., Braun J., McQueen H., 1995, Geophys. J. Int., 122, 837
 Schaap W.E., van de Weygaert R., 2003, in *Statistical challenges in modern astronomy III*,
 Babu G.J., Feigelson E.D. (eds.), Springer-Verlag, New York, USA
 Schmalzing J. et al., 1999, ApJ, 526, 568
 Schuecker P., Böhringer H., Reiprich T.H., Feretti L., 1991, A&A, 378, 408
 Seldner M., Siebers B., Groth E.J., Peebles P.J.E., 1977, AJ, 82, 249
 Shandarin S.F., Sheth J.V., Sahni V., 2004, MNRAS, 353, 162
 Sheth R.K., van de Weygaert R., 2004, MNRAS, 350, 517
 Sibson R., 1981, in *Interpreting Multi-Variate Data*, Barnet V. (ed.), Wiley, Chichester, UK
 Soneira R.M., Peebles P.J.E., 1977, ApJ, 211, 1
 Soneira R.M., Peebles P.J.E., 1978, AJ, 83, 7
 Sousbie T. et al., 2006, astroph/0602628
 Springel V., 2005, Nature, 435, 629
 Stoica R.S., Martínez V.J., Mateu J., Saar E., 2005, A&A, 434, 423
 Stoughton C. et al., 2002, AJ, 123, 485
 Sukumar N., 1998, PhD Thesis, Northwestern University, Evanston, IL, USA
 Tegmark M. et al., 2004, ApJ, 606, 702

- Szapudi S., Szalay A., 1998, ApJ, 494, 41
 Watson D.F., 1992, *Contouring: a Guide to the Analysis and Display of Spatial Data*, Pergamon Press, Oxford, UK
 van de Weygaert R., Icke V., 1989, A&A, 213, 1
 van de Weygaert R., 1991, PhD Thesis, Leiden University, Leiden, The Netherlands
 van de Weygaert R., 1994, A&A, 283, 361
 van de Weygaert R., Bertschinger E., 1996, MNRAS, 281, 84
 van de Weygaert R., 2001, in Large Scale Structure in the X-ray Universe, Plionis M. et al. (eds.), Atlantisciences, Paris, France
 van de Weygaert R., 2002, *Froth across the Universe*, in ‘2nd Hellenic Cosmology Meeting’, Plionis M., Cotsakis S. (eds.), Kluwer Academic Publishers, Dordrecht, NL
 White S.D.M., 1976, MNRAS, 177, 717
 Whitehurst R., 1995, MNRAS, 277, 655
 Zaroubi S., Hoffman Y., Fisher K.B., Lahav O., 1995, ApJ, 449, 446
 Zel'dovich Y.B., 1982, Soviet Astron. Lett., 8, 102

4.A Density estimation

To determine/reconstruct a continuous density field $\widehat{\rho}$ from a set of discrete points within the sample points which are considered to be representative for the underlying field we have used three different filtering and interpolation schemes:

- 1 TSC, rigid kernel size, rigid kernel shape;
- 2 SPH, adaptive kernel size, rigid kernel shape;
- 3 DTFE, fully adaptive to spatial point process.

4.A.1 TSC density field reconstruction procedure

The TSC method (Hockney & Eastwood 1981) interpolates an irregularly sampled field to a fixed and regular grid using the Triangular Shaped Cloud kernel. The TSC smoothing kernel distributes the mass corresponding to each point over the fixed grid points according to a weighting function W ,

$$\widehat{\rho}\left(\frac{\mathbf{n}}{N_G}\right) = \frac{N_G^3}{N} \sum_{i=1}^N m_i W\left(\mathbf{x}_i - \frac{\mathbf{n}}{N_G}\right). \quad (4.16)$$

Here $\mathbf{n} = (n_x, n_y, \dots, n_{n_M})$ denotes the grid cell, N_G is the number of cells of the grid in each dimension, N is the number of sampling points and m_i is the mass of sampling point i . The units of the box have been chosen such that its volume is unity.

The TSC weight/kernel function W is a product of one-dimensional weight functions w , specified by

$$W(\mathbf{x}) = \prod_{i=1}^M w(x_i), \quad (4.17)$$

where M is the number of spatial dimensions. The one-dimensional weight functions w are given by

$$w(x_i) = \begin{cases} \frac{3}{4} - N_G^2 \delta x_i^2, & N_G |\delta x_i| \leq \frac{1}{2}; \\ \frac{1}{2}(\frac{3}{2} - N_G |\delta x_i|)^2, & \frac{1}{2} \leq N_G |\delta x_i| \leq \frac{3}{2}. \end{cases} \quad (4.18)$$

The size of the grid may be arbitrarily set according to the specifications of the user. A reasonable conservative procedure is to take a grid cell density in accordance with the number of points, with each grid cell containing one point. With $n = N/V$ the number density of points in the box it implies a rigid TSC resolution in the order of $n^{-1/M}$.

4.A.2 SPH density field reconstruction procedure

In the SPH procedure (see Monaghan 1992 and references therein) the mass of each point is distributed according to a smoothing function W whose size adapts to the local value of the density of points. This leads to an expression of the form

$$\widehat{\rho}(\mathbf{x}) = \sum_{i=1}^N m_i W(\mathbf{x} - \mathbf{x}_i; h). \quad (4.19)$$

Usually the smoothing kernel W is spherically symmetric, while the kernel scale h adapts itself to the local density of points. The smoothing scale h is chosen such that the smoothing volume contains M points. By fixing the number M the resolution is set by the local point density. This renders the method to be adaptive in its spatial resolution, yet rigid with respect to the shape of the point distribution.

We chose the smoothing kernel W to be the following spline kernel (see e.g. Monaghan & Lattanzio 1985),

$$W(r, h) = \frac{10}{7\pi h^2} \begin{cases} 1 - \frac{3}{2}(\frac{r}{h})^2 + \frac{3}{4}(\frac{r}{h})^3, & 0 \leq \frac{r}{h} \leq 1; \\ \frac{1}{4} \left[2 - (\frac{r}{h}) \right]^3, & 1 < \frac{r}{h} \leq 2; \\ 0, & \frac{r}{h} > 2. \end{cases} \quad (4.20)$$

In our study we have chosen h such that the effective smoothing volume contains 16 points for the two-dimensional Soneira-Peebles point sets (see section 4.2) and 64 for the three-dimensional Voronoi clustering model datasets (see section 4.5).

4.A.3 DTFE density field reconstruction procedure

The density $\widehat{\rho}$ is evaluated at the locations of the sample points on the basis of the $V(\mathcal{W}_i)$ of the contiguous Voronoi cell of each point. The contiguous Voronoi cell of a sample point is the union of all Delaunay tetrahedra (or triangles in two dimensions) of which the point forms one of the four vertices. The density $\widehat{\rho}(\mathbf{x}_i)$ at the location of point i is given by

$$\widehat{\rho}(\mathbf{x}_i) = (1 + D) \frac{m_i}{V(\mathcal{W}_i)}. \quad (4.21)$$

Subsequently, the DTFE defines the density field $\widehat{\rho}$ by linear interpolation over the Delaunay tessellation of the point distribution. Once the (linear) gradient $\widehat{\nabla f}|_m$ inside the Delaunay tetrahedron m has been determined from the $(1 + D)$ field values $\widehat{\rho}_j$ at the sample points constituting its vertices, it is straightforward to determine the DTFE density field value $\widehat{\rho}(\mathbf{x})$ for each location \mathbf{x} within the tetrahedron. By means of straightforward linear interpolation within the Delaunay tetrahedron in which \mathbf{x} is located we find

$$\widehat{f}(\mathbf{x}) = \widehat{f}(\mathbf{x}_i) + \widehat{\nabla f}|_m \cdot (\mathbf{x} - \mathbf{x}_i). \quad (4.22)$$

In two dimensions the contiguous Voronoi cell is defined by on average 7 points: the point itself and its natural neighbours. In two dimensions each point has on average 6 natural neighbours, irrespective of the spatial point distribution (see e.g. Icke & van de Weygaert 1987), and also belongs to 6 Delaunay triangles. In three dimensions, the number of natural neighbours varies somewhat as a function of the underlying point distribution. For a Poisson distribution, it is on average ~ 13.4 (see e.g. van de Weygaert 1994). Unlike for the two-dimensional case the number of Delaunay tetrahedra to which a point belongs is rather different from the number of natural neighbours, and is in ~ 27.07 for a Poisson distribution.

Not only the size, but also the shape of the Delaunay simplices is fully determined by the spatial point distribution. As a result the resolution of the DTFE procedure depends on both the density and geometry of the local point distribution. Not only does the DTFE kernel embody an optimal spatial resolution, it also involves a high level of adaptivity to the local geometry of the point distribution.

4.B Voronoi clustering model

Each Voronoi clustering model is defined by a distribution of M cell nuclei and N initially randomly distributed model galaxies. The set of M nuclei or *expansion centres* within the volume V are the ‘expansion centres’ driving an evolving matter distribution. The nuclei have locations \mathbf{y}_m ($m = 1, \dots, M$). Within the same volume N model galaxies are distributed, usually with $N \gg M$. The initial spatial distribution of these N galaxies is purely random within the confines of the sampling volume V , their initial locations $\mathbf{x}_n(0)$ ($n = 1, \dots, N$) defined by a homogeneous Poisson process.

The essential first step of the Voronoi clustering model formalism is to determine for each galaxy n the Voronoi cell \mathcal{V}_α in which it is initially located. This involves the trivial task of finding the nucleus j_α which is closest to the initial position $\mathbf{x}_n(0)$.

Dependent on how far the galaxy is moved away from its initial location $\mathbf{x}_n(0)$, the galaxies’ path $\mathbf{x}_n(t)$ consists of cell displacement \mathbf{s}_α directed radially outward from the expansion centre j_α , a wall displacement $\mathbf{s}_{\alpha\beta}(t)$ within the Voronoi wall $\Sigma_{\alpha\beta}$ defined by the nucleus j_α and its natural neighbour j_β , and an edge displacement $\mathbf{s}_{\alpha\beta\gamma}(t)$ along the Voronoi edge $\Lambda_{\alpha\beta\gamma}$, defined by j_α and its natural neighbours j_β and j_γ (see Fig. 4.22),

$$\begin{aligned} \mathbf{x}_n(t) &= \mathbf{y}_\alpha + \mathbf{s}_{n\alpha}(t) + \mathbf{s}_{n\alpha\beta}(t) + \mathbf{s}_{n\alpha\beta\gamma}(t) \\ &= \mathbf{y}_\alpha + s_{n\alpha}(t)\hat{\mathbf{e}}_{n\alpha} + s_{n\alpha\beta}(t)\hat{\mathbf{e}}_{n\alpha\beta} + s_{n\alpha\beta\gamma}(t)\hat{\mathbf{e}}_{n\alpha\beta\gamma}, \end{aligned} \quad (4.23)$$

in which $\hat{\mathbf{e}}_{n\alpha}$ is the unity vector along the path of the galaxy within the interior of Voronoi cell \mathcal{V}_α , $\hat{\mathbf{e}}_{n\alpha\beta}$ the unity vector along the galaxies’ path within the Voronoi wall $\Sigma_{\alpha\beta}$ and $\hat{\mathbf{e}}_{n\alpha\beta\gamma}$ the unity vector directed along the Voronoi edge $\Lambda_{\alpha\beta\gamma}$.

The identity of the nuclei j_α , j_β , j_γ and j_δ , and thus the identity of the cell \mathcal{V}_α , the wall $\Sigma_{\alpha\beta}$, the edge $\Lambda_{\alpha\beta\gamma}$ and the vertex $\Xi_{\alpha\beta\gamma\delta}$, depends on the initial location $\mathbf{x}_n(0)$ of the galaxy, the position \mathbf{y}_α of its closest nucleus and the definition of the galaxies’ path within the Voronoi skeleton. As for the latter, our procedure is based on the radial path

$$\mathbf{r}_{n\alpha}(t) \equiv \mathbf{y}_\alpha + R_n(t)\hat{\mathbf{e}}_{n\alpha} \quad (4.24)$$

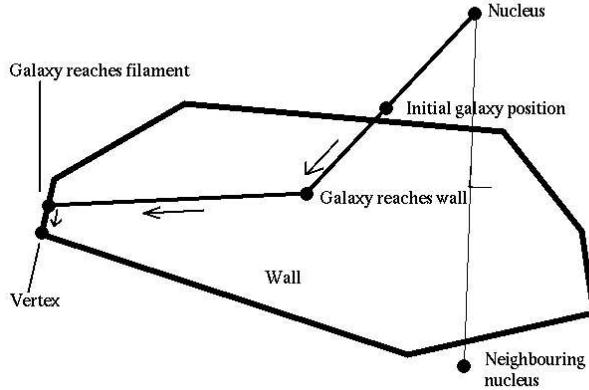


Figure 4.22 — Schematic illustration of the Voronoi kinematic model. Courtesy: Jacco Dankers.

which defines the line starting at \mathbf{y}_α and pointing radially outward from j_α along the direction of unity vector $\hat{\mathbf{e}}_{n\alpha}$,

$$\hat{\mathbf{e}}_{n\alpha} \equiv \frac{\mathbf{x}_n(0) - \mathbf{y}_\alpha}{|\mathbf{x}_n(0) - \mathbf{y}_\alpha|}, \quad (4.25)$$

defined by the initial location of galaxy n and \mathbf{y}_α . The wall $\Sigma_{\alpha\beta}$ is the wall in the Voronoi skeleton which yields the nearest intersection point along the line $\mathbf{r}_n(t)$. Following the intersection, the formalism involves the projection $r_{n\alpha\beta}$ of the radial path $r_{n\alpha}(t)$ onto the wall $\Sigma_{\alpha\beta}$. The boundary of the wall is defined by Voronoi edges. The galaxy n moves towards the one edge $\Lambda_{\alpha\beta\gamma}$ which defines the nearest intersection with the wall path $r_{n\alpha\beta}$. Finally, the path of the galaxy n towards one of its two vertices of the edge $\Lambda_{\alpha\beta\gamma}$ corresponds to the projection of $r_{n\alpha}(t)$ along the edge. The direction along the edge determines the vertex $\Xi_{\alpha\beta\gamma\delta}$.

4.B.1 Procedure outline

Following the above, the recepy for generating Voronoi model galaxy distributions proceeds in four major stages

- Initial Conditions.
- Path within Voronoi cell \mathcal{V}_α , until intersection Voronoi wall $\Sigma_{\alpha\beta}$.
- Path within Voronoi wall $\Sigma_{\alpha\beta}$, until intersection Voronoi edge $\Lambda_{\alpha\beta\gamma}$.
- Path along Voronoi edge $\Lambda_{\alpha\beta\gamma}$, until arrival at vertex $\Xi_{\alpha\beta\gamma\delta}$.

4.B.2 Initial conditions

The initial conditions for the galaxy distribution are generated as follows:

- Distribution of M nuclei, *expansion centres*, within the simulation volume V . The location of nucleus m is \mathbf{y}_m .

- Generate N model galaxies whose initial locations, $\mathbf{x}_n(0)$ ($n = 1, \dots, N$), are randomly distributed throughout the sample volume V .
- Of each model galaxy n determine the Voronoi cell \mathcal{V}_α in which it is located, ie. determine the closest nucleus j_α .

4.B.3 Path within Voronoi cell

- Move galaxy n from its initial position \mathbf{x}_{n0} along the radial path emanating from its expansion centre j_α , ie. along the direction defined by the unity vector $\hat{\mathbf{e}}_{n\alpha}$,

$$\hat{\mathbf{e}}_{n\alpha} \equiv \frac{\mathbf{x}_{n0} - \mathbf{y}_\alpha}{|\mathbf{x}_{n0} - \mathbf{y}_\alpha|}. \quad (4.26)$$

The path \mathbf{x}_n of galaxy n within the cell \mathcal{V}_α is

$$\mathbf{x}_n(t) = \mathbf{y}_\alpha + s_{n\alpha}(t) \hat{\mathbf{e}}_{n\alpha}, \quad (4.27)$$

where $\mathbf{s}_{n\alpha}(t) = s_{n\alpha}(t) \hat{\mathbf{e}}_{n\alpha}$ is its displacement with respect to the expansion centre j_α .

- Identify the Voronoi wall $\Sigma_{\alpha\beta}$ of Voronoi cell \mathcal{V}_α which will be intersected by the path of galaxy n , ie. determine the identity j_β of the corresponding natural neighbour of j_α . This wall is shared with one of the natural neighbours j_β and delineates the region of space equidistant to its two closest nuclei, j_α and j_β .
- Determine the parameter value

$$s_{n\alpha}(t) = v_n \quad (4.28)$$

for which the radial path $\mathbf{x}_n(t)$ crosses $\Sigma_{\alpha\beta}$, yielding the intersection point

$$\mathbf{x}_{n\alpha\beta} \equiv \mathbf{y}_\alpha + v_n \hat{\mathbf{e}}_{n\alpha}. \quad (4.29)$$

4.B.4 Path within Voronoi wall

Subsequently, the path $\mathbf{x}_n(t)$ of galaxy n is constrained to the wall $\Sigma_{(\alpha\beta)}$ (see Fig. 4.22). The path is a continuation of the radial path within \mathcal{V}_α (Eqn. 4.27), with the displacement perpendicular to the wall set to zero.

- The path within the wall is specified by

$$\begin{aligned} \mathbf{x}_n(t) &= \mathbf{x}_{n\alpha\beta} + s_{n\alpha\beta}(t) \hat{\mathbf{e}}_{n\alpha\beta} \\ &= \mathbf{y}_\alpha + v_n \hat{\mathbf{e}}_{n\alpha} + s_{n\alpha\beta}(t) \hat{\mathbf{e}}_{n\alpha\beta}, \end{aligned} \quad (4.30)$$

where $\hat{\mathbf{e}}_{n\alpha\beta}$ is the unity vector along the direction of the projection of the vector $\hat{\mathbf{e}}_{n\alpha}$ onto the wall $\Sigma_{\alpha\beta}$. Defining the unity vector $\hat{\mathbf{e}}_{\alpha\beta}$ perpendicular to $\Sigma_{\alpha\beta}$ as

$$\hat{\mathbf{e}}_{\alpha\beta} \equiv \frac{\mathbf{y}_\alpha - \mathbf{y}_\beta}{|\mathbf{y}_\alpha - \mathbf{y}_\beta|}, \quad (4.31)$$

the vector $\hat{\mathbf{e}}_{n\alpha\beta}$ is defined by

$$\hat{\mathbf{e}}_{n\alpha\beta} \equiv \frac{\hat{\mathbf{e}}_{n\alpha} - (\hat{\mathbf{e}}_{n\alpha} \cdot \hat{\mathbf{e}}_{\alpha\beta}) \hat{\mathbf{e}}_{\alpha\beta}}{|\hat{\mathbf{e}}_{n\alpha} - (\hat{\mathbf{e}}_{n\alpha} \cdot \hat{\mathbf{e}}_{\alpha\beta}) \hat{\mathbf{e}}_{\alpha\beta}|}. \quad (4.32)$$

- Identify the Voronoi edge $\Lambda_{\alpha\beta\gamma}$, ie. identify the nucleus j_γ with whom the nuclei j_α and j_β share a Voronoi edge. The edge $\Lambda_{\alpha\beta\gamma}$ delineates the region of space equidistant to its three closest nuclei, j_α , j_β and j_γ , marking the boundary between their Voronoi cells and as such is also one of the edges defining the boundary of the polygonal Voronoi wall $\Sigma_{\alpha\beta}$.

- Determine the parameter value

$$s_{n\alpha\beta}(t) = \sigma_n \quad (4.33)$$

for which the path $\mathbf{x}_n(t)$ crosses $\Lambda_{\alpha\beta\gamma}$, yielding the crossing point $\mathbf{x}_{n\alpha\beta\gamma}$,

$$\mathbf{x}_n = \mathbf{x}_{n\alpha\beta} + \sigma_n \hat{\mathbf{e}}_{n\alpha\beta}. \quad (4.34)$$

4.B.5 Path along Voronoi edge

Subsequently, the path $\mathbf{x}_n(t)$ of galaxy n is constrained to proceed along the edge $\Lambda_{\alpha\beta\gamma}$, proceeding towards one of the two vertices at its tip.

- The path of galaxy n along the edge is given by

$$\begin{aligned} \mathbf{x}_n(t) &= \mathbf{x}_{n\alpha\beta\gamma} + s_{n\alpha\beta\gamma}(t) \hat{\mathbf{e}}_{n\alpha\beta\gamma} \\ &= \mathbf{y}_\alpha + v_n \hat{\mathbf{e}}_{n\alpha} + \sigma_n \hat{\mathbf{e}}_{n\alpha\beta} + s_{n\alpha\beta\gamma}(t) \hat{\mathbf{e}}_{n\alpha\beta\gamma}, \end{aligned} \quad (4.35)$$

where $\hat{\mathbf{e}}_{n\alpha\beta\gamma}$ is the unity vector directed along the axis of the edge $\Lambda_{\alpha\beta\gamma}$,

$$\hat{\mathbf{e}}_{n\alpha\beta\gamma} \equiv \frac{(\mathbf{y}_\beta - \mathbf{y}_\alpha) \times (\mathbf{y}_\gamma - \mathbf{y}_\alpha)}{|(\mathbf{y}_\beta - \mathbf{y}_\alpha) \times (\mathbf{y}_\gamma - \mathbf{y}_\alpha)|}. \quad (4.36)$$

4.B.6 Voronoi vertex

- Identify the Voronoi vertex $\Xi_{\alpha\beta\gamma\delta}$, ie. identify the nucleus j_δ with whom the nuclei j_α , j_β and j_γ share a Voronoi vertex. That is, the nucleus j_δ with whom j_α , j_β and j_γ define a Delaunay polyhedron and a corresponding circumscribing sphere which does not contain any of the other $M - 4$ nuclei and of which $\Xi_{\alpha\beta\gamma\delta}$ is the centre.
- The path of galaxy n along the edge $\Lambda_{\alpha\beta\gamma}$ proceeds up to vertex $\Xi_{\alpha\beta\gamma\delta}$, which it reaches when the parameter $s_{n\alpha\beta\gamma}(t)$ attains the value

$$s_{n\alpha\beta\gamma}(t) = \lambda_n, \quad (4.37)$$

such that its coordinates \mathbf{x}_n are those of the vertex $\Xi_{\alpha\beta\gamma\delta}$, i.e.

$$\begin{aligned} \mathbf{x}_n &= \mathbf{x}_{\alpha\beta\gamma\delta} \\ &= \mathbf{x}_{n\alpha\beta\gamma} + \lambda_n \hat{\mathbf{e}}_{n\alpha\beta\gamma} \\ &= \mathbf{y}_\alpha + v_n \hat{\mathbf{e}}_{n\alpha} + \sigma_n \hat{\mathbf{e}}_{n\alpha\beta} + \lambda_n \hat{\mathbf{e}}_{n\alpha\beta\gamma}. \end{aligned} \quad (4.38)$$

4.C Voronoi wall density and mass distribution

In order to estimate the volume occupied by a particular mass fraction of the galaxy distribution we need to compute the ‘raw’ density values in the underlying Voronoi galaxy models. The enclosed mass may then be computed by integrating over the regions of space with a density value larger than a given threshold ρ_c .

With the situation and calculation being quite similar for the Voronoi wall, filament and cluster models we here concentrate on the explicit calculation of the mass volume occupation of the Voronoi wall model.

4.C.1 Single Voronoi wall

In the Voronoi wall model the surface density σ_k of a Voronoi wall Σ_k is uniform, its value solely dependent on the distance h_k of the wall to the nuclei j_α and j_β . These are the nuclei which have Σ_k as one of the polygonal Voronoi walls marking their boundary. This follows from the fact that the volume within the pyramidal volume $V_{k\alpha}$ defined by the nucleus j_α and the wall Σ_k ,

$$V_{k\alpha} = V_{k\beta} = \frac{1}{6} h_k A_k, \quad (4.39)$$

is linearly proportional to the value of h_k and the surface area A_k of the wall (see van de Weygaert 1991, 1994). As the whole volume $V_{k\alpha}$ is projected onto the wall, the surface density of the latter simply follows from

$$\sigma_k = 2 \frac{V_{k\alpha}}{A_k} \propto h_k. \quad (4.40)$$

The factor two is a result of the projection of both volumes $V_{\alpha\beta}$ and $V_{k\beta}$ onto the wall Σ_k .

In the Voronoi wall models used in this study the density distribution around each wall is assumed to follow a Gaussian distribution perpendicular to the wall. The density profile around each wall is given by

$$\rho_k(\mathbf{x}) = \frac{\sigma_k}{R_w \sqrt{2\pi}} e^{-z^2/2R_w^2}, \quad (4.41)$$

in which z is the perpendicular distance of a point to the wall Σ_k . Note that the central wall density $\rho_{k,0} = \rho_k(z=0)$ is given by

$$\rho_k(z=0) = \frac{\sigma_k}{R_w \sqrt{2\pi}}. \quad (4.42)$$

The width of the wall is set by the scale R_w (which was chosen to be 1/400th of the box-size). For reasons of consistency, the wall scale R_w needs to be smaller than the perpendicular Voronoi cell scale h_k ,

$$R_w \ll h_k. \quad (4.43)$$

From this density profile we can immediately infer the perpendicular distance $z_{k,c}$ corresponding to a density contour value ρ_c ,

$$z_{k,c} = R_w \sqrt{2} \sqrt{\ln \frac{\sigma_k}{R_w \sqrt{2\pi} \rho_c}}. \quad (4.44)$$

Evidently the implicit assumption is that the central density $\rho_{k,0} > \rho_c$. The mass $M_k(z)$ in and around Σ_k enclosed within the density contour ρ_c is equal to

$$M_{k,c} = A_k \times \int_{-z_{k,c}}^{z_{k,c}} dz \frac{\sigma_k}{R_w \sqrt{2\pi}} e^{-z^2/2R_w^2}. \quad (4.45)$$

4.C.2 Multiple Voronoi walls

If all Voronoi walls would have the same surface density, the total mass enclosed by the complete network of Voronoi walls would be obtained by simply multiplying the integral in equation 4.45 by the surface area of all Voronoi walls combined.

In reality, the calculation is slightly complicated by the each wall having a different surface density, varying according to the perpendicular distance wall-nucleus (Eqn. 4.39). As a consequence the perpendicular distance $z_{k,c}$ corresponding to the density threshold ρ_c will be different for each wall. To obtain the total mass enclosed within a density contour level ρ_c we therefore need to consider each Voronoi wall individually and add the individual wall mass contributions $M_{k,c}$,

$$M_{total,c} = \sum_{\substack{k \\ \rho_{k,0} > \rho_c}} M_{k,c} \quad (4.46)$$

$$= \sum_{\substack{k \\ \rho_{k,0} > \rho_c}} A_k \times \int_{-z_{k,c}}^{z_{k,c}} dz \frac{\sigma_k}{R_w \sqrt{2\pi}} e^{-z^2/2R_w^2}. \quad (4.47)$$

In the above only walls with a core density $\rho_{k,0} > \rho_C$ may be taken into account.

To find the contour value ρ_c for which a fraction α of the total mass (or total number of galaxies) is enclosed within the contour's interior,

$$M_{total,c} = \alpha N_{gal}, \quad (4.48)$$

we iteratively solve the above integral Eqn. 4.47, taking into account the dependence of the integral limits $z_{k,c}$ for each individual wall on the contour value ρ_c . The volume $V_{f,c}$ enclosed by the mass fraction α can then be computed from the sum

$$V_f = \sum_{\substack{k \\ \rho_{k,0} > \rho_c}} A_k \times 2z_{k,c}. \quad (4.49)$$

For the Voronoi filament and cluster models a similar calculation yields the value of the corresponding volume filling fractions.

4.D Voronoi cell volume distribution

When estimating the density values in a density field by means of a Voronoi tessellation we need to take into account the fact that a particle distribution forms a Poisson sample of the underlying density field. As a result the number density of points varies and at some locations

may be higher than expected while at others it is lower than expected. In regions where the number density is higher the volumes of the Voronoi cells are proportionally smaller. This brings about fluctuations in the DTFE density field estimates, extensively studied and treated in Chapter 8.

Regions which as a result of the Poisson sampling have a larger number density will thus correspond to smaller Voronoi cells. As a result we find an uncharacteristically large amount of mass in the highest density areas, occupying a relatively small fraction of the total volume. Conversely, at locations where the density is smaller than expected a relatively large volume contains a relatively small amount of mass. This results in a larger volume filling fraction for the lowest density regions (with high value for α).

We may obtain a quantitative estimate of this effect by considering the distribution of Voronoi cell sizes in a homogeneous Poissonian point sampling. Recall that in such a distribution the underlying density field is homogeneous. Kiang (1966) has shown that this distribution may be approximated by

$$dp(\tilde{a}) = \frac{c}{\Gamma(c)} (c \tilde{a})^{c-1} e^{-c\tilde{a}} d\tilde{a}, \quad (4.50)$$

with $\tilde{a} = a/\langle a \rangle$ the size of the Voronoi cell in units of the average Voronoi cell size (also see Icke & van de Weygaert 1987). The value of c , a numerical constant, depends on the dimension of space. In three dimensions $c = 6$.

As an example, consider the volume filling fraction of the 10% smallest cells, the ones with the highest density,

$$V_{10\%} = \int_0^{\tilde{a}_{10\%}} dp(\tilde{a}) \tilde{a}. \quad (4.51)$$

In this expression $\tilde{a}_{10\%}$ is the volume of the Voronoi cell at the 10 percentile level: 10% of the Voronoi cells have a volume smaller than $\tilde{a}_{10\%}$. This volume may be found by solving

$$\int_0^{\tilde{a}_{10\%}} dp(\tilde{a}) = 0.10. \quad (4.52)$$

From this we find that $V_{10\%} = 0.042$, ie. the 10% smallest cells occupy 4% of the total volume, only 40% of the expected volume.

5

Density Estimators in Particle Hydrodynamics: DTFE versus SPH[†]

F.I. Pelupessy, W.E. Schaap & R. van de Weygaert

ABSTRACT — We present the results of a study comparing density maps reconstructed by the Delaunay Tessellation Field Estimator (DTFE) and by SPH kernel-based techniques. The density maps are constructed from the outcome of an SPH particle hydrodynamics simulation of a multi-phase interstellar medium. The comparison between the two methods clearly demonstrates the superior performance of the DTFE with respect to conventional SPH methods, in particular at locations where SPH appears to fail. SPH overestimates the density in the outer parts of voids, filaments and sheet-like structures, whereas it underestimates the density in the interior of filaments, sheet-like structures and in high-density clumps. Through the improvement in local density estimates, calculations invoking the DTFE will yield a much better representation of physical processes which depend on density. This will be crucial in the case of feedback processes, which play a major role in galaxy and star formation. The presented results form an encouraging step towards the incorporation of the DTFE in astrophysical hydrocodes. We describe an outline for the construction of a particle hydrodynamics code in which the DTFE replaces kernel-based methods. Further discussion addresses ‘moving grid’ hydrocodes invoking the DTFE and Delaunay tessellations in an attempt to combine the virtues of the Eulerian and Lagrangian approaches.

[†]based on Pelupessi F.I., Schaap W.E., van de Weygaert R., 2003, A&A, 403, 389

5.1 Introduction

Smoothed particle hydrodynamics (SPH) has established itself as the workhorse for a variety of astrophysical fluid-dynamical computations (Lucy 1977, Gingold & Monaghan 1977). In a wide range of astrophysical environments this Lagrangian scheme offers substantial and often crucial advantages over Eulerian, usually grid-based, schemes. Astrophysical applications such as cosmic structure formation and galaxy formation, the dynamics of accretion disks and the formation of stars and planetary systems are examples of its versatility and successful performance (for an enumeration of applications and corresponding references, see e.g. the reviews by Monaghan 1992 and Bertschinger 1998).

A crucial aspect of the SPH procedure concerns the proper estimation of the local density, i.e. the density at the location of the particles which are supposed to represent a fair discrete sampling of the underlying continuous density field. The basic feature of the SPH procedure for density estimation is based on a convolution of the discrete particle distribution with a particular user-specified kernel function W . For a sample of N particles, with masses m_j and locations \mathbf{r}_j , the density $\widehat{\rho}(\mathbf{r}_i)$ at the location \mathbf{r}_i of particle i is given by

$$\widehat{\rho}(\mathbf{r}_i) = \sum_{j=1}^N m_j W(\mathbf{r}_i - \mathbf{r}_j, h_i), \quad (5.1)$$

in which the kernel resolution is determined through the smoothing scale h_i . Notice that generically the scale h_i may be different for each individual particle, and thus may be set to adapt to the local particle density. Usually the functional dependence of the kernel W is chosen to be spherically symmetric, so that it is a function of $|\mathbf{r}_i - \mathbf{r}_j|$ only.

The evolution of the physical system under consideration is fully determined by the movement of the discrete particles. Given a properly defined density estimation procedure, the equations of motion for the set of particles are specified through a suitable Lagrangian, if necessary including additional viscous forces (see e.g. Rasio 1999).

In practical implementations, however, the SPH procedure involves a considerable number of artefacts. These stem from the fact that SPH particles represent functional averages over a certain Lagrangian volume. This averaging procedure is further aggravated by the fact that it is based on a rather arbitrary user-specified choice of both the adopted resolution scale(s) h_i and the functional form of the kernel W . Such a description of a physical system in terms of user-defined fuzzy clouds of matter is known to lead to considerable complications in realistic astrophysical circumstances. Often these environments involve fluid flows exhibiting complex spatial patterns and geometries. In particular in configurations characterized by strong gradients in physical characteristics – of which density, pressure and temperature discontinuities in and around shock waves represent the most frequently encountered example – SPH has been hindered by its relative inefficiency in resolving these gradients.

Given the necessity for the user to specify the characteristics and parameter values of the density estimation procedure, the accuracy and adaptability of the resulting SPH implementation hinges on the ability to resolve steep density contrasts and the capacity to adapt itself to the geometry and morphology of the local matter distribution. A considerable improvement with respect to the early SPH implementations, which were based on a uniform smoothing length h , involves the use of adaptive smoothing lengths h_i (Hernquist & Katz 1989), which provides the SPH calculations with a larger dynamic range and higher spatial resolution. The

mass distribution in many (astro)physical systems and circumstances is often characterized by the presence of salient anisotropic patterns, usually identified as filamentary or planar features. To deal with such configurations, additional modifications in a few sophisticated implementations attempted to replace the conventional – and often unrealistic and restrictive – spherically symmetric kernels by ones whose configuration is more akin to the shape of the local mass distribution. The corresponding results do indeed represent a strong argument for the importance of using geometrically adaptive density estimates. A noteworthy example is the introduction of ellipsoidal kernels by Shapiro et al. (1996). Their shapes are stretched in accordance with the local flow. Yet, while evidently being conceptually superior, their practical implementation does constitute a major obstacle and has prevented widespread use. This may be ascribed largely to the rapidly increasing number of degrees of freedom needed to specify and maintain the kernel properties during a simulation.

Even despite their obvious benefits and improvements, these methods are all dependent on the artificial parametrization of the local spatial density distribution in terms of the smoothing kernels. Moreover, the specification of the information on the density distribution in terms of extra non-physical variables, necessary for the definition and evolution of the properties of the smoothing kernels, is often cumbersome to implement and may introduce subtle errors (Hernquist 1993, see however Nelson & Papaloizou 1994 and Springel & Hernquist 2002). In many astrophysical applications this may lead to systematic artefacts in the outcome for the related physical phenomena. Within a cosmological context, for example, the X-ray visibility of clusters of galaxies is sensitively dependent on the value of the local density, setting the intensity of the emitted X-ray emission by the hot intergalactic gas. This will be even more critical in the presence of feedback processes, which for sure will be playing a role when addressing the amount of predicted star formation in simulation studies of galaxy formation.

Here we seek to circumvent the complications induced by the kernel parametrization and propose an alternative to the use of kernels for the quantification of the density within the SPH formalism. This new method, based on the Delaunay Tessellation Field Estimator (DTFE, see Chapters 2 and 3), has been devised to mould and fully adapt itself to the configuration of the particle distribution. Unlike conventional SPH methods, it is able to deal self-consistently and naturally with anisotropies in the matter distribution, even when it concerns caustic-like transitions. In addition, it manages to successfully treat density fields marked by structural features over a vast (dynamic) range of scales.

The DTFE produces density estimates on the basis of the particle distribution, which is supposed to form a discrete spatial sampling of the underlying continuous density field. As a linear multi-dimensional field interpolation algorithm it may be regarded as a first-order version of the natural neighbour algorithm for spatial interpolation (Sibson 1981, see also Okabe et al. 2000). In general, applications of the DTFE to spatial point distributions have demonstrated its success in dealing with the complications of anisotropic geometry and dynamic range (see Chapters 3 and 4). The key ingredient of the DTFE procedure is the Delaunay triangulation, serving as the complete covering of a sample volume by mutually exclusive multi-dimensional linear interpolation intervals.

Delaunay tessellations (Delone 1934; see also Okabe et al. 2000 and references therein) form the natural framework in which to discuss the properties of discrete point sets, and thus also of discrete samplings of continuous fields. Their versatility and significance have been underlined by their widespread application in areas such as computer graphics, geo-

graphical mapping and medical imaging, as well as in a variety of ‘conventional’ grid-based fluid-dynamical computation schemes. The latter may concern their use as a non-regular application-oriented grid covering of physical systems, which represents a prominent procedure in technological applications. More innovating has been their use in Lagrangian ‘moving-grid’ implementations (see Mavriplis 1997 for a review and Whitehurst 1995 for a promising astrophysical application).

It therefore seems a good idea to explore the possibilities of applying the DTFE in the context of a numerical hydrodynamics code. Here, as a first step, we wish to obtain an idea of the performance of a hydrocode involving the use of DTFE estimates with respect to an equivalent code involving regular SPH density estimates. The quality of the new DTFE method with respect to the conventional SPH estimates and their advantages and disadvantages under various circumstances are evaluated by a comparison between the density field which would be yielded by a DTFE processing of the resulting SPH particle distribution and that of the regular SPH procedure itself. In this study we compare the resulting matter distributions in the situation of a representative stochastic multi-phase density field. This allows us to make a comparison between both density estimates in a regime for which an improved method for density estimates would be of great value. We should point out a major drawback of our approach, in that we do not really treat the DTFE density estimate in a self-consistent fashion. Instead of being part of the dynamical equations themselves we only use it as an analysis tool of the produced particle distribution. Nevertheless, it will still show the value of the DTFE in particle gasdynamics and give an indication of what kind of differences may be expected when incorporating in a fully self-consistent manner the DTFE estimate in an hydrocode.

On the basis of our study we elaborate on the potential benefits of a hydrodynamics scheme based on the DTFE. Specifically, we outline how we would set out to develop a complete particle hydrodynamics code whose artificial kernel-based nature is replaced by the more natural and self-adaptive approach of the DTFE. Such a DTFE-based particle hydrodynamics code would form a promising step towards the development of a fully tessellation-based quasi-Eulerian moving-grid hydrodynamical code. Such would yield a major and significant step towards defining a much needed alternative and complement to currently available simulation tools.

5.2 DTFE and SPH density estimates

The methods we use for the SPH and DTFE density estimates have been extensively described elsewhere (Hernquist & Katz 1989, Chapter 3). Here we summarize their main aspects.

5.2.1 SPH density estimate

Amongst the various density recipies employed within available SPH codes, we use the Hernquist & Katz (1989) symmetrized form of Eq. 5.1, using adaptive smoothing lengths,

$$\widehat{\rho}(\mathbf{x}_i) = \frac{1}{2} \sum_j m_j \left\{ W(|\mathbf{r}_i - \mathbf{r}_j|, h_i) + W(|\mathbf{r}_i - \mathbf{r}_j|, h_j) \right\}. \quad (5.2)$$

The smoothing lengths h_i are chosen such that the sum involves the 40 nearest neighbours. For the kernel W we take the conventional spline kernel described by Monaghan (1992). Other variants of the SPH estimate produce comparable results.

5.2.2 DTFE density estimate

The DTFE density estimating procedure consists of three basic steps. Starting from the sample of particle locations, the first step involves the computation of the corresponding Delaunay tessellation. Each Delaunay cell \mathcal{T}_m is the uniquely defined tetrahedron whose four vertices are the sampling particles whose circumscribing sphere does not contain any of the other sampling particles. The Delaunay tessellation is the full covering of space by the complete set of these mutually disjunct tetrahedra. Delaunay tessellations are well known concepts in stochastic and computational geometry (Delone 1934, for further references see e.g. Okabe et al. 2000, Møller 1994 and van de Weygaert 1991).

The second step involves estimating the density at the location of each of the particles in the sample. From the definition of the Delaunay tessellation, it may be evident that there is a close relationship between the volume of a Delaunay tetrahedron and the local density of the generating point process. Evidently, the ‘empty’ cirumscribing spheres corresponding to the Delaunay tetrahedra, and the volumes of the resulting Delaunay tetrahedra, will be smaller as the number density of sample points increases, and vice versa. Following this observation, a proper density estimate $\widehat{\rho}(\mathbf{r}_i)$ at the location \mathbf{r}_i of a sampling point i is obtained by determining the properly calibrated inverse of the volume of the corresponding ‘contiguous’ Voronoi cell \mathcal{W}_i . The latter is defined as the union of all Delaunay tetrahedra $\mathcal{T}_{j,i}$ of which particle i forms one of the four vertices, i.e. $\mathcal{W}_i = \bigcup_j \mathcal{T}_{j,i}$. In general, when a particle i is surrounded by $N_{\mathcal{T},i}$ Delaunay tetrahedra, each with a volume $V(\mathcal{T}_{j,i})$, the volume of the resulting contiguous Voronoi cell is equal to

$$V(\mathcal{W}_i) = \sum_{j=1}^{N_{\mathcal{T},i}} V(\mathcal{T}_{j,i}). \quad (5.3)$$

Note that $N_{\mathcal{T},i}$ is not a constant, but in general may acquire a different value for each point in the sample. Assuming that each particle i has been assigned a mass m_i , the estimated density $\widehat{\rho}(\mathbf{r}_i)$ at the location of particle i is given by (see Chapters 2 and 3)

$$\widehat{\rho}(\mathbf{r}_i) = \frac{4m_i}{V(\mathcal{W}_i)}. \quad (5.4)$$

The factor 4 is a normalization factor, accounting for the four different contiguous Voronoi cells to which each Delaunay tetrahedron is assigned, one for each vertex of a Delaunay tetrahedron.

The third step is the interpolation of the estimated densities $\widehat{\rho}(\mathbf{r}_i)$ over the full sample volume. In this, the DTFE bases itself on the fact that each Delaunay tetrahedron may be considered the natural multi-dimensional equivalent of a linear interpolation interval (see e.g Bernardeau & van de Weygaert 1996). Given the four vertices of a Delaunay tetrahedron with corresponding density estimates $\widehat{\rho}(\mathbf{r}_i)$, the value $\widehat{\rho}(\mathbf{r})$ at any location \mathbf{r} within the tetrahedron can be straightforwardly determined by simple linear interpolation,

$$\widehat{\rho}(\mathbf{r}) = \widehat{\rho}(\mathbf{r}_{i0}) + \left. \widehat{\nabla \rho} \right|_j \cdot (\mathbf{r} - \mathbf{r}_{i0}), \quad (5.5)$$

in which \mathbf{r}_{i0} is the location of one of the four Delaunay vertices i . This is a trivial evaluation once the value of the (linear) density gradient $\left. \widehat{\nabla \rho} \right|_j$ has been estimated. For each Delaunay tetrahedron \mathcal{T}_j this is accomplished by solving the system of three linear equations

corresponding to each of the remaining three Delaunay vertices constituting the Delaunay tetrahedron \mathcal{T}_j . The ‘minimum triangulation’ property of Delaunay tessellations underlying this linear interpolation, minimum in the sense of representing a volume-covering network of optimally compact multi-dimensional tetrahedra, has been a well-known property utilized in a variety of imaging and surface rendering applications such as geographical mapping and various computer imaging algorithms.

5.2.3 Comparison

Comparing the two methods, we see that in the case of SPH the particle ‘size’ and ‘shape’ (i.e. its domain of influence) is determined by some arbitrary kernel $W(r, h_i)$ and a fortuitous choice of the smoothing length h_i (assuming, along with the major share of SPH procedures, a radially symmetric kernel). In the case of the DTFE method the influence region of a particle is fully determined by the sizes and shapes of its surrounding Delaunay tetrahedra $\mathcal{T}_{j,i}$, which are completely determined by the local particle distribution. In other words, in regular SPH the density is determined through the kernel function $W(\mathbf{x})$, while in DTFE it is solely the particle distribution itself setting the estimated values of the density. Contrary to the generic situation for the kernel-dependent methods, there are no extra variables left to be determined. One major additional advantage is that it is therefore not necessary to worry about the evolution of the kernel parameters.

Both methods do display some characteristic artefacts in their density reconstructions (see Fig. 5.1). To a large extent these may be traced back to the implicit assumptions involved in the interpolation procedures, a necessary consequence of the finite amount of information contained in a discrete representation of a continuous field. SPH density fields implicitly contain the imprint of the specified and applied kernel which, as has been discussed before, may seriously impart its resolving power and capacity to trace the true geometry of structures. The DTFE technique, on the other hand, does produce triangular artefacts. At instances conspicuously visible in the DTFE reconstructed density fields, they are the result of the linear interpolation scheme employed for the density estimation at the locations not coinciding with the particle positions. In principle, this may be substantially improved by the use of higher-order interpolation schemes. Such higher-order schemes have indeed been developed, and the ones based on the natural neighbour interpolation prescription of Sibson (1981) have already been successfully applied to two-dimensional problems in the field of geophysics (Sambridge et al. 1995, Braun & Sambridge 1995) and solid state physics (Sukumar 1998).

5.3 Case study: two-phase interstellar medium

For the sake of testing and comparing the SPH and DTFE methods, we assess a snapshot from a simulation of the neutral ISM. The model of the ISM is chosen as an illustration rather than as a realistic model. The ‘simulation’ sample of the ISM consists of H I gas confined in a periodic simulation box with a size $L = 0.6 \text{ kpc}^3$. The initially uniform density of the gas is $n_{\text{H}} = 0.3 \text{ cm}^{-3}$, while its temperature is taken to be $T = 10000 \text{ K}$. No fluctuation spectrum is imposed to set the initial featureless spatial gas distribution. To set the corresponding initial spatial distribution of the $N = 64000$ simulation particles, we start from relaxed initial conditions according to a ‘glass’ distribution (see e.g. White 1994).

The evolution of the gas is solely a consequence of fluid-dynamical and thermodynamical

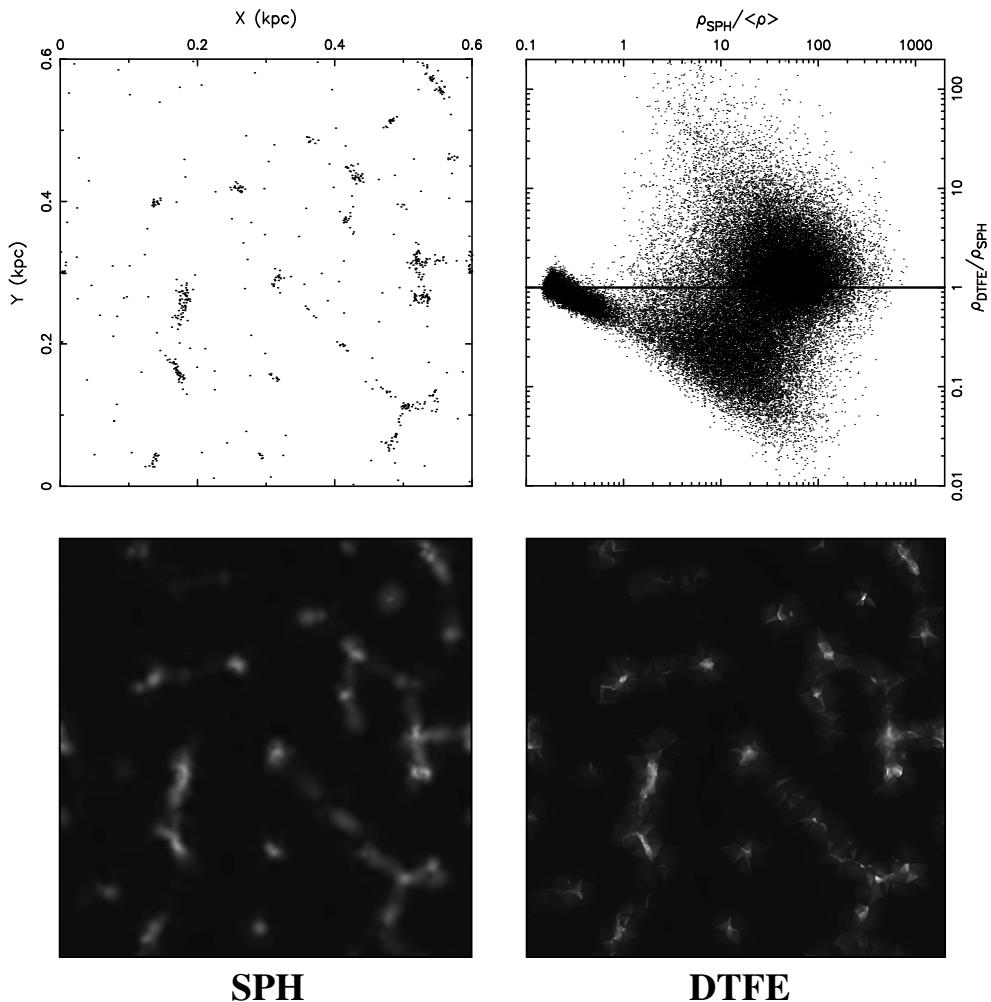


Figure 5.1 — Comparison of the DTFE performance versus that of the regular SPH method in a characteristic configuration, a hydrodynamic simulation of the multi-phase interstellar medium. Top left-hand frame: the particle distribution in a 0.6×0.6 kpc simulation region, within a slice with a width of 0.005 kpc. Bottom left-hand frame: two-dimensional slice through the SPH density field reconstruction. Bottom right-hand frame: the corresponding slice through the DTFE density field reconstruction. Top right-hand frame: summary, in terms of a quantitative point-by-point comparison between the DTFE and SPH density estimates, ρ_{DTFE} and ρ_{SPH} . Abscissa: the value of the SPH density estimate (normalized by the average density $\langle\rho\rangle$). Ordinate: the ratio of DTFE to the SPH density estimate, $\rho_{\text{DTFE}}/\rho_{\text{SPH}}$. These quantities are plotted for each particle location in the full simulation box.

processes. No self-gravity is included. As for the thermodynamical state of the gas, cooling is implemented using a fit to the Dalgarno-McCrory cooling curve (1972). The heating of the gas is accomplished through photo-electric grain heating, attributed to a constant FUV background radiation field ($1.7 G_0$, with G_0 the Habing field). The parameters are chosen such that after about 15 Myrs a two-phase medium forms which consists of warm (10 000 K) and cold (> 100 K) H I gas.

The stage at which a two-phase medium emerges forms a suitable point to investigate the performance of the SPH and DTFE methods. At this stage we took a snapshot from the simulation and subjected it to further analysis. For a variety of reasons the spatial gas distribution of the snapshot is expected to represent a challenging configuration. The multi-phase character of the resulting particle configuration is likely to present a problem for regular SPH. Density contrasts of about four orders of magnitude separate dense clumps from the surrounding diffuse medium through which they are dispersed. Note that a failure to recover the correct density may have serious repercussions for the computed effects of cooling. In addition, we notice the presence of physical structures with conspicuous, aspherical geometries (see Figs. 5.1 and 5.2), such as anisotropic sheets and filaments as well as dense and compact clumps, which certainly do form a challenging aspect for the different methods.

5.3.1 Results

Fig. 5.1 offers a visual impression of the differences in performance between the SPH and DTFE density reconstructions. The grayscale density maps in Fig. 5.1 (lower left-hand frame: SPH, lower right-hand frame: DTFE) represent two-dimensional cuts through the corresponding three-dimensional density field reconstructions (note that contrary to the finite width of the corresponding particle slice in the upper left-hand frame these constitute planes with zero thickness).

Immediately visible is the more crispy appearance of the DTFE density field, displaying substantially more contrast in conjunction with more pronounced structural features. Look e.g. at the compact clump in the lower right-hand corner ($X \approx 0.5, Y \approx 0.12$), forming a prominent and tight spot in the DTFE density field. The clump at ($X \approx 0.48, Y \approx 0.52$) represents another telling example, visible as a striking peak in the DTFE rendering while hardly noticeable in the SPH reconstruction. Structures in the SPH field have a more extended appearance than their counterparts in the DTFE field, whose matter content has been smeared out more evenly, over a larger volume, yielding features with a significantly lower contrast. In this assessment it becomes clear that the DTFE reconstruction adheres considerably closer to the original particle distribution (top left-hand frame). Apparently the DTFE succeeds better in rendering the shapes, the coherence and the internal composition in the displayed particle distribution. At various locations the DTFE manages to capture structural details which seem to be absent in the SPH density field.

To quantify the visual impressions of Fig. 5.1 and to analyze the nature of the differences between the two methods we have plotted the ratio $\rho_{\text{DTFE}}/\rho_{\text{SPH}}$ as a function of the SPH density estimate $\rho_{\text{SPH}}/\langle \rho \rangle$ (in units of the average density $\langle \rho \rangle$). Doing so for all particles in the sample (Fig. 5.1, top right-hand frame, Fig. 5.2, top left-hand frame) immediately reveals interesting behaviour. The scatter diagram does show that the discrepancies between the two methods may be substantial, with density estimates at various instances differing by a factor of 5 or more.

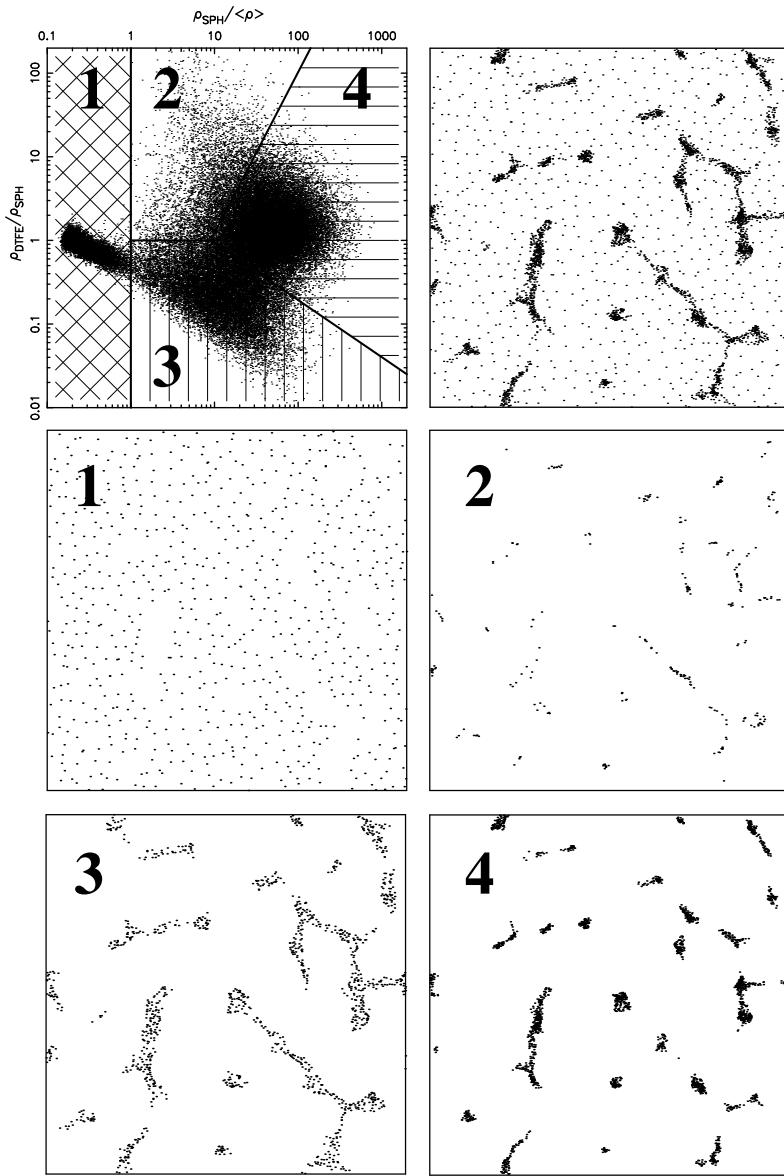


Figure 5.2 — Systematic analysis of the differences between the DTFE and SPH density estimates, ρ_{DTFE} and ρ_{SPH} . Basis of the analysis is a point-by-point comparison of these two density estimates. Top left-hand frame: diagram of the value of the ratio $\rho_{\text{DTFE}}/\rho_{\text{SPH}}$ (ordinate) versus $\rho_{\text{SPH}}/\langle \rho \rangle$ (abscissa) for each of the points in the simulation volume. Indicated in this scatter diagram are four sectors, each of which corresponds to particles residing in a physically different regime/phase. On the basis of this identification, the full set of particles is dissected into the corresponding four composing particle samples. Top right-hand frame: the spatial distribution of the full set of particles in a 0.04 kpc wide slice. The subsequent 4 frames (from central left-hand to bottom right-hand) show, for each indicated sector in the scatter diagram, the spatial distribution of the corresponding particles (within the same 0.04 kpc slice).

Most interesting is the finding that we may distinguish clearly identifiable and distinct regimes in the scatter diagram of $\rho_{\text{DTFE}}/\rho_{\text{SPH}}$ versus $\rho_{\text{SPH}}/\langle\rho\rangle$. Four different sectors may be identified in the scatter diagram. Allowing for some arbitrariness in their definition, and indicating these regions by digits 1 to 4, we may organize the particles according to density-related criteria, roughly specified as (we refer to Fig. 5.2, top left-hand frame, for the precise definitions of the domains):

1. low density regions:

$$\rho_{\text{SPH}}/\langle\rho\rangle < 1;$$

2. medium density regions, DTFE smaller than SPH:

$$\rho_{\text{DTFE}} < \rho_{\text{SPH}}; \quad 1 < \rho_{\text{SPH}}/\langle\rho\rangle < 10;$$

3. medium density regions, DTFE larger than SPH:

$$\rho_{\text{DTFE}} > \rho_{\text{SPH}}; \quad 1 < \rho_{\text{SPH}}/\langle\rho\rangle < 10;$$

4. high density regions:

$$\rho_{\text{DTFE}} \gtrsim \rho_{\text{SPH}}; \quad \rho_{\text{SPH}}/\langle\rho\rangle > 10.$$

The physical meaning of the distinct sectors in the scatter diagram becomes apparent when relating the various regimes with the spatial distribution of the corresponding particles. This may be appreciated from the five subsequent frames in Fig. 5.2, each depicting the related particle distribution in the same slice of width 0.04 kpc. The central and bottom frames, numbered 1 to 4, show the spatial distribution of each group of particles, isolated from the complete distribution (top right-hand frame, Fig. 5.2). These particle slices immediately reveal the close correspondence between any of the sectors in the scatter diagram and typical features in the spatial matter distribution of the two-phase interstellar medium. This systematic behaviour seems to point to truly fundamental differences in the workings of the SPH and DTFE methods, and would be hard to understand in terms of random errors. The separate spatial features in the gas distribution seem to react differently to the use of the DTFE method.

We argue that the major share of the disparity between the SPH and DTFE density estimates has to be attributed to SPH, mainly on the grounds of the known fact that SPH is poor in handling non-trivial configurations such as encountered in multi-phase media. By separately assessing each regime, we may come to appreciate how these differences arise. In sector 1, involving the diffuse low density medium, the DTFE and SPH estimates are of comparable magnitude, be it that we do observe a systematic tendency. In the lowest density realms, whose relatively smooth density does not raise serious obstacles for either method, DTFE and SPH are indeed equal (with the exception of variations to be attributed to random noise). However, near the edges of the low density regions, SPH starts to overestimate the local density as the kernels do include particles within the surrounding high density structures. The geometric interpolation of the DTFE manages to avoid this systematic effect, which explains the systematic linear decrease of the ratio $\rho_{\text{DTFE}}/\rho_{\text{SPH}}$ with increasing $\rho_{\text{SPH}}/\langle\rho\rangle$. To the other extreme, the high density regions in sector 4 are identified with compact dense clumps as well as with their extensions into connecting filaments and walls. On average DTFE yields higher density estimates than SPH, frequently displaying superior spatial resolution (see also grayscale plot in Fig. 5.1). Note that the repercussions may be far-reaching in the context of a

wide variety of astrophysical environments characterized by strongly density-dependent physical phenomena and processes! The intermediate regime of sectors 2 and 3 clearly connects to the filamentary structures in the gas distribution. Sector 2, in which the DTFE estimates are larger than those of SPH, appears to select the inner parts of the filaments and walls. By contrast, the higher values for the SPH-produced densities in sector 3 are related to the outer realms of these features. This characteristic distinction can be traced back to the failure of the SPH procedure to cope with highly anisotropic particle configurations. While it attempts to maintain a fixed number of neighbours within a spherical kernel, it smears out the density in a direction perpendicular to the filament. This produces lower estimates in the central parts, which are compensated for with higher estimates in the periphery. Evidently, the adaptive nature of DTFE does not appear to produce similar deficiencies.

5.4 The DTFE particle method

Having demonstrated the improvement in quality of the DTFE density estimates, this suggests a considerable potential for incorporating the DTFE in a self-consistent manner within a hydrodynamical code. Here, we first wish to indicate a possible route for accomplishing this in a particle hydrodynamics code through replacement of the kernel-based density estimates (5.1) by the DTFE density estimates. The formalism which is needed for to implement this can be easily derived, involving non-trivial yet minor modifications. Essentially, it uses the same dynamical equations for gas particles as those in the regular SPH formalism, the fundamental adjustment being the insertion of the DTFE densities instead of the regular SPH ones. In addition, a further difference may be introduced through a change in treatment of viscous forces. Ultimately, this will work out into different equations of motion for the gas particles. A fundamental property of a DTFE-based hydrocode, by construction, is that it conserves mass exactly and therefore obeys the continuity equation. This is not necessarily true for SPH implementations (Hernquist & Katz 1989, see also Chapter 3).

The start of the suggested DTFE particle method is formed by the discretized expression for the Lagrangian L for a compressible, non-dissipative flow,

$$L = \sum_i m_i \left(\frac{1}{2} v_i^2 + u_i(\rho_i, s_i) \right), \quad (5.6)$$

where m_i is the mass of particle i , v_i its velocity, s_i the corresponding entropy and u_i its specific internal energy. In this expression, ρ_i is the density at location i , as yet unspecified. The resulting Euler-Lagrange equations are

$$m_i \frac{dv_i}{dt} = - \sum_j m_j \left(\frac{\partial u_j}{\partial \rho_j} \right)_s \frac{\partial \rho_j}{\partial x_i}. \quad (5.7)$$

The standard SPH equations of motion then follow after inserting the SPH density estimate (Eq. 5.1). Instead, insertion of the DTFE density (Eq. 5.4) will lead to the corresponding equations of motion for the DTFE-based formalism. Note that the usual conservation properties related to Eq. 5.6 remain intact. After some algebraic manipulation, thereby using the basic thermodynamic relation for a gas with equation of state $P(\rho)$,

$$\left(\frac{\partial u_i}{\partial \rho_i} \right)_s = \frac{P_i}{\rho_i^2}, \quad (5.8)$$

we finally obtain the equations of motion for the gas particles,

$$m_i \frac{dv_i}{dt} = \frac{1}{4} \sum_{j=1}^{N_{\mathcal{T}_i}} P(\mathcal{T}_{j,i}) \frac{\partial V(\mathcal{T}_{j,i})}{\partial x_i}. \quad (5.9)$$

This expression involves a summation over all $N_{\mathcal{T}_i}$ Delaunay tetrahedra $\mathcal{T}_{j,i}$, with volumes $V(\mathcal{T}_{j,i})$, which have the particle i as one of its four vertices. The pressure term $P(\mathcal{T}_{j,i})$ is the sum over the pressures P_j at the four vertices m of tetrahedron $\mathcal{T}_{j,i}$, $P(\mathcal{T}_{j,i}) = \sum P_m$.

As an interesting aside we point out that unlike in the conventional SPH formalism this procedure implies an exactly vanishing acceleration dv_i/dt in the case of a constant pressure P at each of the vertices of the Delaunay tetrahedra containing particle i as one of their vertices. The reason for this is that one can then invoke the definition of the volume of the contiguous Voronoi cell corresponding to point i (Eqn. 5.3), yielding

$$m_i \frac{dv_i}{dt} = \frac{1}{D+1} P \frac{\partial V(W_i)}{\partial x_i}. \quad (5.10)$$

Since the volume of the contiguous Voronoi cell does not depend on the position of particle i itself (it lies in the interior of the contiguous Voronoi cell), the resulting acceleration vanishes. Another interesting notion, which was pointed out by Icke (2002), is that Delaunay tessellations also provide a unique opportunity to include a natural treatment of the viscous stresses in the physical system (see also Ritzerveld et al. 2003).

5.5 Delaunay tessellations and ‘moving grid’ hydrocodes

Ultimately, the ideal hydrodynamical code would combine the advantages of both the Eulerian and the Lagrangian approach. In their simplest formulation, Eulerian algorithms cover the volume of study with a fixed grid and compute the fluid transfer through the faces of the (fixed) grid-cell volumes to follow the evolution of the system. Lagrangian formulations, on the other hand, compute the system by following the ever changing volume and shape of a particular individual element of gas (interestingly, the ‘Lagrangian’ formulation is also due to Euler, who employed this formalism in 1759 in a letter to Lagrange, who in 1762 proposed these ideas in a publication by himself; see Whitehurst 1995).

For a substantial part the success of the DTFE may be ascribed to the use of Delaunay tessellations as an optimally covering grid. This suggests that they may also be ideal for the use in moving grid implementations for hydrodynamical calculations. As in our SPH application hydrocodes with Delaunay tessellations at their core would warrant a close connection to the underlying matter distribution. Indeed, attempts towards such implementations have already been introduced in the context of a few specific, mainly two-dimensional applications (Whitehurst 1995, Braun & Sambridge 1995, Sukumar 1998). Alternative attempts towards the development of moving grid codes, in an astrophysical context, have shown their potential (Gnedin 1995, Pen 1998).

For a variety of astrophysical problems it is indeed essential to have such advanced codes at one’s disposal. An example of high current interest may offer a good illustration. Such an example is the reionization of the intergalactic medium by the ionizing radiation emitted by the first generation of stars, (proto)galaxies and/or active galactic nuclei. These radiation

sources will form in the densest regions of the universe. To be able to resolve these in sufficient detail, it is crucial that the code is able to focus in onto these densest spots. Their emphasis on mass resolution makes Lagrangian codes – including SPH – usually better equipped to do so, be it not yet optimally. On the other hand, it is in the low density regions that most radiation is absorbed at first. In the early stages the reionization process is therefore restricted to the huge underdense fraction of space. Simulation codes should therefore properly represent and resolve the gas density distribution within these void-like regions. The uniform spatial resolution of the Eulerian codes is better suited to accomplish this. Ideally, however, a simulation code should be able to combine the virtues of both approaches, yielding optimal mass resolution in the high density source regions and a proper coverage of the large underdense regions. Moving grid methods, of which Delaunay tessellation-based ones will be a natural example, may indeed be the best alternative, as the reionization simulations by Gnedin (1995) appear to indicate. There have been many efforts in the context of Eulerian codes towards the development of adaptive mesh refinement (AMR) algorithms (Berger & Colella 1989), which have achieved a degree of maturity. Their chief advantage is their ability to concentrate computational effort on regions based on arbitrary refinement criteria, where, in the basic form at least, moving grid methods refine on a mass resolution criterion. However, they are still constrained by the use of regular grids, which may introduce artefacts due to the presence of preferred directions in the grid.

The advantages of a moving grid fluid-dynamics code based on Delaunay tessellations have been most explicitly demonstrated by the implementation of a two-dimensional Lagrangian hydrocode (FLAME) by Whitehurst (1995). These advantages will in principle apply to any such algorithm, in particular also for three-dimensional implementations (of which we are currently unaware). Whitehurst enumerated various potential benefits in comparison with conventional SPH codes, most importantly the following:

1. SPH needs a smoothing length h ;
2. SPH needs an arbitrary kernel function W ;
3. The moving grid method does not need an (unphysical) artificial viscosity to stabilize solutions.

The validity of the first two claims has of course also been demonstrated in this study for particle methods based on DTFE. Whitehurst showed that there is an additional advantage of moving grid methods over Eulerian grid-based ones. The implementation of Whitehurst, which used a first-order solver and a limit on the shape of the grid-cells to control the effect of shearing of the grid, was far superior to all tested first-order Eulerian codes, and superior to many second-order ones as well. The adaptive nature of the Langrangian method and the fact that the resulting grid has no preferred directions are key factors in determining the performance of moving grid methods such as FLAME. For additional convincing arguments, including the other claims, we may refer the reader to the truly impressive case studies presented by Whitehurst.

5.6 Summary and discussion

In this chapter we have introduced the DTFE as an alternative density estimator for particle fluid-dynamics. Its principle asset is that it is fully self-adaptive, resulting in a density field

reconstruction which closely reproduces the characteristics of the spatial particle distribution. It may do so because of its complete independence of arbitrary user-specified smoothing functions and parameters. Unlike conventional methods, such as SPH-based procedures, it manages to faithfully reproduce the anisotropies in the local particle distribution. It therefore automatically reflects the genuine geometry and shape of the structures present in the underlying density field. This is in marked contrast with kernel-based methods, which almost without exception produce distorted shapes of density features, the result of the convolution of the real structure with the intrinsic shape of the smoothing function. Its adaptive and local nature also makes it optimally suited for reconstructing the hierarchy of scales present in the density distribution. In kernel-based methods the internal structural richness of density features is usually suppressed on scales below that of the characteristic (local) kernel scale. DTFE, however, is solely based on the particle distribution itself and follows the density field wherever the discrete representation by the particle distribution allows it to do so. Its capacity to resolve structures over a large dynamic range may prove to be highly beneficial in many astrophysical circumstances, quite often involving environments in which we encounter a hierarchical embedding of small-scale structures within more extended ones.

In this study we have investigated the performance of the DTFE density estimator in the context of a smooth particle hydrodynamics simulation of a multi-phase interstellar medium of neutral gas. The limited spatial resolution of current particle hydrodynamics codes are known to implicate considerable problems near regions with e.g. steep density and temperature gradients. Especially their handling of shocks forms a source of considerable concern. SPH often fails in and around these regions, which often play a critical and vital role in the evolution of physical systems. Our study consists of a comparison and confrontation of the conventional SPH kernel-based density estimation procedure with the corresponding DTFE density field reconstruction method.

The comparison of the density field reconstructions demonstrated convincingly the considerable improvement embodied by the DTFE procedure. This is in particular true at locations and under conditions where SPH appears to fail. SPH overestimates the density in the outer parts of voids, filaments and sheet-like structures, whereas it underestimates the density in the interior of filaments, sheet-like structures and in high-density clumps.

Having shown the success of the DTFE, we are convinced that its application towards the analysis of the outcome of SPH simulations will prove to be highly beneficial. This may be underlined by considering a fitting illustration. Simulations of the settling and evolution of the X-ray emitting hot intracluster gas in forming clusters of galaxies do represent an important and cosmologically relevant example (see Borgani & Guzzo 2001 and Rosati et al. 2002 for recent reviews). The X-ray luminosity is strongly dependent on the density of the gas. The poor accuracy of the density determination in regular SPH calculations therefore yields deficient X-ray luminosity estimates (see Bertschinger 1998 and Rosati et al. 2002 for relevant recent reviews). Despite a number of suggested remedies, such as separating particles according to their temperature, their ad hoc nature does not evoke a strong sense of confidence in the results. Numerical limitations will of course always imply a degree of artificial smoothing, but by invoking tools based on the DTFE technique there is at least a guarantee of an optimal retrieval of information contained in the data.

Despite its promise for the use in a variety of analysis tools for discrete data samples, such as particle distributions in computer simulations or galaxy catalogues in an observational

context, its potential would be most optimally exploited by building it into genuine new fluid-dynamics codes. Some specific (two-dimensional) examples of successful attempts in other scientific fields were mentioned, and we argue for a similar strategy in astrophysics. One path may be the upgrade of current particle hydrodynamics codes by inserting DTFE technology. In this chapter we have outlined the development of such an SPH-like hydrodynamics scheme in which the regular kernel estimates are replaced by DTFE estimates. One could interpret this in terms of the replacement of the user-specified kernel by the self-adaptive contiguous Delaunay cell, solely dependent on the local particle configuration. An additional benefit is that on the basis of the localized connections in a Delaunay tessellation it will be possible to define a more physically motivated artificial viscosity term.

The ultimate hydrodynamics algorithm would combine the virtues of Eulerian and Lagrangian techniques. Considering the positive experiences with DTFE, it appears to be worthwhile within the context of ‘moving grid’ or ‘Lagrangian grid’ methods to investigate the use of Delaunay tessellations for solving the Euler equations. With respect to a particle hydrodynamics code, the self-adaptive virtues of DTFE and its ability to handle arbitrary density jumps with only one intermediate point may lead to significant improvements in the resolution and shock handling properties. Yet, for grid-based methods major complications may be expected in dealing with the non-regular nature of the corresponding cells, complicating the handling of flux transport along the boundaries of the Delaunay tetrahedra.

The computational cost of DTFE resembling techniques is not overriding. The CPU time necessary for generating the Delaunay tessellation corresponding to a point set of N particles is in the order of $O(N \log N)$, comparable to the cost of generating the neighbour list in SPH. Within an evolving point distribution these tessellation construction procedures may be made far more efficient, as small steps in the development in the system will induce a correspondingly small number of tetrahedron (identity) changes.

In summary, in this work we have argued for and demonstrated the potential and promise of a natural computational technique which is based on one of the most fundamental and natural tilings of space, the Delaunay tessellation. Although the practical implementation will undoubtedly encounter a variety of complications, dependent on the physical setting and scope of the code, the final benefit of a natural moving grid hydrodynamics code for a large number of astrophysical issues may not only represent a large progress in a computational sense. Its major significance may be found in its ability to address fundamental astrophysical problems in a new and truly natural way, leading to important new insights in the workings of the cosmos.

References

- Berger M.J., Colella P., 1989, J. Comp. Phys., 82, 64
- Bernardeau F., van de Weygaert R., 1996, MNRAS, 279, 693
- Bertschinger E., 1998, ARAA, 36, 599
- Borgani S., Guzzo L., 2001, Nature, 409, 39
- Braun J., Sambridge M., 1995, Nature, 376, 655
- Dalgarno A., McCray R.A., 1972, A&A, 10, 375
- Delone B.N., 1934, Bull. Acad. Sci. USSR: Classe Sci. Mat., 7, 793
- Euler L., 1862, Letter dated 1759 October 27 from Euler to Lagrange, Opera Postuma, 2, 559

- Gingold R.A., Monaghan J.J., 1977, MNRAS, 181, 375
Gnedin N.Y., 1995, ApJS, 97, 231
Hernquist L., Katz N., 1989, ApJS, 70, 419
Hernquist L., 1993, ApJ, 404, 717
Icke V., 2002, *private communication*
Lagrange J.L., 1762, Oeuvres de Lagrange, 1, 151
Lucy L.B., 1977, AJ, 82, 1013
Mavriplis D.J., 1997, Ann. Rev. Fluid Mech., 29, 473
Møller J., 1994, *Lecture notes on Random Voronoi Tessellations*, Lecture Notes in Statistics 87,
Springer-Verlag, New York, USA
Monaghan J.J., 1992, ARAA, 30, 543
Nelson R.P., Papaloizou J.C.B., 1994, MNRAS, 270, 1
Okabe A., Boots B., Sugihara K., Nok Chiu S., 2000, *Spatial Tessellations, Concepts and
Applications of Voronoi Diagrams*, 2nd edition, Wiley, Chichester, UK
Pen U., 1998, ApJS, 115, 19
Rasio F.A., 1999, astro-ph/9911360
Ritzerveld J., Icke V., Rijkhorst E.-J., 2003, astro-ph/0312301
Rosati P., Borgani S., Norman C., 2002, ARAA, 40, 539
Sambridge M., Braun J., McQueen H., 1995, Geophys. J. Int., 122, 837
Shapiro P.R., Martel H., Villumsen J.V., Owen J.M., 1996, ApJS, 103, 269
Sibson R., 1981, in *Interpreting Multi-Variate Data*, Barnett V. (ed.), Wiley, Chichester, UK
Springel V., Hernquist L., 2002, MNRAS, 333, 649
Sukumar N., 1998, PhD Thesis, Northwestern University, Evanston, IL, USA
van de Weygaert R., 1991, PhD Thesis, Leiden University, Leiden, The Netherlands
White S.D.M., 1994, in *Cosmology and Large Scale Structure*, Les Houches Session LX, Schaeffer R.,
Silk J., Spiro M., Zinn-Justin J. (eds.), Elsevier Science, Amsterdam, NL
Whitehurst, R. 1995, MNRAS, 277, 655

6

Foam Dynamics: on Matter and Motions around the Cosmic Foam[†]

W.E. Schaap, R. van de Weygaert & E. Romano-Díaz

ABSTRACT — The Delaunay Tessellation Field Estimator (DTFE) is a method for reconstructing density or intensity fields from a discrete set of irregularly distributed points sampling this field. In this chapter we show that it can also be used to reconstruct other continuous fields which are sampled at the locations of these points. The main advantage of the DTFE is that the fields are reconstructed locally without the application of an artificial or user-dependent smoothing procedure, resulting in an optimal resolution and the suppression of shot-noise effects. The reconstructed fields are volume-covering and allow for a direct comparison with theoretical predictions. In this chapter we focus on the simultaneous reconstruction of the density and velocity fields corresponding to cosmological N -body simulations. The resulting fields closely adhere to the continuity equation. The DTFE reconstruction results in realistic void density and velocity profiles. We present analytical models for voids in cosmologies with and without a cosmological constant and show that voids can be used to constrain cosmological parameters. We also discuss the reconstruction of the density and velocity fields in filamentary environments and show that the DTFE is able of capturing both the almost caustic behaviour of the density and velocity field as well as the large scale infalling motions towards and around these objects. The DTFE reconstruction of the cosmic velocity field automatically provides a reconstruction of the velocity divergence, shear and voricity fields, which we shortly discuss. Our results show that the DTFE represents a major step forward for studies of the dynamics and evolution of the cosmic foam.

[†]part of this chapter is based on Schaap W.E., van de Weygaert R., 2003, *Delaunay recovery of cosmic density and velocity probes*, in ‘Statistical challenges in modern astronomy III’, Babu G.J, Feigelson E.D. (eds.), Springer-Verlag, New York, USA

6.1 Introduction

The peculiar velocity field provides important information on the formation and evolution of structures in our universe. On cosmological scales it supplies direct information on the dynamics of structure formation, while on smaller scales it provides information on the local environment. In general, the peculiar velocity field is sampled at a discrete set of points. These points correspond to the locations of galaxies in the case of observational data or to particle positions in the case of N -body simulations. In both these cases the sampling density is related to the cosmic density field, which is highly non-uniform. Large scale galaxy redshift surveys have shown that galaxies are not randomly distributed in the universe but reside in an intricate network, the cosmic web (Bond et al. 1996, van de Weygaert 2002). Dense and compact galaxy clusters containing up to thousands of galaxies reside at the nodes of this network. The Virgo and Coma clusters form the most nearby examples. The clusters are connected by filamentary and wall-like structures of intermediate densities, usually consisting of a large number of galaxy groups, each containing tens of galaxies. The most outstanding of these structures are usually identified with superclusters. The Local Supercluster, with the Local Group amongst its members, forms the most nearby example. The extended and elongated superclusters surround large and empty regions, voids. The result is a foamy network which is pervading throughout the whole observable universe.

The complex nature of the large scale matter distribution may be understood in terms of the standard framework of structure formation, gravitational instability theory. In this theory physical objects originate from small primordial fluctuations in the density field. Gravity acts as the driver of structure formation by exerting a larger than average pull towards overdense regions and a smaller than average pull towards underdense regions. The density of overdense regions continues to grow by attracting more and more matter, causing an ever stronger gravitational pull on its surroundings. Once these regions surpass a given density threshold they collapse and virialize, attaining a high density excess. Conversely, regions of low density expand and empty with time, gradually approaching a density contrast of $\delta \approx -1$.

The complex cosmic matter distribution poses a formidable challenge for studies of the cosmic velocity field. The main reason for this is that the cosmic velocity field is sampled by galaxies which predominantly reside in high density environments, such as filaments and clusters. Whereas such regions are compact, the extended and low density voids contain very few galaxies and are very poorly sampled. As a consequence, the peculiar velocity field is sampled in a highly non-uniform fashion. An additional complication concerns the fact that peculiar velocity fields are measured at galaxy positions. Due to the filtering which is applied in reconstruction procedures reconstructed fields are usually mass-weighted, while most analytical relations concern volume-weighted quantities. The a priori unknown bias between dark matter and baryonic matter makes the comparison between measured and theoretical values even more difficult.

Given these issues the central problem becomes how to reconstruct a continuous volume-covering field from a discrete and highly non-uniform sampling without losing relevant information. This is closely related to the reconstruction of a continuous and volume-covering density field from a discrete set of particles, which has been discussed in detail in the previous chapters. The difference is that in the case of the cosmic velocity field one starts with a set of velocity field samples, while in the case of a density field all one has is a set of sampling point

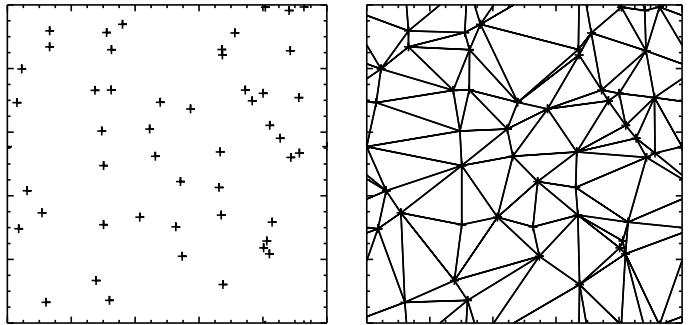


Figure 6.1 — A set of 50 points and the corresponding Delaunay tessellation in two dimensions. Periodic boundary conditions are assumed.

locations, which itself is supposed to represent a fair sample of the underlying density field.

In this thesis we have described the Delaunay Tessellation Field Estimator (DTFE), a method for reconstructing density fields from a discrete set of sampling points. The DTFE forms an elaboration of the formalism first proposed by Bernardeau & van de Weygaert (1996) for the case of assessing the statistical properties of cosmic velocity fields. In the previous chapters we have shown that the main advantage of the DTFE over conventional fixed grid-based and adaptive SPH-like procedures is that it is intrinsically self-adaptive to both the density and the geometry of the point distribution and does not make use of any artificial smoothing. This makes it a superior tool for studying the large scale galaxy distribution. It attains the highest spatial resolution. It can accurately describe the full hierarchy of spatial scales and densities present in the large scale galaxy distribution. Finally, it recovers the complex cellular geometry of the cosmic web involving structures of widely varying densities and anisotropies.

The DTFE consists of three main steps. In the first step the Delaunay tessellation (Delone 1934) of the point sample is constructed. In the second step the properties of this tessellation are employed to obtain an estimate of the density at the location of the sampling points. In the final step these estimates are interpolated to the rest of space, again by making use of the Delaunay tessellation. The resulting density field is volume-covering and continuous and can be used as a starting point for a variety of subsequent analyses.

The central element of the DTFE is the Delaunay tessellation of the point sample. This is a volume-covering tiling of space into tetrahedra whose vertices are formed by four specific points in the dataset. An example of a Delaunay tessellation is shown in Fig. 6.1. The four vertices of each Delaunay tetrahedron are uniquely selected such that their circumscribing sphere does not contain any of the other datapoints. Another important property of the Delaunay tessellation is that among all possible triangulations of a given point distribution the Delaunay tetrahedra are on average objects of minimal size and elongation (Lawson 1977, Okabe et al. 2000). These properties make the network of Delaunay cells the multi-dimensional Delaunay tessellations have been abundantly applied in surface rendering applications such as geographical mapping and various computer imaging algorithms (see Okabe 2000 for a review).

Given the advantages of the DTFE in the reconstruction and analysis of cosmological density fields, as well as the promising results of using Delaunay tessellations for the interpolation of various other physical quantities, it is highly interesting to combine these techniques for a simultaneous reconstruction of the cosmic density and velocity field. In this chapter we

explicitly focus on those environments for which conventional analysis tools fail: low density voids and anisotropic filamentary and wall-like objects of intermediate density.

The outline of this chapter is as follows. In section 6.2 we discuss the reconstruction of a cosmic velocity field by means of a conventional interpolation technique. In section 6.3 we describe the DTFE velocity field reconstruction procedure. In section 6.4 we apply this formalism to a GIF N -body simulation of cosmic structure formation. We discuss the simultaneous reconstruction of the corresponding cosmic density and velocity field and assess the quality of the reconstructions. In section 6.5 we focus on voids, while in section 6.6 we describe filamentary and wall-like environments. In section 6.7 we discuss the reconstruction of the velocity shear and vorticity fields. Finally, in section 6.8 we summarize and discuss our results.

6.2 Case study: a cosmic velocity field

Several methods have been proposed for reconstructing a volume-covering field from a discrete set of data-points. We can divide these methods into three main groups. The first is that of the grid-based schemes, also known as Eulerian, in which the desired field estimates are confined to a set of locations on a pre-defined fixed grid, which in principle does not depend on the point distribution itself (e.g. Hockney & Eastwood 1988; Efstathiou et al. 1985). In cosmology these grids are usually regular but various other options can be applied depending on the system under consideration. Their main limitation is that their spatial resolution is set by the cell-size. This limitation is quite severe in cosmology, since many relevant physical processes are acting over a range of scales, so that once a grid-size has been defined effective smoothing over the Eulerian cell-size usually erases important information on smaller scales.

The second group is formed by the Lagrangian schemes, in which the locations of the interpolation are confined to or defined by the point distribution itself. Of this type, the ‘SPH-type’ schemes are the best known in astronomy. The latter refers to the abundantly used smooth particle hydrodynamics technique (Lucy 1977, Gingold & Monaghan 1977) to follow the hydrodynamical evolution of astrophysical systems. The main difference between Lagrangian and Eulerian methods is the fact that the former is not restricted to a specific geometry because it does not make use of a mesh. Instead, it follows the trajectories of the displacing matter aggregating into high density regions and ideally with an unlimited spatial resolution. In practice, this is not possible and one has to make use of a ‘softening length’. This length is usually much smaller than the cell-size of an Eulerian scheme, resulting in a significantly higher spatial resolution of Lagrangian schemes. The main difficulty of this technique is that it relies on stochastic arguments which means that it yields only approximate solutions at a given spatial position.

The third class seeks to combine the virtues of Lagrangian and Eulerian methods. These were first introduced by Noh (1964) and referred to as arbitrary Lagrangian-Eulerian (ALE) methods. This technique incorporates the high resolution of the Lagrangian scheme by letting a grid move along with the system, combined with the Eulerian scheme for computing the physical state of the system within each of the resulting distorted grid-cells. Although as yet reluctantly applied in cosmology, there are some noteworthy promising efforts (e.g. Gnedin 1995, Xu 1997, Pen 1998).

For the purpose of merely providing a comparison, we restrict ourselves to a grid-based

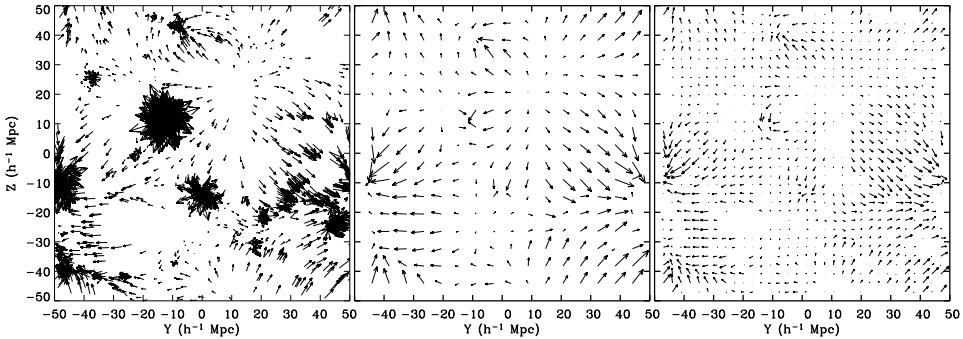


Figure 6.2 — Interpolated TSC velocity fields. The left-hand frame shows the N -body velocities at the particle positions. The central and right-hand frames depict the TSC interpolated velocity field onto a 64^3 and 128^3 cubic grid.

TSC scheme as a representative velocity field reconstruction algorithm. All other Eulerian grid-based methods perform in a similar fashion. For a general revision of the performance of several grid and Lagrangian methods we refer the reader to Chapters 4 and 5.

In Fig. 6.2 the outcome of an N -body simulation is shown (left-hand frame) together with two TSC velocity field reconstructions (central and right-hand frames) at different resolutions. The region depicted corresponds to a $1h^{-1}\text{Mpc}$ thick slice through a cosmological simulation box of size $100h^{-1}\text{Mpc}$. This box contains several common large scale structures. At the top right-hand and bottom left-hand part of the slice two large voids are present, close to the center a cluster is located and at the bottom right-hand part a filamentary structure is located. For the central frame the resolution of the TSC grid is the standard choice of on average one galaxy per grid-cell for the full galaxy sample, in the right-hand plot it is twice as large.

Clear differences can be distinguished between the two maps. The low density interpolated map assures full volume-coverage, but at the price of blurring out high density regions. Although the main trend of the flows can be distinguished, many of the features visible in the particle velocities are not recovered in the low resolution interpolated map. Even though a trace of related shearing flow is visible at the bottom right-hand part of the map, it is only marginal. The high resolution map of Fig. 6.2 (right-hand frame) illustrates the second point. This map does recover structures that the low resolution map did not. The high density regions can be better noticed in this map. The general flow resembles that visible in the discrete velocity map. Particularly noteworthy are the low density regions, devoid of any objects to base the velocity interpolation on. Here the high resolution TSC procedure is unable to recover any significant information. A possible solution to circumvent this problem is to apply a smoothing procedure in order to extrapolate information to those regions where no information is available. However, the cost of such an operation would be to lose the high sensitivity of the map, degrading it to an equivalent of the low resolution TSC reconstruction.

From the map we may infer that: (1) At high density regions the interpolation smoothes away interesting velocity information because one effectively averages over the velocities within a grid-cell. (2) Highly crucial with respect to the ability to reproduce the one-point velocity distribution function, at low density regions the sampling is so poor and sparse that

there is no strong and significant signal in these regions to be recovered. (3) The field estimates are beset by shot-noise effects. Note that if one has less than one or two particles within the kernel, estimates of the local bulk flow and/or velocity dispersion would not be possible. (4) Cosmic velocity fields are much more quiet than density fields. Hence, features are intrinsically less prominent and therefore require a better sampling to obtain a sufficient contrast.

6.3 DTFE reconstruction of velocity fields

The DTFE was introduced for rendering continuous and fully volume-covering density fields from a discrete set of points sampling this field. It adapts to the density and geometry of the point distribution and does not make use of artificial smoothing procedures. It is an extension of the pioneering work by Bernardeau & van de Weygaert (1996) who used the Delaunay tessellation of a point set as a natural and self-adaptive interpolation frame for recovering the continuous velocity field sampled at those points. The DTFE generalized this formalism to the recovery of a density or intensity field on the assumption that it is fairly sampled by the spatial point distribution.

Mathematically the DTFE interpolation procedure for a field which is sampled at a number of discrete locations can be formulated as follows. Let the value of a field f be known at a set of N irregularly distributed locations \mathbf{x}_i . Let the corresponding Delaunay tessellation consist of N_T tetrahedra. Let point \mathbf{x} lie inside tetrahedron j . Let this tetrahedron consist of the 4 vertices $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2$ and \mathbf{x}_3 . Then the value of the function f at location \mathbf{x} is defined as

$$\widehat{f}(\mathbf{x}) = f(\mathbf{x}_0) + \widehat{\nabla f} \Big|_j \cdot (\mathbf{x} - \mathbf{x}_0). \quad (6.1)$$

Here the gradient of f in tetrahedron j is defined by the following set of 3 equations:

$$f(\mathbf{x}_i) = f(\mathbf{x}_0) + \widehat{\nabla f} \Big|_j \cdot (\mathbf{x}_i - \mathbf{x}_0), \quad i = 1, 2, 3. \quad (6.2)$$

In the case of the velocity field the function f involves a vector field $\mathbf{v} = (v_x, v_y, v_z)$ and Eqns 6.1 and 6.2 may be written down for each of the 3 components of the velocity field. Explicitly, the velocity field is given by

$$\widehat{v}_\alpha(\mathbf{x}) = v_\alpha(\mathbf{x}_0) + \widehat{\nabla v_\alpha} \Big|_j \cdot (\mathbf{x} - \mathbf{x}_0), \quad \alpha = x, y, z. \quad (6.3)$$

Here each of the components α of the gradient is defined by the following set of three equations:

$$v_\alpha(\mathbf{x}_i) = v_\alpha(\mathbf{x}_0) + \widehat{\nabla v_\alpha} \Big|_j \cdot (\mathbf{x}_i - \mathbf{x}_0), \quad i = 1, 2, 3; \quad \alpha = x, y, z. \quad (6.4)$$

By applying the DTFE interpolation formalism the reconstructed field will have the same mathematical properties as a DTFE reconstructed density field: it is volume-covering and continuous everywhere. Its gradient is constant inside Delaunay tetrahedra, and hence discontinuous at the edges of the tetrahedra. The local resolution (smoothing kernel) is set by the size of the Delaunay tetrahedra. It leads to an optimal resolution of the reconstructed field in the sense that no additional information on the field is present inside the tetrahedra.

6.3.1 Case study: DTFE reconstruction of a velocity field

The DTFE velocity field reconstruction procedure is illustrated in Fig. 6.3. The top left-hand frame represents a discretely sampled velocity field. The length of the arrows is proportional to the velocity amplitude of each object, while the arrows' head indicates the direction of the motions. The velocity field gives the impression of a velocity flow emerging from the bottom right-hand corner of the box and spanning over the whole box.

From the locations of the sampling points the Delaunay tessellation is constructed (top right-hand frame). The computed velocity gradient for each triangle is depicted in the central frames. The gray scale in both panels is proportional to the amplitude of the determinant of the velocity gradient matrix. In the central left-hand frame a clear dark strip can be recognized, where the velocity field experiences maximal changes. The central right-hand frame represents the corresponding Delaunay surface of the gradient plot. The saddle-like shape of the surface is determined by the velocity distribution. The two maxima are located at the two extremes of the cube, while the minima form the saddle point.

After the gradient matrix has been calculated, the volume-covering velocity field can be calculated. In this case the field has been calculated on a rectangular grid of 8 grid-cells per dimension (bottom left-hand frame). For each grid-point the surrounding triangle has to be identified, indicated by the hatched triangles in the frame. The interpolation of the velocity field is linear inside each triangle, resulting in the velocity field shown in the bottom right-hand frame. This interpolated velocity field clearly displays the same characteristics as the input field: it recovers the two maxima regions located at the opposite extremes as well as the low velocity region at the bottom right-hand corner of the box. Also note that due to the linear interpolation scheme velocity information has been recovered at the top left-hand corner of the box where almost no sampling points were available.

6.3.2 Virtues and limitations

The main virtue of the DTFE over conventional reconstruction procedures is that it is intrinsically self-adaptive and does not involve artificial smoothing. Another important advantage of the DTFE procedure over conventional velocity reconstruction procedures is the fact that it produces a volume-weighted as opposed to a mass-weighted velocity field. Bernardeau & van de Weygaert (1996) have shown that this leads to critical improvements in recovering theoretical predictions. An additional advantage is that the simultaneous use of the DTFE for the reconstruction of both the density and velocity field at the same resolution allows for a coherent dynamical analysis of numerical simulations of cosmic structure formation.

The DTFE also involves a number of limitations. The first is that it assumes a linear variation inside the Delaunay tetrahedra, while in reality the cosmic density and velocity field are smoothly varying. More importantly, the DTFE assumes the field to be continuous. This implies severe limitations for the reconstruction of the velocity field in areas where the cosmic density field is non-linear and orbit crossing has occurred. In such circumstances the velocity field is not uniquely defined and the DTFE interpolation procedure is not applicable. The DTFE also involves a number of technical limitations. The first is that the sampling point distribution should not be degenerate, because in such a case the Delaunay tessellation is not uniquely defined. In practice one rarely encounters degenerate distributions. Secondly, one has to be careful about the boundary conditions. In the case where a sample is spatially

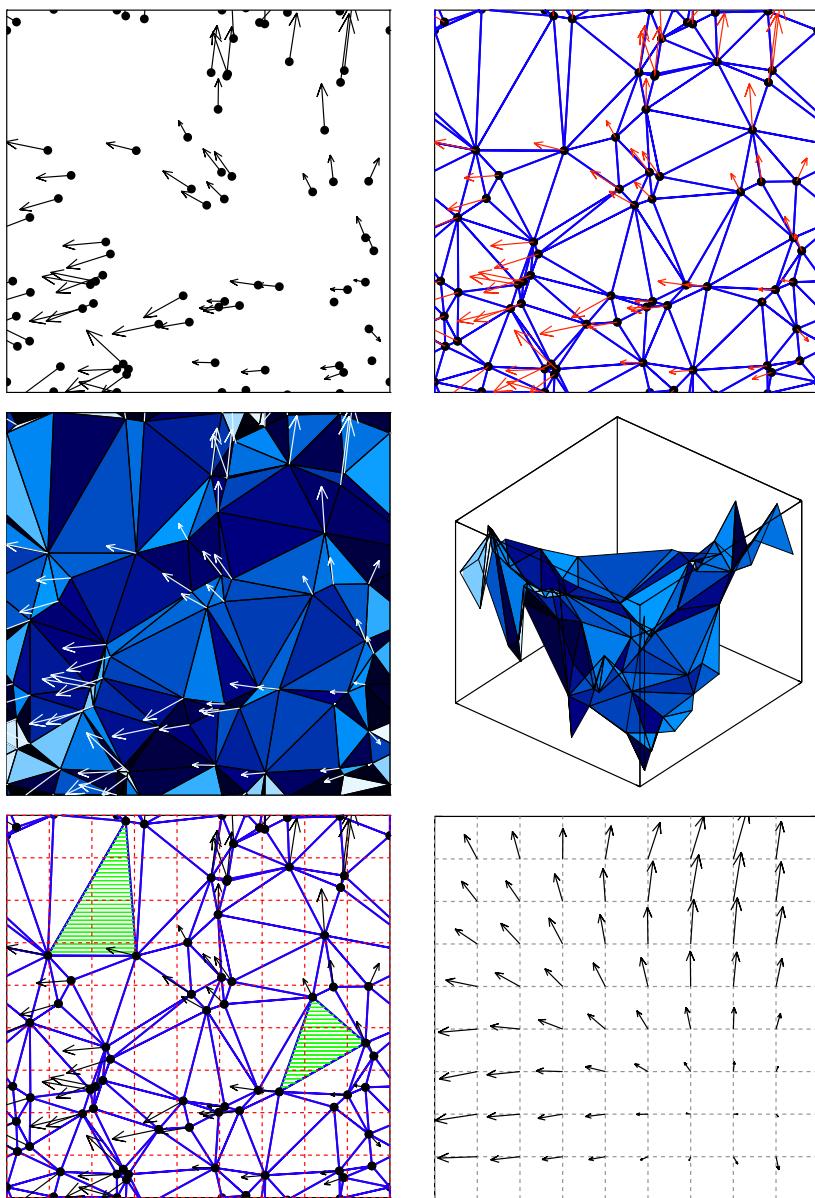


Figure 6.3 — DTDE velocity field reconstruction procedure. The top right-hand frame presents the Delaunay triangulation of the discrete particle positions of the velocity field presented in the top left-hand frame. The central frames show the velocity gradient computed for each triangle. The gray scale corresponds to the amplitude of the determinant of the gradient matrix. The right-hand frame depicts the three-dimensional representation of the gradient surface. The height of each point corresponds to its velocity amplitude. The DTDE velocity field is estimated at the grid-points indicated by the gray-colored grid of the bottom left-hand frame by assuming a linear variation of the velocity field inside each triangle. The bottom right-hand frame presents the outcome of the DTDE velocity field reconstruction procedure.

limited, the boundary points do not have natural neighbours on one side. The outermost Delaunay tetrahedra will then stretch out into infinity. In this chapter we have used periodic boundary conditions for which case this issue is not relevant.

Note that the resolution of the DTFE reconstruction procedure is dependent on the number of sampling points. Whereas variations in a reconstructed density field are directly related to variations in the local density of sampling points (we refer the reader to Chapter 3 of this thesis for a discussion of the properties of the DTFE kernel), this is not the case for the reconstructed velocity field.

6.3.3 Divergence, shear and vorticity

The DTFE velocity field reconstruction procedure involves the calculation of the nine elements of the velocity gradient (see section 6.3). From the elements of the gradient matrix it is straightforward to compute any related quantity. The DTFE estimates of the velocity divergence $\nabla \cdot \mathbf{v}$ (the trace of the velocity gradient matrix), the shear σ_{ij} (with $i, j = x, y, z$) (the symmetric and traceless part) and the vorticity ω_{ij} (its antisymmetric part) directly follow:

$$\widehat{\nabla \cdot \mathbf{v}} = \left(\frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y} + \frac{\partial v_z}{\partial z} \right), \quad (6.5)$$

$$\widehat{\sigma}_{ij} = \frac{1}{2} \left\{ \frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i} \right\} - \frac{1}{3} (\nabla \cdot \mathbf{v}) \delta_{ij}, \quad (6.6)$$

$$\widehat{\omega}_{ij} = \frac{1}{2} \left\{ \frac{\partial v_i}{\partial x_j} - \frac{\partial v_j}{\partial x_i} \right\}. \quad (6.7)$$

Here δ_{ij} represents the Kronecker delta pseudo-tensor and the vorticity ω_{ij} follows the usual definition $(\nabla \times \mathbf{v})_k = \epsilon_{ijk} \omega_{ij}$, with ϵ_{ijk} the Levi-Civita tensor.

This results in a fully volume-covering reconstruction of the divergence, shear and vorticity fields. However, notice that while the velocity field itself is continuous, the DTFE reconstruction involves a discontinuous velocity gradient. The velocity gradient is constant within each Delaunay tetrahedron and varies at the edges of the Delaunay tetrahedra. The same holds for the divergence, shear and vorticity fields.

6.4 GIF simulations: DTFE density and velocity field reconstructions

The potential of the DTFE formalism can be fully exploited for the analysis of N -body simulations. Here we illustrate this on the basis of a 256³ particles GIF N -body simulation of cosmic structure formation (Kauffmann et al. 1999). Such a simulation offers a challenging variety of objects whose characteristics are widely differing not only in density and geometry, but also with respect to their dynamical properties. The GIF simulation was performed using an adaptive P³M N -body code in a cubic box with length 141 h^{-1} Mpc and periodic boundary conditions were imposed. A concordance Λ CDM cosmology was assumed ($\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, $H_0 = 70$ km/s/Mpc).

In the left-hand frame of Fig. 6.4 a slice of thickness 1 h^{-1} Mpc through the simulation box is shown. The depicted area has a size of 141 × 141 h^{-1} Mpc and corresponds to a cosmic epoch for which the redshift $z = 3$. This redshift has been chosen to ensure that most structures are still in the linear or semi-linear phases. The right-hand frame shows a two-dimensional

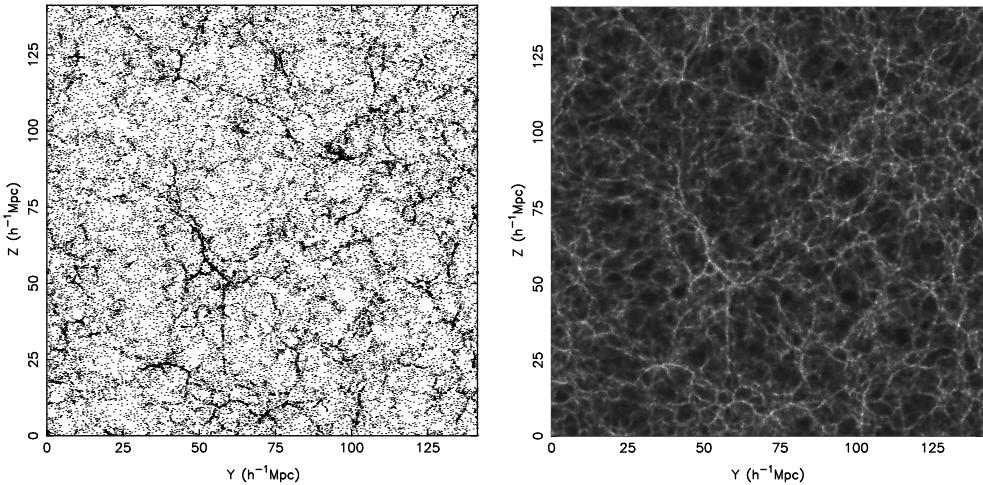


Figure 6.4 — Slice through a GIF N -body simulation (left) and the corresponding DTDE reconstructed density field (right).

slice through the three-dimensional DTDE reconstructed density field. All structural details present in the particle distribution have a corresponding counterpart in the slice through the reconstructed density field. The foam-like network of structures which the large scale matter distribution forms is clearly visible and more pronounced than in the particle distribution. It consists of such characteristic elements as extended void-like regions, compact regions of high density and filamentary structures of varying degree of anisotropy.

6.4.1 Velocity vector fields

In Fig. 6.5 three mutually perpendicular slices through the center of the three-dimensional DTDE reconstructed density field are shown, as well as similar slices through the corresponding DTDE reconstructed velocity field. The squares plotted on top of the velocity reconstructions denote interesting regions of which magnifications are shown in Fig. 6.6. In these figures we have plotted the reconstructed velocity field on a 40^3 cubic grid. Note that we may depict the velocity field at arbitrary resolution. The magnifications are therefore real magnifications of the reconstructed field in Fig. 6.5 and not distinct reconstructions. This is in stark contrast with conventional grid-based reconstructions, for which the resolution is arbitrarily set by the user and whose properties are dependent on the adopted resolution of the reconstruction procedure.

Especially when looking at Fig. 6.6 the impressive rendering of both fields becomes apparent, revealing the tight physical relation between these two fields. The top frames of Fig. 6.6 show a proto-cluster region (Kauffmann et al. 1999). The density field reconstruction shows a rather complex configuration with multiple filamentary objects connecting to the central node located at $(X, Y) = (52, 57)$. The velocity field shows the conspicuous matter flows onto the proto-cluster.

The central frames depict an expanding void whose center almost coincides with the geometrical center of its respective box. Theoretical models of voids (Icke 1984, Sheth & van de

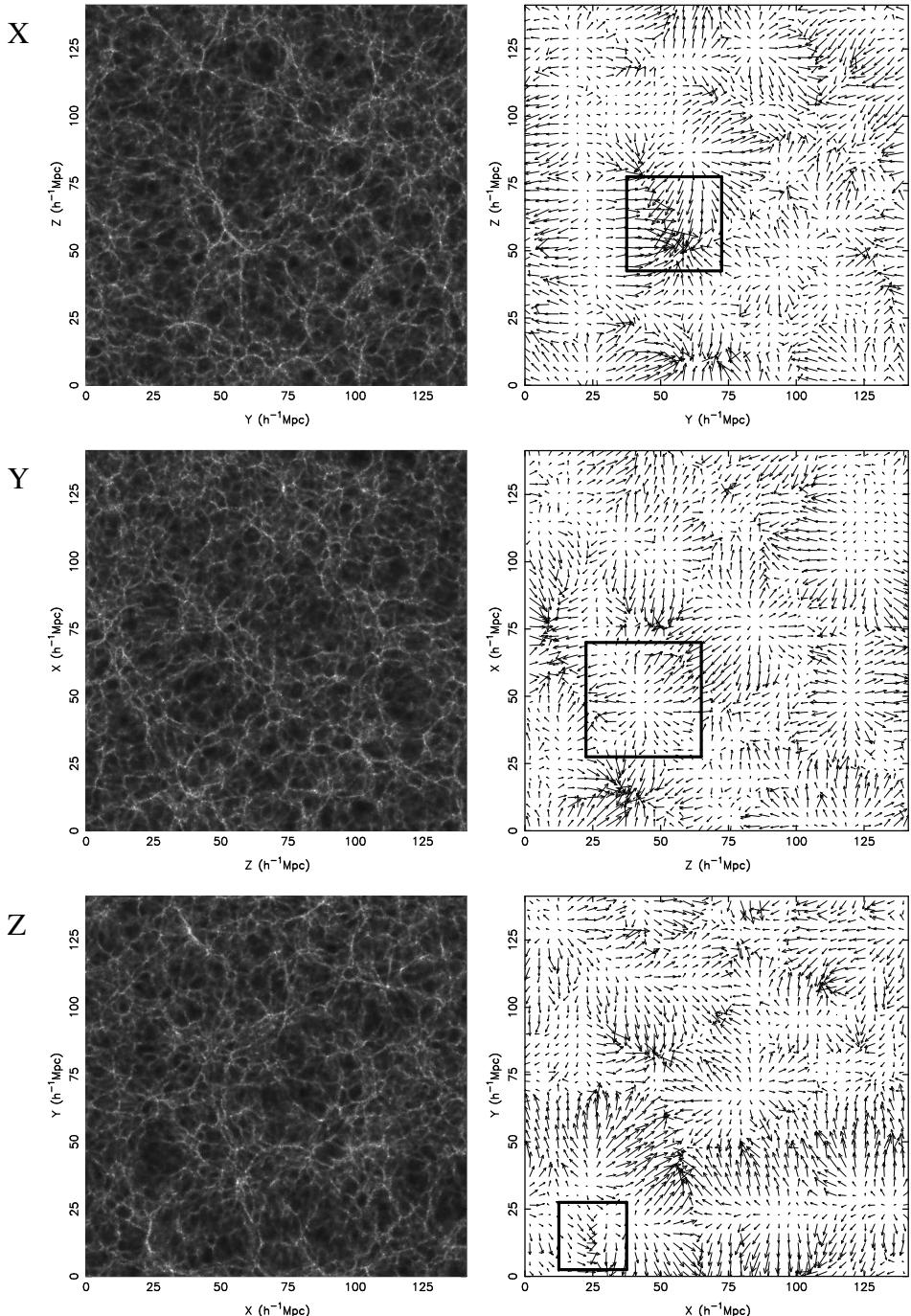


Figure 6.5 — DTFE density (left-hand column) and velocity (right-hand column) reconstructions through the GIF N -body simulation shown in Fig. 6.4. The squares plotted on top of the velocity reconstructions denote interesting regions of which magnifications are shown in Fig. 6.6

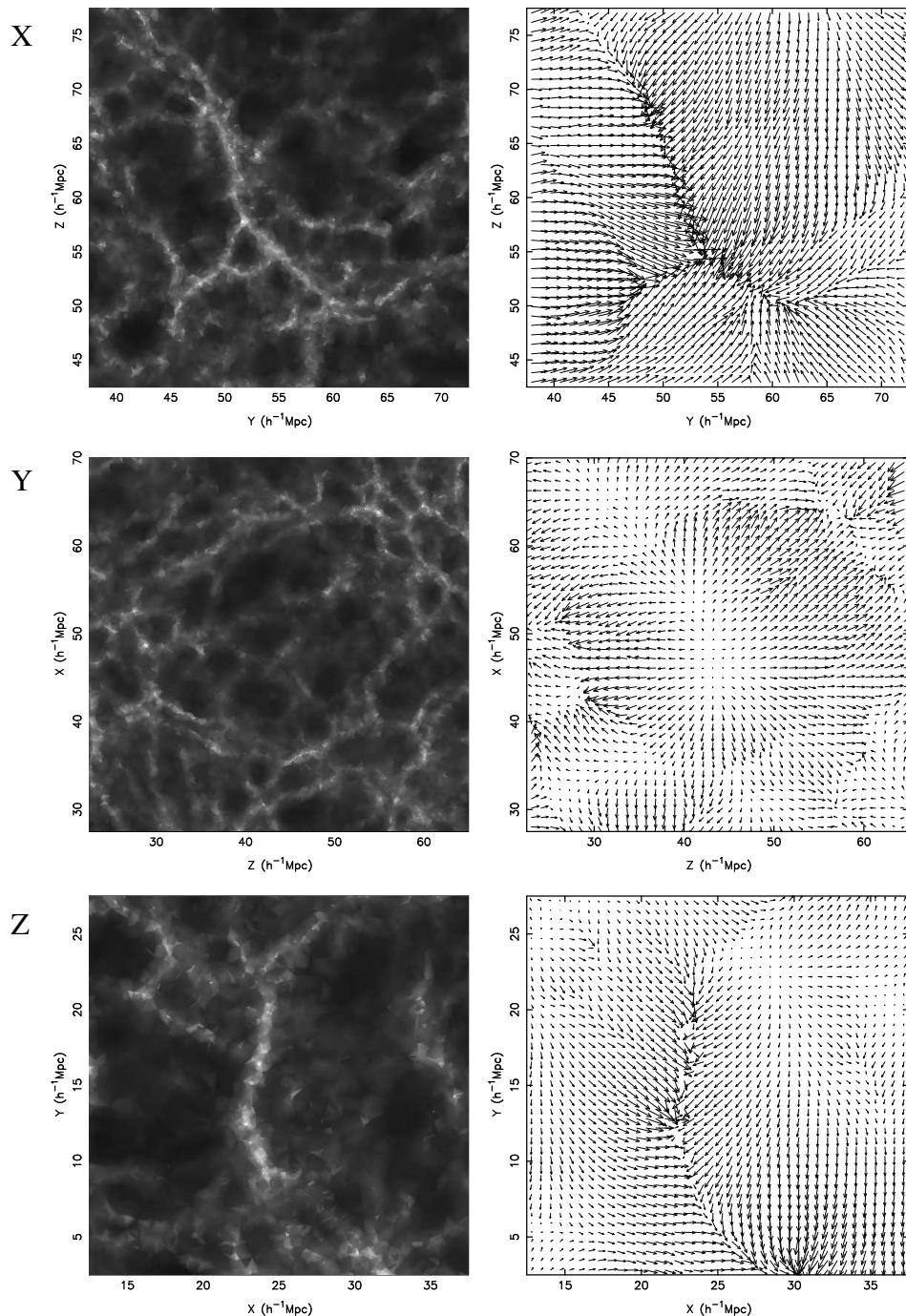


Figure 6.6 — Magnifications of a number of characteristic regions in the slices through the GIF N -body simulation shown in Fig. 6.5.

Weygaert 2004) predict that such empty regions are characterized by an inverse top-hat density profile. Their dynamics correspond to a low Ω_m universe, represented as ‘super-Hubble’ expanding bubbles with an almost constant velocity divergence. The expansion of the void is clearly recognizable in the velocity field reconstruction. Notice also that substructure is even present in the velocity field of the void. In section 6.5 we analyze the dynamics of these void regions and test how well the analytic predictions are reproduced by the DTFE reconstructions.

The bottom frames show a filamentary feature running almost parallel to the Y -axis along the center of its respective box. The velocity field shows that matter is flowing from its surroundings onto the filament. A clear shearing flow can be identified along the filament’s ridge. Notice however also the strong flow towards the clump of matter located at $(X, Y) = (30, 2.5)$. In section 6.6 we discuss the dynamics of filamentary structures in more detail.

A further test of the combined DTFE density and velocity reconstruction is to compare the densities and velocities directly. In Fig. 6.7 the density and velocity profiles are shown along a one-dimensional section through the simulation box. Plotted on top of the particle distribution and the slice through the reconstructed density field is a thick solid line, along which the density and velocity field as a function of distance have been reconstructed with the DTFE procedure. The resulting density and velocity profiles are plotted in the bottom frame of Fig. 6.7. Only the velocity component parallel to the direction of the cut is shown. The density scale is plotted on the left, while the velocity scale is plotted on the right ordinate.

A visual inspection of the density and velocity profiles plotted in Fig. 6.7 shows that the DTFE reconstruction procedure is capable of recovering all the structures present in the point distribution. High density peaks show up at places where the profile intersects with filamentary structures or proto-clusters visible in the particle distribution. There the velocity drops almost instantaneously (notice e.g. the peaks at about $7h^{-1}\text{Mpc}$ and $128h^{-1}\text{Mpc}$ along the cut). The drop in velocity indicates infalling motions onto these structures. Conversely, empty regions are rendered as slowly-varying regions of low density. The velocity appears to increase almost linearly across these regions (notice e.g. the void centered at about $124h^{-1}\text{Mpc}$ along the cut). This linear increase reflects the expansion of voids. The specific properties of voids and filamentary structures are analyzed in sections 6.5 and 6.6. Here we note that the DTFE reconstructed density and velocity fields are indeed tightly correlated and appear to give a realistic image of the large scale structure environment. This result has been obtained without having to invoke any subsequent filtering.

6.4.2 The continuity equation

The velocity divergence and the density contrast are related through the continuity equation (Peebles 1980). Within linear theory this relationship is given by

$$\nabla \cdot \mathbf{v}(\mathbf{x}) = -H f(\Omega_m, \Omega_\Lambda, z) \delta(\mathbf{x}). \quad (6.8)$$

Here $H = H(z)$ is the Hubble parameter, Ω_m the cosmic density parameter at the current epoch and Ω_Λ the cosmic vacuum energy density parameter at the current epoch. The function f may be approximated by $f(z) \approx \Omega_m^{0.6}(z)$ (Lahav et al. 1991). In terms of the cosmological parameters at the present epoch this approximation can be written as

$$f(z) \approx \left[\frac{\Omega_m(1+z)^3}{\Omega_m(1+z)^3 - (\Omega_m + \Omega_\Lambda - 1)(1+z)^2 + \Omega_\Lambda} \right]^{0.6} \quad (6.9)$$

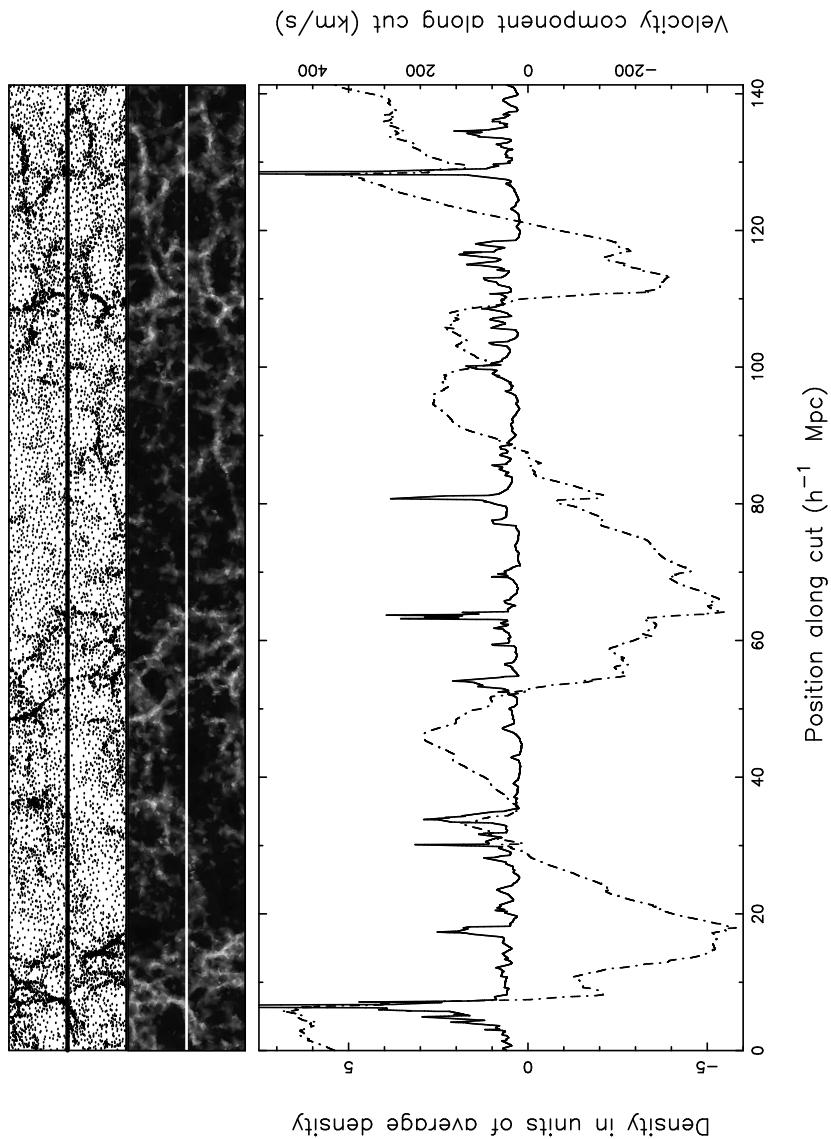


Figure 6.7 — Slice through a GIF N -body simulation (top), the DTFE reconstructed density field (center) and density and velocity profiles along the central line (bottom).

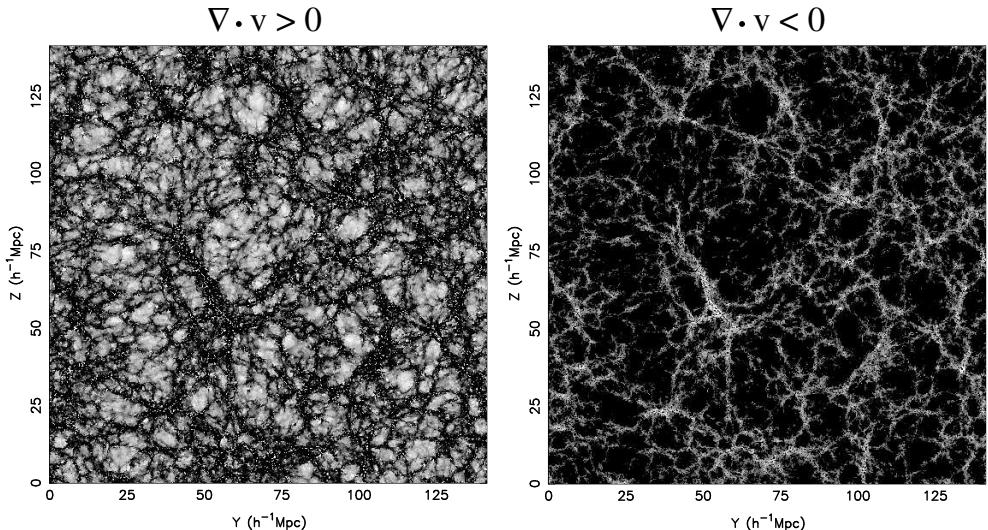


Figure 6.8 — Velocity divergence maps. The plotted slices correspond to the slice shown in Fig. 6.4. The left-hand frame shows regions with positive divergence, i.e. with outflowing motions. The right-hand frame shows regions with negative divergence, i.e. with infalling motions.

A lot of theoretical effort has been devoted to obtain accurate density-velocity divergence relations beyond the linear regime (see the review of Bernardeau et al. 2002). On the basis of perturbation theory Bernardeau (1992) derived a second-order relationship for the quasi-linear regime,

$$\frac{\nabla \cdot \mathbf{v}(\mathbf{x})}{H} = \frac{3}{2} f(\Omega_m, \Omega_\Lambda, z) [1 - (1 + \delta(\mathbf{x}))^{2/3}]. \quad (6.10)$$

Other authors have derived even higher-order corrections (e.g. Bernardeau et al. 1999, Kudlicki et al. 2000, Bernardeau et al. 2002). As the DTDE is capable of simultaneously recovering the cosmic density and velocity fields at the same resolution, the one-to-one relationship between the velocity divergence and the density contrast provides a unique opportunity of testing the capacity and consistency of the DTDE procedure with respect to probing the dynamics in cosmic structure formation scenarios.

Fig. 6.8 gives a first impression of the strong relationship between the density and velocity divergence fields as reconstructed by the DTDE. The figure shows the velocity divergence for the slice presented in Fig. 6.4. For a better appreciation of the correspondence between high and low density regions, we have splitted the divergence field into two parts, positive and negative divergence regions. No additional smoothing has been applied to the density and velocity reconstructions. In the right-hand frame regions with negative divergence are shown. They clearly correspond to regions where infall motions are present, such as matter accretion onto clusters and filaments. In the left-hand frame regions with positive divergence are shown. These regions are expanding and may be identified with low density voids. The substructures which are visible in the density fields inside voids (see Fig. 6.4) have conspicuous counterparts in the velocity divergence field, where they present themselves as regions of positive divergence.

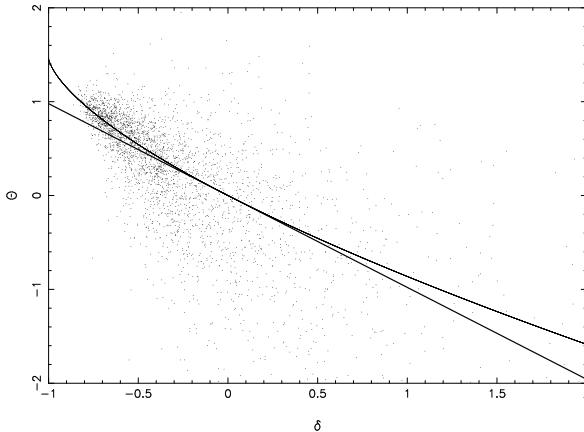


Figure 6.9 — The relationship between the velocity divergence and the density contrast. The points are DTRE reconstructed field values. The continuous lines represent predictions from linear theory (thick solid line) and a second-order approximation (thin solid line). The velocity divergence θ is expressed in units of the Hubble parameter.

Fig. 6.9 shows a scatter plot of the relationship between the velocity divergence and the density contrast for the GIF simulation shown in Fig. 6.4. Each dot corresponds to the DTRE reconstructed density contrast and velocity divergence at the center of gravity of a given Delaunay tetrahedron. The thick continuous line represents the prediction from linear theory (Eqn. 6.8), while the thin continuous line represent the prediction from a second-order approximation (Eqn. 6.10). The DTRE reconstructed fields have not been smoothed.

The agreement of the DTRE reconstructions with the predictions from theory is remarkably good, especially when considering the fact that we have not smoothed the fields at all. It appears that the second-order approximation is a better fit to the data than linear theory, which is not surprising given the relatively advanced dynamical state of the cosmic network and the fact that we have not smoothed the data. Nevertheless there is a large amount of scatter present in the data. This scatter is present for several reasons:

1. Physical deviations from (semi-)linearity. Clearly both the low density voids as the intermediate and high density filaments and clusters have entered the non-linear regime. In this regime the one-to-one relation between the density and the velocity relation is not valid.
2. Inconsistency of a combined DTRE density and velocity reconstruction. Strictly speaking such a combined reconstruction is not self-consistent because the DTRE implies a linear density and velocity field, whereas the continuity equation demands a constant density field for a linearly varying velocity field.
3. Poisson scattering of density field values. A Poisson sampling of the density field will lead to the presence of sampling noise in the density field (which is larger for higher density values). The velocity field is not beset by this effect because the values at the location of the sampling points are exact.

Given the presence of these effects, it is remarkable how well the data fits the theoretical predictions. This clearly indicates that the combined DTFE density and velocity reconstruction works very well. Note that each of these effects may be diminished by smoothing over the data. Smoothing of the data does indeed lead to a tighter correlation at the price of a more linear relation (e.g. Bernardeau & van de Weygaert 1996, Romano-Díaz 2004).

6.5 Voids

By nature void-like regions contain very few galaxies or simulation particles. This poses specific problems for the analysis of their density and velocity fields in observations or numerical simulations. Conventional methods are in general unable of accurately describing the density and velocity field inside voids over the whole range of relevant scales. The rigid non-adaptive nature of grid-based schemes leads to a loss of information in the reconstructed field on scales smaller than the size of the effective kernel or smoothing radius. Anisotropic and caustic features will therefore not be recovered if their scale is smaller than the size of the grid-cells. Also, if the resolution of the grid is set too high, the sampling particle density will on average be less than one particle per cell. In such a case the interpolated fields will be dominated by shot-noise. In the case of the velocity field the problems are even more severe than for the density field, because the sampling is based on variations in the density field. This means that in low density environments the velocity field is severely undersampled and previous studies had to rely on heavy smoothing or other specifically designed techniques in order to obtain any useful information in such environments see (e.g. van de Weygaert & van Kampen 1993, Dubinski et al. 1993). Although SPH-like procedures whose resolution adapt to the local density of sampling points certainly do represent a significant improvement over the traditional methods, SPH-kernels are almost exclusively spherically symmetric (see however Shapiro et al. 1996 and Owen et al. 1998) and do not adapt to the local geometry of the point distribution. In Chapters 4 and 5 we have studied the differences between DTFE and SPH-like reconstruction methods and showed that SPH-like methods are not well-suited for studying anisotropic structures like the filaments and wall-like structures present in the large scale galaxy distribution. The SPH kernel has a somewhat larger effective smoothing kernel and thus a lower resolution. The spherical smoothing tends to blur out anisotropic structures into their surrounding environment, affecting the outer parts of voids. We have shown that DTFE reconstructed fields are not affected by such problems.

6.5.1 Structure and dynamics

In Fig. 6.10 a typical void-like region is shown, together with the DTFE density and velocity field reconstructions. The solid line running from the bottom to the top of these fields indicates the one-dimensional section along which the density and velocity field are shown in the bottom right-hand frame of the figure. The DTFE procedure clearly manages to render the void as a realistically slowly varying region of low density. Notice the clear distinction between the empty (dark) interior regions of the void and its edges. In the interior of the void several smaller ‘sub-voids’ may be recognized, whose boundaries consist of low density filamentary or wall-like structures. The presence of a hierarchy of voids, with large voids composed of the merging of smaller ones is in agreement with theories of void evolution (Regos & Geller 1991, Dubinski et al. 1993, van de Weygaert & van Kampen 1993, Sheth &

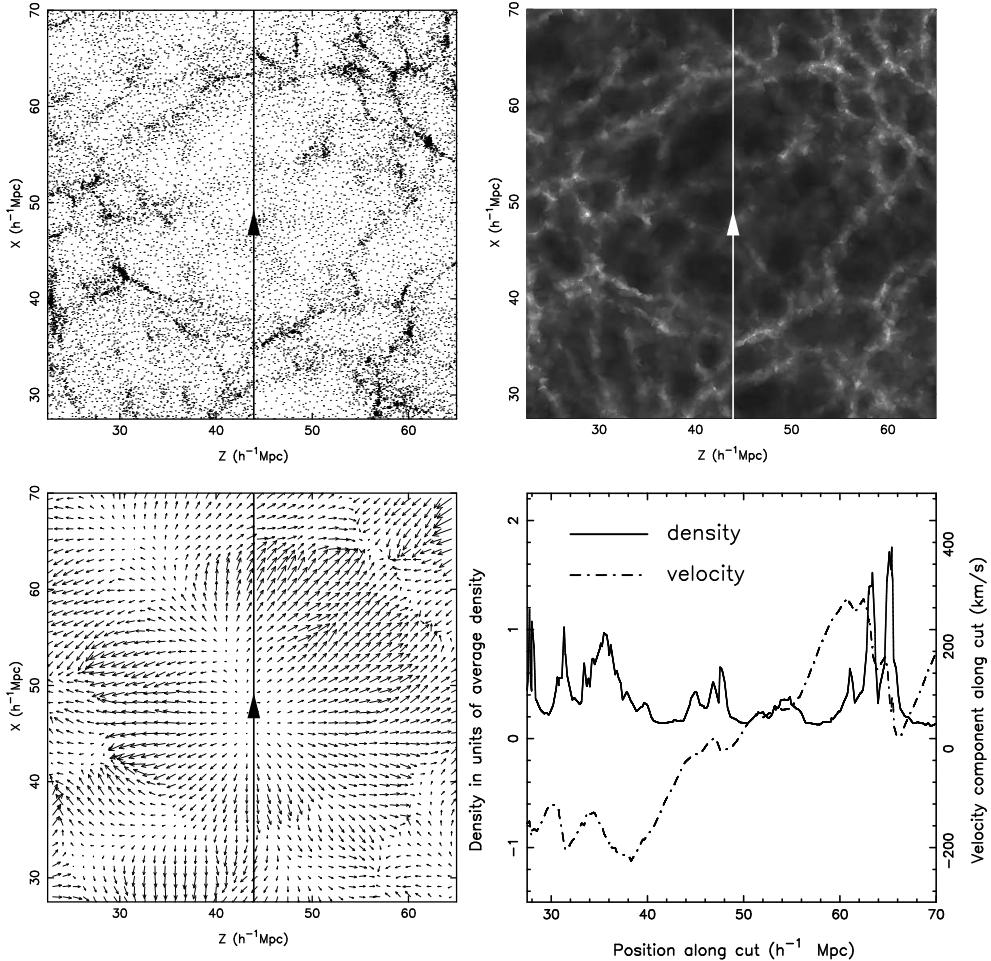


Figure 6.10 — A typical void-like region in the simulation shown in Fig. 6.7. Top left-hand frame: particle distribution in a thin slice through the simulation box. Top right-hand frame: two-dimensional slice through the three-dimensional DTFE density field reconstruction. Bottom left-hand frame: two-dimensional slice through the three-dimensional DTFE velocity field reconstruction. Bottom right-hand frame: density and velocity reconstructions along the thick line shown in the other frames.

van de Weygaert 2004). The velocity field shows that the void is expanding. This expansion is rather uniform as can be observed in the one-dimensional sections through the density and velocity reconstruction shown in the bottom right-hand frame of Fig. 6.10.

The reconstructed velocity field shows a consistent behavior over the entire void region. The velocity increases approximatively linearly along the one-dimensional profile, while it suddenly drops after crossing the edge of relatively high density. This linear ‘super-Hubble’ expansion of voids is well understood in terms of gravitational dynamics. According to Birkhoff’s theorem the internal dynamics of a spherically symmetric system is independent of the dynamics of the outside universe (Birkhoff 1923). This theorem has been applied in the development of spherically symmetric infall models for galaxies (e.g. Gunn & Gott 1973, Schechter 1980), large-scale inhomogeneities in general (Silk 1974) and cluster infall regions (Regos & Geller 1989). According to Birkhoff’s theorem voids can be approximated as expanding, isolated universes unto themselves that do not accrete matter from the universe at large (e.g. van de Weygaert & van Kampen 1993, Goldberg & Vogeley 2004). Because voids are emptier than the rest of the universe they will expand faster than the rest of the universe with a net velocity divergence equal to

$$\frac{1}{3} \nabla \cdot \mathbf{v} = H_{\text{void}} - H. \quad (6.11)$$

When we define the normalized velocity divergence $\theta = \nabla \cdot \mathbf{v}/H$ and the ratio of the Hubble expansion of the void universe and the Hubble expansion of the universe $\alpha = H_{\text{void}}/H$ this equation may simply be written as

$$\theta = 3(\alpha - 1). \quad (6.12)$$

The super-Hubble expansion of voids has been observed in numerical simulations of void-like regions (e.g. van de Weygaert & van Kampen 1993, Dubinski et al. 1993), be it that a large amount of artificial smoothing or other specifically designed techniques had to be imposed to derive credible results. The DTRE reconstruction yields natural and continuous void density and velocity profiles without the application of smoothing procedures and thus allows a study of its physical structure and dynamics.

6.5.2 Constraining cosmological parameters using void dynamics

The expansion velocity of voids is directly related to the cosmology of the background universe, which determines the expansion ratio α . The velocity divergence inside voids follows from Eqn. 6.12. This relation thus allows the possibility of using the largest voids to constrain the cosmic density parameter (van de Weygaert & van Kampen 1993, Dekel & Rees 1994). Bernardeau et al. (1997) used the cut-off and the peak of the probability distribution function (PDF) of the velocity divergence field to measure the normalized velocity divergence of the largest voids to constrain the cosmic density parameter (see also Bernardeau 1994, Bernardeau et al. 1995).

6.5.2.1 Empty voids in cosmologies with no cosmological constant

The largest expansion ratios will occur for the emptiest voids. For cosmologies with no cosmological constant the age of the universe may be approximated by (see e.g. Peacock 1999)

$$t \approx H^{-1} \left(1 + \frac{1}{2} \Omega_m^{0.6} \right)^{-1}. \quad (6.13)$$

According to Birkhoff's theorem the largest voids behave like empty FRW universes and their age is therefore equal to

$$t_{\text{void}} = H_{\text{void}}^{-1}. \quad (6.14)$$

Setting the age of voids equal to the age of the universe one finds after some algebraic manipulations for the expansion ratio of empty voids,

$$\alpha = 1 + \frac{1}{2} \Omega_m^{0.6}. \quad (6.15)$$

6.5.2.2 Empty voids in cosmologies with a non-zero cosmological constant

For cosmologies with a non-zero cosmological constant Eqn. 6.15 is not valid. Not only is the age of the universe different from the value given by Eqn. 6.13, which evidently is true for voids too. The dynamics of voids is also affected by the cosmological constant and their age will differ from the value given by Eqn. 6.14. Romano-Díaz (2004) nevertheless applied Eqn. 6.15 to measure the cosmological density parameter in a Λ CDM cosmology, which may explain part of the differences between their measurements and the imposed value. The age of the universe with a non-zero cosmological constant can be approximated by (Carroll, Press & Turner 1992)

$$t \approx \frac{2}{3} H^{-1} (0.7 \Omega_m - 0.3 \Omega_\Lambda + 0.3)^{-0.3}. \quad (6.16)$$

Similarly, the age of voids can be approximated by

$$t_{\text{void}} \approx \frac{2}{3} H_{\text{void}}^{-1} (-0.3 \Omega_{\Lambda, \text{void}} + 0.3)^{-0.3}. \quad (6.17)$$

Here $\Omega_{\Lambda, \text{void}}$ is the vacuum energy density parameter inside voids. Although the cosmological constant Λ clearly is the same inside voids as inside the background universe, the corresponding vacuum energy density parameters are not. The reason for this is that voids are expanding faster than the background universe and the Hubble parameter is correspondingly larger. It follows that the value of the vacuum energy density parameter inside empty voids is given by

$$\Omega_{\Lambda, \text{void}} = \frac{\Lambda}{3H_{\text{void}}^2} = \Omega_\Lambda \frac{H^2}{H_{\text{void}}^2} = \Omega_\Lambda \alpha^{-2}. \quad (6.18)$$

Equating the age of voids with the age of the universe leads to the following expression:

$$\alpha(-\Omega_\Lambda \alpha^{-2} + 1)^{0.3} = (2.33 \Omega_m - \Omega_\Lambda + 1)^{0.3}. \quad (6.19)$$

This equation has to be solved numerically for the expansion ratio α . The velocity divergence in the interior of voids can subsequently be found by applying Eqn. 6.12.

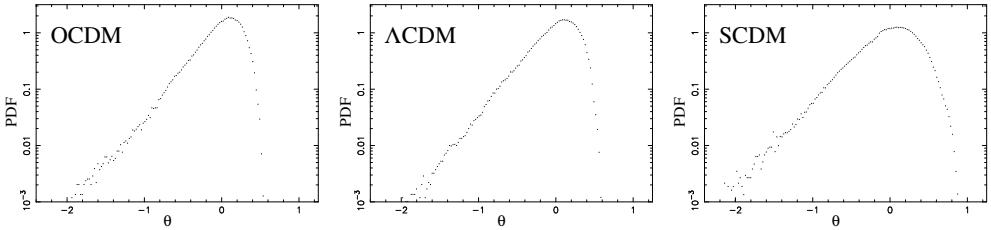


Figure 6.11 — Probability distribution functions (PDFs) of the normalized DTDE velocity divergence θ at actual time for three different GIF simulations (see text for description). The velocity divergence field has been smoothed with a Gaussian kernel with $R_G = 5h^{-1}\text{Mpc}$.

Table 6.1 — Parameters of the GIF simulations.

Model	Ω_m	Ω_Λ	h	$L (h^{-1}\text{Mpc})$
SCDM	1.0	0.0	0.5	84.5
OCDM	0.3	0.0	0.7	141.3
Λ CDM	0.3	0.7	0.7	141.3

We have tested this formalism for a number of cosmologies by means of cosmological N -body simulations. Following Bernardeau et al. (1997) we have determined the velocity divergence inside empty voids by measuring the position of the cut-off of the probability distribution function (PDF) of the velocity divergence field. In Fig. 6.11 the PDFs of the normalized velocity divergence θ at the present epoch is shown for three GIF-simulations, each corresponding to a different cosmology (see Table 6.1 for their parameters). They correspond to a standard cold dark matter model (SCDM), an open model (OCDM) and a concordance model (Λ CDM). The velocity divergence field has been computed by means of the DTDE procedure. Negative values of the velocity divergence correspond to overdense regions where inflows occur, while positive values correspond to underdense regions where outflows occur. The strongest outflows occur for the emptiest voids.

For all three simulations the PDF clearly deviates from Gaussianity and shows a pronounced negative skewness. This is because in the non-linear regime the inflows to high density regions are faster than the outflows from low density regions. Notice that the shapes of the OCDM and the Λ CDM PDFs are very similar, while the SCDM PDF appears very different. In particular the SCDM PDF is non-zero for higher values of the velocity divergence, while negative values of the velocity divergence are much more prevalent. This difference is due to the fact that the SCDM model contains more mass than both the OCDM and Λ CDM models, resulting in stronger infall and outflow motions. The predicted sharp cut-off at positive values of θ is clearly present in all three models. To obtain this sharp cut-off we have smoothed the velocity divergence field with a Gaussian filter of radius $R_G = 5h^{-1}\text{Mpc}$. This smoothing has been applied to obtain a uniform scale for the measured divergence values and because the non-linear velocity field inside high density environments such as clusters and filaments cannot be represented by the DTDE interpolation procedure, which assumes a linear variation of the field between neighbouring sampling points. Because of visualization and orbit cross-

Table 6.2 — Comparison of measured and theoretical expansion ratios. Expansion ratios are given both for voids modeled as empty universes ($\alpha_{\text{theory, empty}}$) and as low density universes ($\alpha_{\text{theory, low density}}$)

Model	θ_{\max}	$\alpha_{\text{theory, empty}}$	α_{measured}	δ_{\min}	$\alpha_{\text{theory, low density}}$
SCDM	0.90	1.50	1.30	0.19	1.32
OCDM	0.55	1.24	1.18	0.11	1.18
Λ CDM	0.59	1.21	1.20	0.065	1.20

ing the velocity field is not uniquely defined in these regions, which leads to non-physical estimates of the velocity divergence.

We have measured the position of the maximum velocity divergence by determining the position of the cut-off in the PDFs shown in Fig. 6.11). The results are listed in Table 6.2. By means of Eqn. 6.12 we have calculated the expansion ratios α_{measured} of these empty voids with respect to the corresponding background cosmologies. We have compared these measurements with the analytical predictions α_{theory} calculated from Eqn. 6.15 (SCDM and OCDM models) and Eqn. 6.19 (Λ CDM model). The results are also listed in Table 6.2.

The measured expansion ratios are substantially lower than predicted for the SCDM and the OCDM models, while for the Λ CDM model the measured value coincides with the theoretical prediction. The fact that we find a too low expansion ratio for the SCDM and OCDM models may be due to the fact that the measured expansion ratios should be interpreted as lower bounds, as we have considered empty voids and in reality voids are not empty. The simulations do not contain completely empty voids, which instead should be approximated by low density universes.

6.5.2.3 Non-empty voids

The density of non-empty voids can be parametrized with a non-zero cosmic density parameter $\Omega_{m,\text{void}}$ given by

$$\Omega_{m,\text{void}} = \Omega_m \frac{H^2}{H_{\text{void}}^2} (\bar{\delta}_{\text{void}} + 1) = \Omega_m \alpha^{-2} (\bar{\delta}_{\text{void}} + 1). \quad (6.20)$$

Here $\bar{\delta}_{\text{void}} + 1$ is the mean underdensity inside the void. The age of a non-empty void in a universe without a cosmological parameter is given by Eqn. 6.13 with Ω_m replaced by $\Omega_{m,\text{void}}$. The age of a non-empty void in a universe with a non-zero cosmological parameter is given by Eqn. 6.16, again with Ω_m replaced by $\Omega_{m,\text{void}}$.

To check if this explains the differences between the measured and predicted expansion ratios, we have measured the mean underdensity inside the most empty void for the three GIF simulations. These values are listed in Table 6.2. The SCDM universe has the highest minimal density, in correspondence with the fact that this model has the highest matter density. The OCDM and Λ CDM universe have an equal matter density, but the density of the emptiest void is substantially lower in the Λ CDM universe. This is due to the extra expansion caused by a non-zero cosmological parameter. We have calculated the expansion ratio for the three GIF simulations. These are also listed in Table 6.2. The agreement with the analytically predicted values is now very good for all three models. This shows that voids can be modeled

Table 6.3 — Evolution of measured and theoretical expansion ratios for a characteristic void from $z = 5$ to $z = 0$. The void is modeled as a low density universe.

z	$H(z)$	$\Omega_m(z)$	$\Omega_\Lambda(z)$	$\bar{\delta}_{\text{void}}$	a_{theory}	$\nabla \cdot \mathbf{v}$ [10 ² km/s]	$a_{\text{measured}} \equiv \nabla \cdot \mathbf{v} H(t)$
5.0	$5.7 \cdot 10^2$	0.99	0.01	0.38	1.25	4.4	1.26
3.0	$3.1 \cdot 10^2$	0.96	0.04	0.29	1.29	3.0	1.33
2.0	$2.1 \cdot 10^2$	0.92	0.08	0.24	1.31	2.2	1.36
1.0	$1.2 \cdot 10^2$	0.77	0.23	0.16	1.31	1.4	1.38
0.5	92	0.59	0.41	0.12	1.29	0.96	1.35
0.3	82	0.49	0.51	0.10	1.26	0.77	1.32
0.0	70	0.30	0.70	0.09	1.19	0.53	1.25

as spherically symmetric low density universes. Moreover, the DTFE combined density and velocity reconstruction procedure clearly provides a consistent description of voids.

6.5.3 Evolution

The description of voids as spherically symmetric low density universes is valid not just for most empty voids at the present epoch, but should in principle apply to any void at any cosmic time. To explore and test this notion we have followed the evolution of one particular void. In Fig. 6.12 we show the same typical void-like region as depicted in Fig. 6.10 at a number of cosmic times ranging from $z = 5$ to $z = 0$. At each cosmic time the particle distribution in a thin slice is shown together with the DTFE density and velocity field reconstructions, as well as one-dimensional sections through the density and velocity fields. The figure shows the evolution of the void from a region which is relatively underdense to a region which is almost empty. The expansion of the void is clearly visible as it gets almost twice as big (in comoving coordinates) between $z = 5$ and $z = 0$. The expansion velocity decreases as the void expands. This can be seen in the one-dimensional sections through the velocity field.

At each cosmic time we have measured the average density inside the void, which together with the cosmological parameters provide a theoretical prediction of the expansion ratio α_{theory} of this void with respect to the background cosmology. We have also measured the average velocity divergence inside the void, which translates into a measurement of the expansion ratio α_{measured} of the void. The predicted and measured expansion ratios are listed in Table 6.3. Given the unknown accuracy of the assumption of spherical symmetry of the void the theoretical and measured values match strikingly well. On the other hand, the measured expansion ratio is systematically about 5% larger than the predicted value. This may be due to the tidal forces of the large scale environment outside the void, which may induce a positive velocity divergence. This explanation could be tested by subtracting the model velocity field for the void from the DTFE reconstructed velocity field and checking whether the resulting residual field can be explained by features present in the density field. Such an analysis is beyond the scope of this chapter.

Given the close agreement between the measured and predicted values it should in principle be possible to constrain the values of the cosmological parameters from the density and

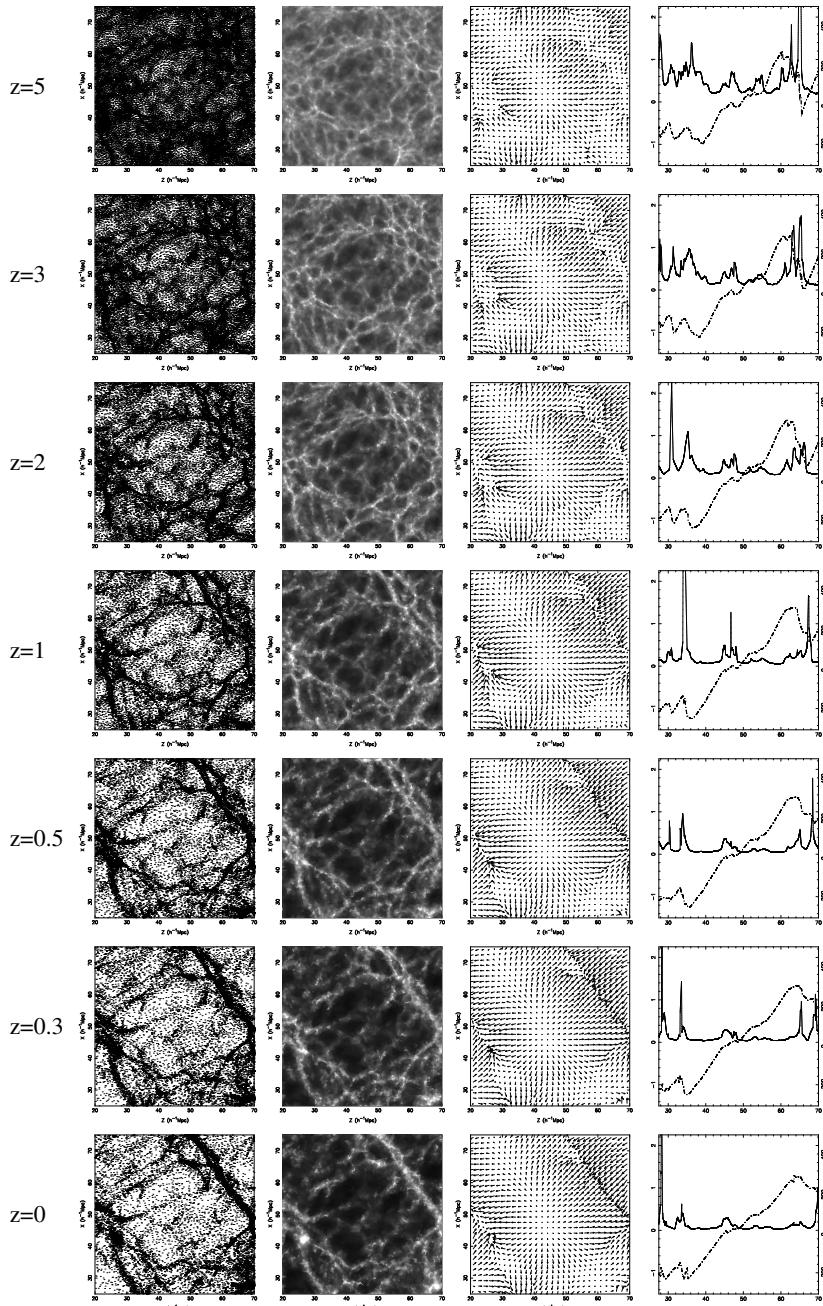


Figure 6.12 — Evolution of a void from $z = 5$ to $z = 0$. The void shown is the same as depicted in Fig. 6.10. At each cosmic time the particle distribution is shown in a thin slice together with a two-dimensional slice through the full three-dimensional DTFE reconstructed density and velocity fields and one-dimensional sections through these fields. The sections are taken along the same directions as shown in Fig. 6.10.

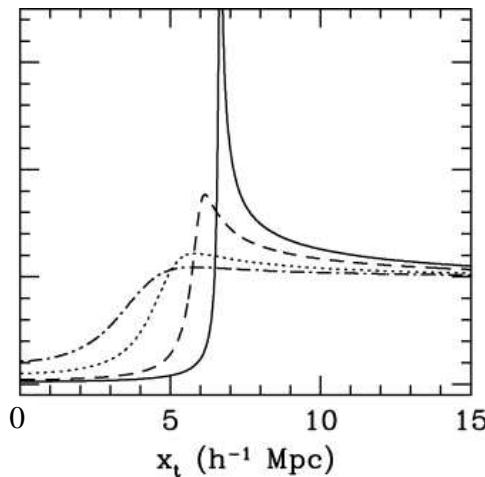


Figure 6.13 — Spherical model for the evolution of voids. A void with an angular averaged SCDM profile evolving up to the epoch of shell crossing. Time steps: $a = 0.05, 0.1, 0.2$ and 0.3 . Source: Sheth & van de Weygaert 2004.

velocity fields in voids. The DTNE technique is clearly able of consistently representing these fields even though relatively few sampling particles are present in these low density environments.

We have modelled voids as uniform low density universes. A more realistic model would take into account that voids do not have a uniform density. Instead, they may be modeled as a series of evolving concentric low density shells. These shells remain concentric and are assumed to be perfectly uniform, without any substructure. The resulting solution of the equation of motion for each shell covers the full non-linear evolution of the void, as long as shell crossing does not occur. This ‘spherical model’ has been described and worked out by Gunn & Gott (1972) and Lilje & Lahav (1991).

In Fig. 6.13 (source: Sheth & van de Weygaert 2004) the evolution of a void in the spherical model is illustrated. It shows the time evolution of the density deficit profile. The evolving density profile bears out the characteristic tendency of voids to expand, with mass streaming out from the interior, and hence for the density to decrease continuously in value (and approach emptiness, $\delta = -1.0$). Notice that this model provides the most straightforward illustration of the formation of a ridge. Looking from the inside out, one sees the interior shells expanding outwards more rapidly than the outer shells. With time the inner matter catches up with the outer shells, leading to a steepening of the density profile in the outer realms and the formation of a ridge. Meanwhile, over a growing area of the void interior, the density distribution is rapidly flattening. This is a direct consequence of the outward expansion of the inner void layers: the ‘flat’ part of the density distribution in the immediate vicinity of the dip gets ‘inflated’ along with the void expansion (Sheth & van de Weygaert 2004).

These characteristics are also visible in Fig. 6.12. At high redshifts the void boundaries are not very conspicuous and only slightly more dense than their surroundings. As matter streams outwards and starts to collect at the boundaries of the voids a ridge is formed, which is clearly

visible at low redshifts. At high redshifts, the void interior still contains a considerable amount of matter and structure, while it empties and flattens almost completely at low redshifts. Note that the void density profile in Fig. 6.12 is not as smooth as the spherical model in Fig. 6.13. The latter is a spherically symmetric model, while in reality voids are not.

6.6 Filaments and walls

Filamentary and wall-like structures are difficult to describe using conventional analysis techniques. These structures are highly anisotropic in one or two spatial dimensions and conventionally used spherically symmetric smoothing kernels are therefore inappropriate. In Chapters 4 and 5 of this thesis we have shown that such kernels tend to blur out anisotropic structures and as a result they obtain a larger volume and become intrinsically less anisotropic. The DTFE kernel is not beset by these problems. It automatically adapts to both the local density and geometry of the distribution of sampling points, resulting in a much more faithful representation of filaments and walls.

6.6.1 Structure and dynamics

In Fig. 6.14 a typical filamentary structure is shown, together with the DTFE density and velocity field reconstructions. The solid line running from the bottom left-hand to the top right-hand part of these fields represents the one-dimensional section along which the density and velocity field are shown in the bottom right-hand frame of the figure. The DTFE procedure renders the filament as a distinct feature in both the density and velocity field. Clearly visible is the shearing flow along the filaments ridge. The one-dimensional section slices roughly perpendicularly through the filament, which shows up as a sharp double peak in the density reconstruction. At the same time the value of the velocity field drops about 900 km/s at the location of the filament. This decrease corresponds to infalling motions towards the filament due to its gravitational pull on its surroundings. Clearly noticeable to the left and right of the filament is the linear super-Hubble expansion of the surrounding voids.

It is known that anisotropic collapse of matter in gravitational instability scenarios leads to the formation of sheets and filaments, which can already be predicted from the Zel'dovich approximation (Zel'dovich 1970, see also Shandarin & Zel'dovich 1989). In Fig. 6.15 the density and velocity profile of a Zel'dovich pancake is shown (source: Shandarin & Zel'dovich 1989). The density and velocity profiles look very similar to those plotted in Fig. 6.14.

A fully non-linear model of the formation and evolution of cosmic walls and filaments is that of isolated homogeneous ellipsoidal overdensities. In particular, the early work by Icke (1972, 1973) elucidated transparently the crucial characteristics of their development and morphology. He showed that any slight asphericity in the initial density field will inevitably amplify during its subsequent evolution and collapse. The work of Icke was followed up upon by White & Silk (1979) who specifically studied the formation of isolated ellipsoidal structures in general background cosmologies (with no cosmological constant). A lot of work has been devoted to generalizing their results to non-isolated and/or non-homogeneous configurations. It was soon recognized that the shear field is closely related to the dynamical evolution of filamentary and wall-like structures and that the dynamical evolution of aspherical density perturbations is rather complicated and determined by the initial ellipticity, the external matter distribution and the background cosmology.

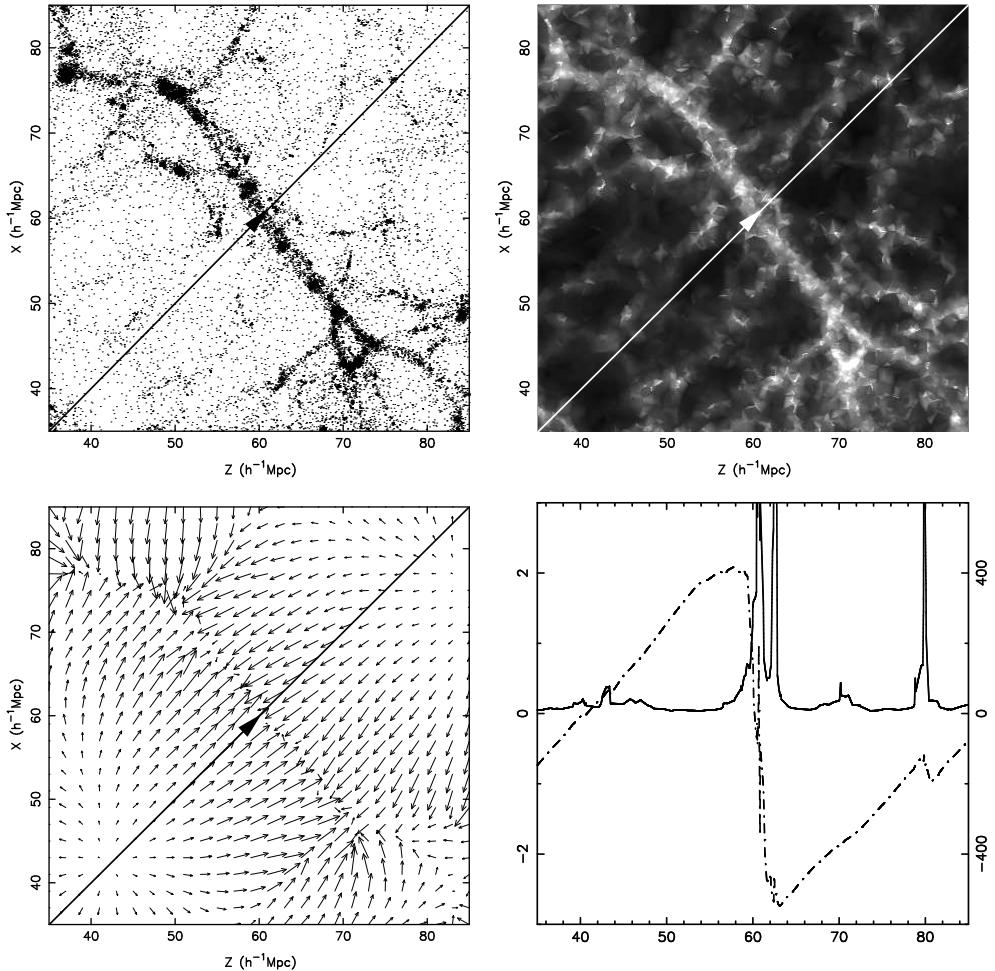


Figure 6.14 — A typical filamentary region in the simulation shown in Fig. 6.7. Top left-hand frame: particle distribution in a thin slice through the simulation box. Top right-hand frame: two-dimensional slice through the three-dimensional DTRE density field reconstruction. Bottom left-hand frame: two-dimensional slice through the three-dimensional DTRE velocity field reconstruction. Bottom right-hand frame: density and velocity reconstructions along the thick line shown in the other frames.

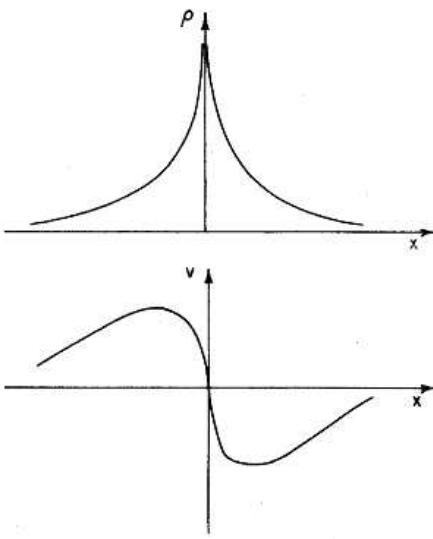


Figure 6.15 — The formation of a Zel'dovich pancake begins with the development of a singularity in the density distribution (top frame). At the same time the velocity field develops a vertical tangent at the position of the singularity (bottom frame). Source: Shandarin & Zel'dovich 1989.

6.6.2 Evolution

The potential of the DTFE is further illustrated by Fig. 6.16, in which the same filamentary region is shown as depicted in Fig. 6.14 at a number of cosmic times, ranging from $z = 5$ to $z = 0$. At each cosmological time the particle distribution in a thin slice is shown together with the DTFE density and velocity field reconstructions, as well as one-dimensional sections through these fields. The figure shows the evolution of the filamentary region from a relatively extended overdense region to a compact and collapsed filament. The velocity profile evolves from a rough sinusoid to a distinct Z when the filament has collapsed. Using high resolution DTFE reconstructions to study the evolution of the density and velocity field of a large sample of representative filaments and walls may significantly increase our understanding of these complex structures and the relation with their environment.

6.7 Shear and vorticity

So far we have only considered the velocity field itself and its divergence. In section 6.3 we have shown that the DTFE velocity field reconstruction technique also provides the shear and vorticity of the velocity field. Shear and vorticity patterns are expected to be prominent near high density regions, where distortions in the velocity flows are relatively strong. In general, vorticity measures the speed of rotation of a fluid element, while shear measures the anisotropy in its expansion rate.

Shear is a dominant factor in the shaping of large scale structure (e.g. Hoffman 1986, van de Weygaert & Babul 1994, Bond et al. 1996, Bond & Myers 1996a, b, c). The shear in the velocity field can be due to the intrinsic asphericity of evolving structures (due to anisotropic collapse) and/or tidal forces exerted on the local matter distribution by the surrounding large scale matter distribution. Shear is therefore expected to be present in linear, quasi-linear and non-linear regions as a result of gravitational interactions and the collapse of matter. In fact, the rate of growth of the density field gets amplified by the presence of shear, which increases

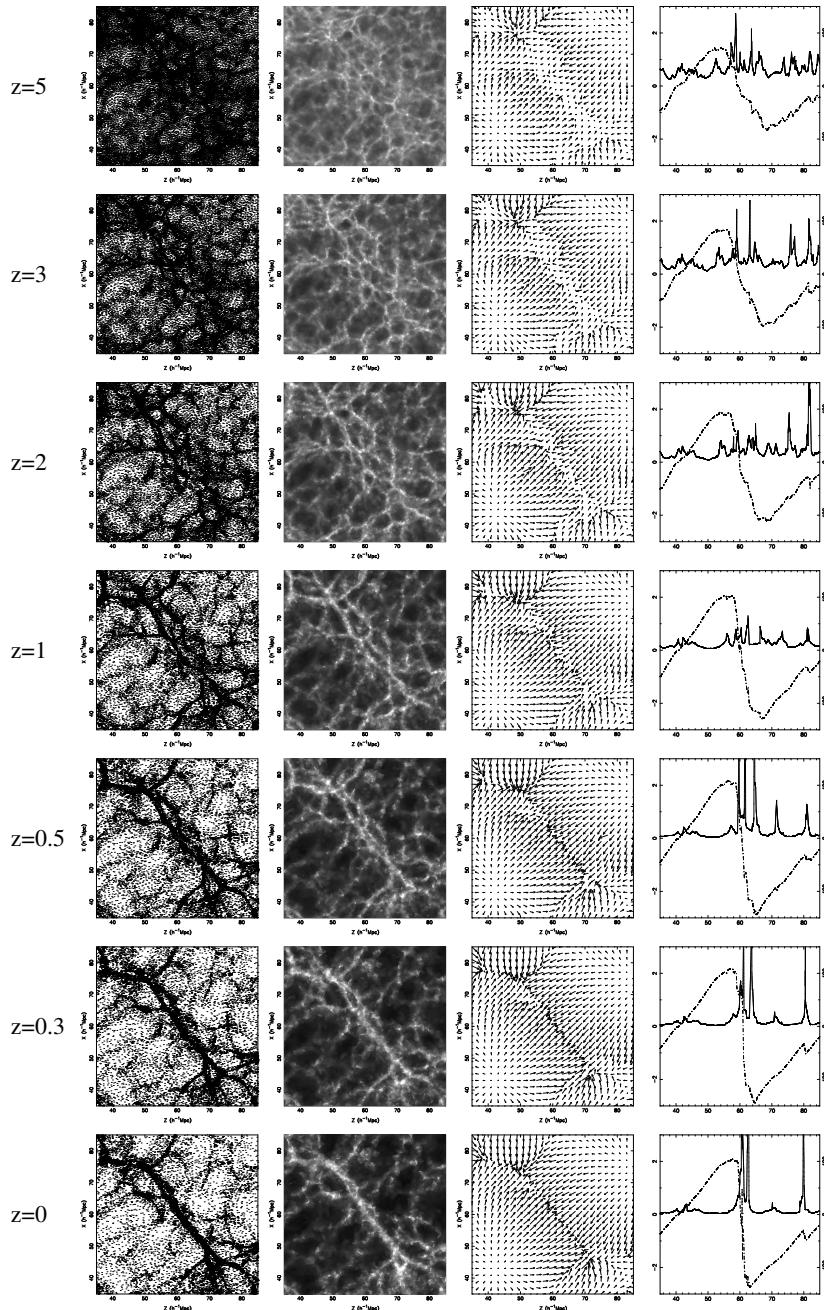


Figure 6.16 — Evolution of a filamentary region from $z = 5$ to $z = 0$. The filament shown is the same as depicted in Fig. 6.14. At each cosmic time the particle distribution is shown in a thin slice together with a two-dimensional slice through the full three-dimensional DTRE reconstructed density and velocity fields and one-dimensional sections through these fields. The sections are taken along the same directions as shown in Fig. 6.14.

the rate of growth of fluid convergence $-\nabla \cdot \mathbf{v}$ following a given fluid element (Hoffman 1986, Bertschinger & Jain 1994). This convergence takes place around collapsing structures, such as filaments and clusters. Bond et al. (1996) pointed out that the cosmic web is in fact a consequence of the distribution and spatial coherence of the shear field in the medium (van de Weygaert 2002). The density field and the velocity shear are therefore expected to be correlated as a function of time. There is however no simple relation between the density field and the velocity shear. In the linear regime the velocity shear is related to the density excess growth factor $D(t)$ via (van de Weygaert & Bertschinger 1996)

$$\sigma_{ij} \propto D(t)H(t)f(\Omega_m). \quad (6.21)$$

In the linear regime the velocity shear therefore does have a direct relation with the density field since $\delta(t) \propto D(t)$.

According to linear theory of gravitational instability (Peebles 1980) the large scale velocity field is irrotational, with $\nabla \times \mathbf{v} = 0$. Any vorticity mode would have decayed away during the linear growth of density fluctuations, while the only growing modes are curl-free. Based on Kelvin's circulation theorem, the flow remains vorticity free as long as it is laminar, i.e., until orbit crossing occurs (Bertschinger & Dekel 1989, Dekel et al. 1990). This condition is expected to hold even in the weakly non-linear regime (Bertschinger & Dekel 1989). It is therefore expected that vorticity is only present around non-linear features in the density field.

Figure 6.17 shows the shear (top left-hand frame) and vorticity (top right-hand frame) amplitudes, $\sigma = (\sum \sigma_{ij} \sigma_{ij})^{1/2}$, $\omega = (\sum \omega_{ij} \omega_{ij})^{1/2}$ in a $100 \times 100 h^{-1} \text{Mpc}$ slice at actual time through a cosmic N -body simulation. This simulation concerns a standard ΛCDM simulation and the slice depicted is the same as shown in Fig. 6.2. In the bottom frame the density contrast and velocity field in this slice are also shown. All fields have been convolved with a Gaussian kernel of $R_G = 1 h^{-1} \text{Mpc}$. The color scales of the shear and vorticity are inverted with respect to the density contrast in order to get a better visual impression of the shear and vorticity patterns. As can be noticed, both the shear and the vorticity trace the matter distribution rather well. As expected, the shear gives a better impression of the large scale matter distribution than the vorticity which is mainly visible around non-linear features in the density field.

6.7.1 Shear tensor and eigenvalues

Above we have considered the amplitude of the shear. The velocity shear is a symmetric and traceless tensor which contains much more information than just its amplitude. The shear tensor describes how the matter distribution is dynamically affected by its surroundings, i.e. if the matter distribution is being stretched or squeezed along a given direction. This information is contained within the eigenvalues and eigenvectors of the diagonalized matrix representation of the velocity shear tensor. The eigenvalues indicate the strength of the stretching or compression, while the direction of stretching or compression is given by the corresponding eigenvectors. Below we have ordered the eigenvalues λ_i (and corresponding eigenvectors) according to their amplitudes, with $\lambda_1 > \lambda_2 > \lambda_3$. Because the shear tensor is traceless, only two of its three eigenmodes will be linearly independent, i.e. $\sum \lambda_i = 0$. By construction the eigenvectors form a unitary and orthogonal system. The amplitude of this shear reference frame will be given by the eigenvalues, and since in general these have different values, they form

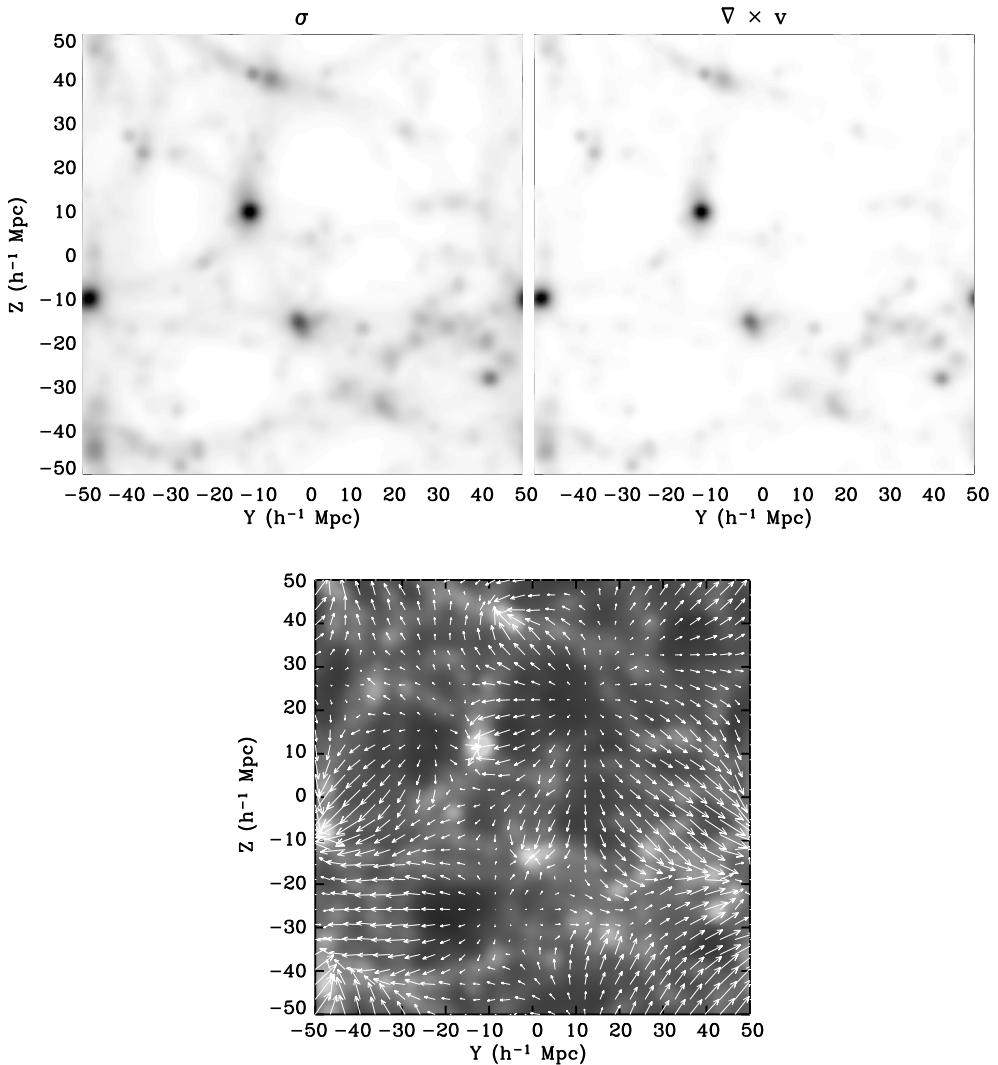


Figure 6.17 — DTFE recovered velocity shear and vorticity amplitudes (top frames) and density and velocity fields (bottom frame) for the same slice through the cosmological N -body simulation as shown in Fig. 6.2. The color scales of the shear and vorticity are inverted with respect to the density contrast in order to get a better visual impression of the shear and vorticity patterns.

an ellipsoid which is known as the *velocity ellipsoid*. The first eigenmode (λ_1) is positively defined and indicates the direction and intensity of the maximum stretching exerted over a given region. The third eigenvalue (λ_3) is always negative and corresponds to a compression along a direction perpendicular to the stretching one. The second eigenmode (λ_2) can be positive and negative. The configuration of the tidal field is determined by this eigenmode, with an extra stretching or compression perpendicular to the first eigenmode. If $\lambda_1 > \lambda_2 > 0$ the

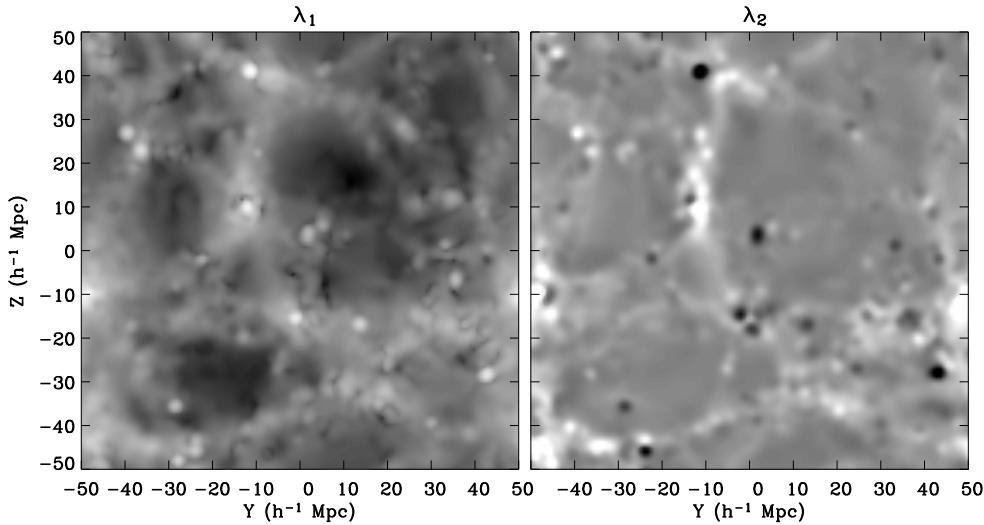


Figure 6.18 — Velocity shear eigenvalue maps for the same slice as shown in Fig. 6.17. The left-hand frame corresponds to the amplitude of the largest eigenvalue (stretching mode) along the slice. The right-hand frame shows the second eigenvalue which can be positive (stretching, light regions) and negative (compression, gray-dark regions).

velocity ellipsoid describes a planar structure, while if $\lambda_3 < \lambda_2 < 0$ it describes a filamentary configuration.

Using the DTFE procedure these eigenmodes can be computed for each Delaunay tetrahedron. However, for reasons of efficiency we did so from the grid-interpolated velocity gradient field. By diagonalizing the shear matrix at each pixel of the image, we find its eigenvectors and eigenvalues. Note that since the velocity-gradient matrix is not continuous, the same holds for the computed eigenvalues and eigenvectors.

Fig. 6.18 depicts the amplitude of the two linearly independent eigenvalues (λ_1 and λ_2) for the same slice as shown in Fig. 6.17. The stretching mode (λ_1 , left-hand frame) traces the filamentary structure present in the map (see bottom panel of Fig. 6.17). Stretching is stronger at the core of high density regions than along filaments. As expected, no stretching is found inside void regions. For the second eigenvalue (right-hand frame) the brightest regions correspond to a positive mode ($\lambda_2 > 0$) where stretching also occurs. Filamentary structures show up as light colored regions. Most regions in the map correspond to $\lambda_2 \approx 0$ (gray areas), predominantly voids where almost no shear is present.

The amplitudes of the eigenvectors are very illustrative in displaying the characteristics of the large scale matter distribution. They are however only a part of the total information contained within the shear tensor. In Fig. 6.19 we present a two-dimensional section through the velocity shear and its projected eigenvalues and eigenvectors. This is equivalent to the intersection of the velocity ellipsoid with the two-dimensional plane. This intersection is an ellipse whose major axis corresponds to the stretching mode, and whose minor axis corresponds to the compression along the bisecting plane. Both modes are presented in Fig. 6.19

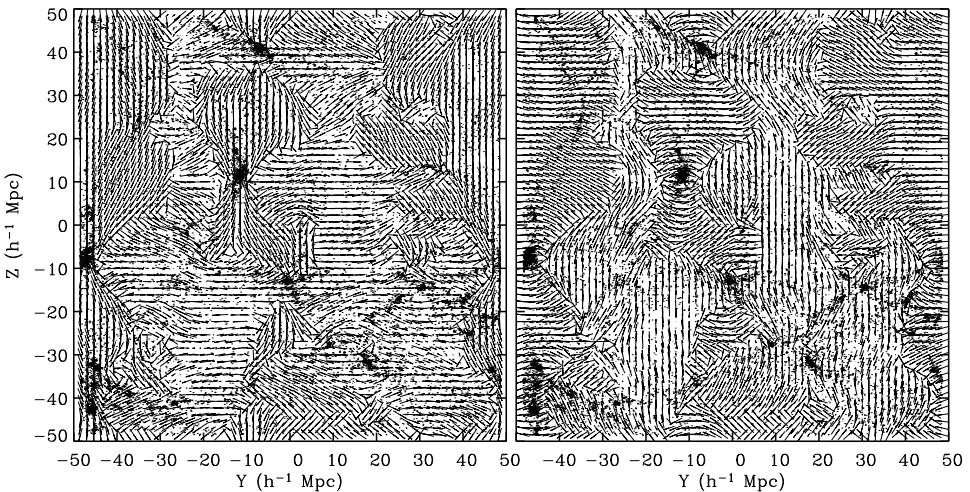


Figure 6.19 — Decomposition of the velocity shear components along the same slice as shown in Fig. 6.18. The dilational (stretching) component is shown in the left-hand frame, while the compressional component is shown in the right-hand frame. The bars are proportional to the intensity of the components and are aligned along the corresponding eigenvectors' direction. The particle distribution is also depicted.

as ‘shear bars’. The left-hand frame represents the stretching mode, the right-hand frame the compressional mode. The bars are proportional to the amplitude of each eigenvalue and oriented according to their corresponding eigenvectors. The point distribution corresponds to the matter distribution in a $5h^{-1}$ Mpc thick slice. One may observe that there is a correlation between the compressional bars and the filamentary structure of the point distribution. This correlation is stronger for the gravitational tidal field (see van de Weygaert 2002). The compressional bars are strong (large) and oriented almost perpendicular to the filamentary structures. The dilational bars tend to be more aligned with the filament. More massive structures like clusters are delineated by the compressional bars as a result of infalling motions around these clusters. In this representation clusters form the ‘nodes’ of the cosmic web. Underdense regions appear very quiet in both representations. They are clearly recognizable as those regions where bars are not perturbed, but coherently aligned instead.

With the shear-bar analysis we have shown the strong interplay between shear and the filamentary structure of the universe. This confirms the theoretical basis of the Zel'dovich formalism. By using a similar approximation Hoffman (1986) noticed that objects with large initial shear collapse sooner than one would predict. This result can also be derived from first principles from the Raychaudhuri equation which connects the velocity divergence, the shear and the vorticity with the density field. According to the Zel'dovich formalism (Zel'dovich 1970), objects collapse first along the eigenmode with the largest eigenvalue to form Zel'dovich pancakes, then along the next eigenmode to form filaments, and finally along the last eigenmode to collapse into clusters. If the primordial Gaussian velocity field would be shear-less and ir-

rotational, the velocity-gradient matrix would be isotropic, allowing for spherical collapse as predicted by the spherical model.

6.8 Summary and discussion

In this chapter we have described the formalism for simultaneously reconstructing the density and velocity fields corresponding to a discrete set of irregularly distributed sampling points. In principle the same technique can be applied to any dynamical field which has been sampled at the locations of these points. The formalism is an essential ingredient of the Delaunay Tessellation Field Estimator (DTFE), which itself is an extension of the work by Bernardeau & van de Weygaert (1996). The DTFE reconstruction procedure yields continuous and volume-covering fields. Its main advantage is that it is intrinsically self-adaptive to the density and geometry of the distribution of sampling points and does not make use of any pre-specified smoothing kernel.

We have shown by the specific example of the simultaneous reconstruction of a density and velocity field corresponding to an N -body simulation of cosmic structure formation that the characteristic elements of the large scale matter distribution are realistically rendered. The reconstructed density and velocity fields adhere closely to analytically predicted density-velocity divergence relation.

We have explicitly discussed the dynamical modeling of voids and filaments, which both are problematic structures for conventional reconstruction methods. Void-like regions contain very few galaxies or simulation particles, which poses specific problems for the analysis of their density and velocity fields in both observations and numerical simulations. Conventional density field reconstructions tend to be dominated by shot-noise effects. The same holds for velocity field reconstructions, which are even more problematic as they are severely under-sampled. The DTFE reconstruction of the velocity field in underdense regions is not beset by this problems. Instead, the density field is realistically rendered as a slowly varying low density valley. At the same time, the void's velocity profile resembles the super-Hubble linear expansion which is predicted by analytical models. We have described these models and explicitly formulated the equations governing the evolution of a void in a Λ CDM universe. A comparison of the model predictions with numerical simulations showed that they are in good agreement. We have also discussed the possibility of using these analytic void models to constrain cosmological parameters.

Next we have discussed the reconstruction of filamentary and wall-like structures. Because these are highly anisotropic in one or two of their spatial dimensions, conventional methods which usually make use of spherically symmetric smoothing kernels fail to give an accurate description of such objects (see Chapters 4 and 5). The DTFE reconstructions of a filamentary structure and its evolution shows that the method is capable of tracing with high resolution both the collapsing structure as well as the infalling motions towards and around these structures. The resulting density and velocity profiles are in good agreement with the Zel'dovich approximation.

Because of the linear interpolation inside each of the Delaunay tetrahedra it is rather straightforward to compute velocity-related quantities such as the velocity divergence, shear and vorticity. We have found that both the shear and vorticity trace the matter distribution very well, which is in accordance with theoretical predictions. The same holds for the eigenvalues

of the velocity shear, which describe whether the matter distribution is squeezed or stretched along a given direction. In particular, we have shown that the compressional mode is strongly related to the filamentary structure of the large scale matter distribution, again in accordance with theoretical predictions (see e.g. van de Weygaert 2002).

The main restriction of the DTDE is that it is a linear reconstruction procedure which assumes the reconstructed fields to vary linearly in between the sampling points. In high density regions shell crossing has occurred and the velocity field is not uniquely defined. In such regions the DTDE interpolation is not valid for the velocity field. However, the reconstructed field may be smoothed in order to obtain physically reliable field estimates.

In conclusion, we have shown that the application of the DTDE to the reconstruction of cosmic velocity fields has important advantages over traditional reconstruction procedures. We have also shown that the dynamics of the cosmic web can be used to constrain cosmological parameters. Applying the DTDE reconstruction procedure to the analysis of large simulations of cosmic structure formation will therefore certainly help us to gain understanding of the formation and evolution of the cosmic web and its characteristic constituent elements.

References

- Bernardeau F., 1992, ApJ, 390, L61
 Bernardeau F., 1994, ApJ, 433, 1
 Bernardeau F., Juszkiewicz R., Dekel A., Bouchet F.R., 1995, MNRAS, 274, 20
 Bernardeau F., van de Weygaert R., 1996, MNRAS, 279, 693
 Bernardeau F., van de Weygaert R., Hivon E., Bouchet F.R., 1997, MNRAS, 290, 566
 Bernardeau F. et al., 1999, MNRAS, 309, 543
 Bernardeau F., Colombi S., Gaztanaga E., Scoccimarro R., 2002, Physics Reports, 367, 1
 Bertschinger E., Dekel A., 1989, ApJ, 336, L5
 Bertschinger E., Jain B., 1994, ApJ, 431, 486
 Birkhoff G.D., 1923, *Relativity and Modern Physics*, Harvard University Press, Cambridge, USA
 Bond J.R., Kofman L., Pogosyan D., 1996, Nature, 380, 603
 Bond J.R., Myers S.T., 1996a, ApJS, 103, 1
 Bond J.R., Myers S.T., 1996b, ApJS, 103, 41
 Bond J.R., Myers S.T., 1996c, ApJS, 103, 63
 Carroll S.M., Press W.H., Turner E.L., ARAA, 1992, 30, 499
 Dekel A., Bertschinger E., Faber S.M., 1990, ApJ, 364, 349
 Dekel A., Rees M.J., 1994, ApJ, 422, L1
 Delone B.N., 1934, Bull. Acad. Sci. USSR: Classe Sci. Mat., 7, 793
 Dubinski J., Da Costa L.N., Goldwirth D.S., Lecar M., Piran T., 1993, ApJ, 410, 458
 Efstathiou G., Davis M., White S.D.M., Frenk C.S., 1985, ApJS, 57, 241
 Gingold R.A., Monaghan J.J., 1977, MNRAS, 181, 375
 Gnedin N.Y., 1995, ApJS, 97, 231
 Goldberg D.M., Vogeley M.S., 2004, ApJ, 605, 1
 Gunn J.E., Gott J.R., 1972, ApJ, 176, 1
 Hockney R., Eastwood J., 1988, *Computer Simulations Using Particles*, Taylor & Francis, Bristol, PA, USA
 Hoffmann Y., 1986, ApJ, 308, 493

- Icke V., 1972, PhD Thesis, University of Leiden, Leiden, The Netherlands
- Icke V., 1973, A&A, 27, 1
- Icke V., 1984, MNRAS, 206, 1
- Kauffmann G. et al., 1999, MNRAS, 303, 188
- Kudlicki A., Chodorowski M., Plewa T., Rozyczka M., 2000, MNRAS, 316, 464
- Lahav O., Rees M.J., Lilje P.B., Primack J.R., 1991, MNRAS, 251, 128
- Lawson C.L., 1977, in *Mathematical Software*, Vol. 3, Rice J. (ed.), Academic Press, New York, USA
- Lilje P.B., Lahav O., 1991, ApJ, 374, 29
- Lucy L.B., 1977, AJ, 82, 1013
- Noh W.F., 1964, Methods Comput. Phys., 3, 117
- Okabe A., Boots B., Sugihara K., Nok Chiu S., 2000, *Spatial Tessellations, Concepts and Applications of Voronoi Diagrams*, 2nd edition, Wiley, Chichester, UK
- Owen J.M., Villumsen J.V., Shapiro P.R., Martel H., 1998, ApJS, 116, 155
- Peacock J.A., 1999, *Cosmological physics*, Cambridge, Cambridge University Press, UK
- Peebles P.J.E., 1980, *The large scale structure of the universe*, Princeton University Press, Princeton, USA
- Pen U., 1998, ApJS, 115, 19
- Regős E., Geller M., 1989, AJ, 98, 755
- Regős E., Geller M., 1991, ApJ, 377, 14
- Romano-Díaz, PhD Thesis, University of Groningen, Groningen, The Netherlands
- Schechter P.L., 1980, AJ, 85, 801
- Shandarin S.F., Zel'dovich Y.B., 1989, Rev. of Mod. Phys., 61, 185
- Shapiro P.R., Martel H., Villumsen J.V., Owen J.M., 1996, ApJS 103, 269
- Sheth R.K., van de Weygaert R., 2004, MNRAS, 350, 517
- Silk J., 1974, ApJ, 193, 525
- van de Weygaert R., van Kampen E., 1993, MNRAS, 263, 481
- van de Weygaert R., Babul A., 1994, ApJ, 425, L59
- van de Weygaert R., Bertschinger E., 1996, MNRAS, 281, 84
- van de Weygaert R., 2002, *Froth across the Universe*, in '2nd Hellenic Cosmology Meeting', Plionis M., Cotsakis S. (eds.), Kluwer Academic Publishers, Dordrecht, NL
- White S.D.M., Silk J., 1979, ApJ, 231, 1
- Whitehurst R., 1995, MNRAS, 277, 655
- Xu G., 1997, MNRAS, 288, 903
- Zel'dovich Y.B., 1970, A&A, 5, 84

7

The Foamy Morphology of the 2dF Galaxy Distribution

W.E. Schaap & R. van de Weygaert

ABSTRACT — We present a reconstruction of the foamy morphology of the galaxy distribution in the 2dF Galaxy Redshift Survey (2dFGRS). The Delaunay Tessellation Field Estimator (DTFE) was used to reconstruct the projected galaxy surface density field as well as the full three-dimensional galaxy density field. The DTFE is a self-adaptive method for a volume-covering reconstruction of continuous density fields which are sampled by a discrete set of points. It is capable of resolving highly complex point distributions such as the large scale galaxy distribution, which is characterized by anisotropic filamentary and planar features, a hierarchy of spatial scales and densities and a cellular geometry with extended and poorly sampled low density voids. Here we present maps of the projected galaxy surface density for different regions in the 2dFGRS as well as three-dimensional renderings of the complete 2dFGRS region. We discuss our results and some of the various possible applications.

7.1 Introduction

Over the last decades ever deeper redshift surveys have provided us with an increasingly refined view of our cosmic neighborhood (examples include the CfA Redshift Survey (Huchra et al. 1990), the Las Campanas Redshift Survey (Shectman et al. 1996), the 2dF Galaxy Redshift Survey (hereafter 2dFGRS, Colless et al. 2001, 2003) and the Sloan Digital Sky Survey (Stoughton et al. 2002). These surveys have shown that galaxies are not randomly distributed in the universe, but reside in an intriguing network, the cosmic web (Bond et al. 1996, van de Weygaert 1991, 2002). At the nodes of this network reside dense and compact galaxy clusters which form deep potential wells trapping hot X-ray emitting gas and containing up to thousands of galaxies. The Virgo and Coma clusters form the most nearby examples. Clusters are connected by filamentary and wall-like structures of intermediate densities ($\delta\rho/\rho \sim 10$), usually consisting of a large number of galaxy groups, each containing tens of galaxies. The largest of these structures are called superclusters, an example of which is the Local Supercluster, which has the Local Group and thus the Milky Way amongst its members. These extended and elongated structures enclose large and empty regions, voids. The result is a foamy network which is pervading the whole observable universe. The highly complex cellular geometry of the cosmic web is one of the most dominant characteristics of the large scale matter distribution in our universe (see for an extensive review van de Weygaert 2002).

The final data release of the 2dFGRS comprises a set of 221 414 galaxies of which the redshift has been determined with a reasonable confidence. The even larger Sloan Digital Sky Survey (Stoughton et al. 2002), in which in total around a million redshifts are being measured, has been released up to the fourth data release (Adelman-McCarthy et al. 2006). It is the mere size of these new samples which is making it finally possible to perform a detailed analysis of the cosmic web and its elements beyond the second-order power spectrum or correlation function analysis on which conventional studies traditionally have focused (see e.g. Park et al. 1994, Lin et al. 1996, Percival et al. 2001, Dodelson et al. 2002, Tegmark et al. 2004, Cole et al. 2005). This task, however, represents a major challenge. An adequate description of the large scale galaxy distribution is not only complicated by its complex cellular geometry, but also by the very different physical and morphological properties of its constituent elements. For example, densities range from values of some tenths in units of the average density in void-like regions (e.g. Hoyle & Vogeley 2002, 2004, Sheth & van de Weygaert 2004, Colberg et al. 2005, Patiri et al. 2006) up to thousands of times the average density at the cores of rich clusters (e.g. Tyson et al. 1998, Rosati et al. 2002 and references therein). In addition, the hierarchical build-up of structure is responsible for another complication in the analysis of large scale structure. Individual objects often contain a lot of substructure and may themselves be embedded within a hierarchy of larger associations. An example is the Local Group, which consists of two smaller groups centered around the Milky Way and Andromeda. At the same time it is a member of the much larger Local Supercluster. Even voids are known to display substructure, with small voids being embedded within more extended voids (Regős & Geller 1991, Dubinski et al. 1993, van de Weygaert & van Kampen 1993, Szomoru et al. 1996, Sheth & van de Weygaert 2004). To complicate things even more, the structural elements of the large scale galaxy distribution have very different shapes. Whereas both voids and clusters are fairly roundish, walls and filaments appear highly anisotropic, having contracted in one (walls) or two (filaments) spatial directions.

Several attempts have been made to quantify and describe the properties of the cosmic web, amongst which are analyses based on minimal spanning tree techniques and percolation studies, for example by employing Minkowski functionals (e.g. Sahni et al. 1997, Sheth et al. 2003, Shandarin et al. 2004). A major downside of most of these methods is that in one or the other way they often make use of filtering aspects, usually a Gaussian smoothing on a typical scale of $5h^{-1}$ or $10h^{-1}$ Mpc. This is also the case for studies which seek to reconstruct the large scale galaxy density field, of which here we mention those based on Wiener filtering techniques (see e.g. Rybicki & Press 1992, Zaroubi et al. 1995, Erdođu et al. 2004). These also pre-impose certain constraints on the underlying density field, e.g. a specific power spectrum.

In this paper we apply the Delaunay Tessellation Field Estimator (hereafter DTFE) to the 2dFGRS public data release. The DTFE is a method to reconstruct continuous density fields which are sampled by a discrete set of points. It is based on the Delaunay tessellation (De-lone 1934) of the given discrete point set. This spatial and volume-covering division of space into mutually disjunct triangular (in two dimensions) or tetrahedral (in three dimensions) cells adapts to the local density and geometry of the point distribution. The DTFE exploits these virtues and adapts fully automatically to changes in the density and geometry of the distribution of sampling points. The DTFE involves an extension of the interpolation procedure described by Bernardeau & van de Weygaert (1996), who used Delaunay tessellations for the specific purpose of estimating the cosmic velocity divergence field. They showed the optimal performance of Delaunay tessellations compared to conventional interpolation methods and proved that the obtained field estimates are volume-weighted in stead of mass-weighted. The DTFE was developed for the purpose of reconstructing fully volume-covering and volume-weighted fields from a discrete set of field values, including the density field as defined by the point sample itself. Instead of involving user-defined filters which are based on artificial smoothing kernels, the main virtue of the DTFE is that it is intrinsically self-adaptive, because it involves filtering kernels which are defined by the local density and geometry of the distribution of sampling points.

The main goal of this chapter is to reconstruct the large scale galaxy density field in the 2dFGRS region. We present both projected two-dimensional surface density maps as well as three-dimensional density maps. Apart from the DTFE surface density field reconstructions at full resolution we also present the same maps after filtering with a Gaussian kernel at a resolution of $5h^{-1}$ Mpc for the purpose of serving as a comparison with conventional density field reconstructions. We also discuss several applications. Note that in this chapter we describe the reconstruction of density maps in redshift space. The transformation between redshift and real space involves a dynamical modelling which falls beyond the scope of this chapter.

This chapter is organized as follows. In section 7.2 we describe the 2dFGRS data set. In section 7.3 we explain the fundamentals of the DTFE, while in section 7.4 its main properties are discussed. In section 7.5 we present a surface map of the DTFE 2dFGRS galaxy density field without making any corrections for observational selection effects. The latter are described in section 7.6. In sections 7.7 and 7.8 we present DTFE reconstructed two-dimensional galaxy surface density and three-dimensional galaxy density maps, which have been corrected for observational selection effects. Finally, in section 7.9 we summarize and discuss our results and explore several of the possible applications of the DTFE reconstruction procedure in the context of studies of the large scale matter distribution.

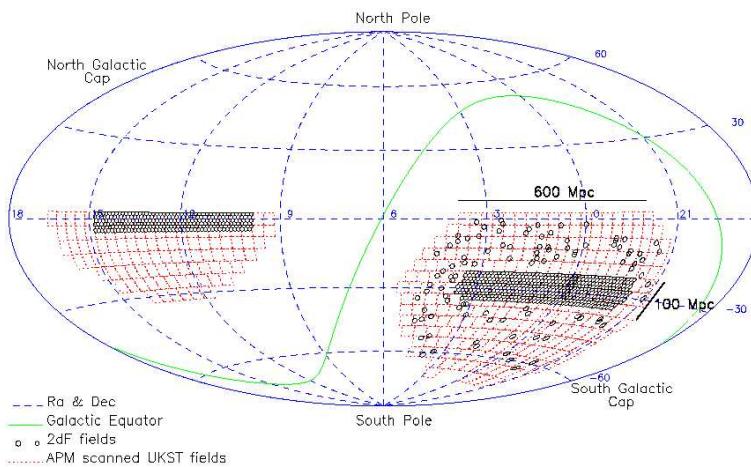


Figure 7.1 — The 2dFGRS regions in Aitoff projection of RA and Dec with individual fields marked as small circles. Also shown are the lines of Galactic latitude $|b| = 0^\circ, 30^\circ$ and 45° . Source: Colless et al. 2001, 2003.

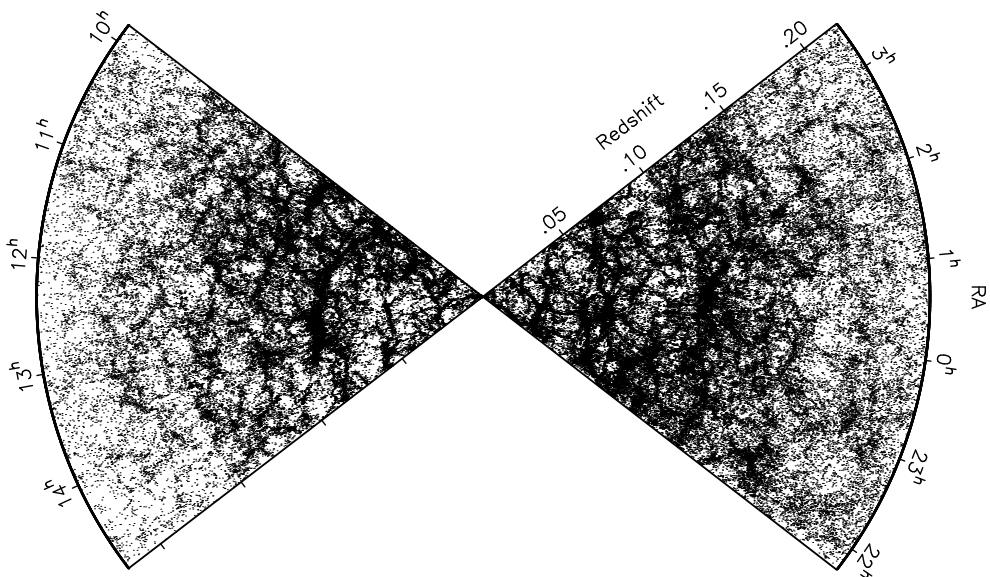


Figure 7.2 — The 2dF galaxy redshift survey. The foamy geometry of the cosmic web is strikingly displayed. Courtesy: the 2dF Galaxy Redshift Survey team.

7.2 The data

The 2dFGRS is one of the major spectroscopic surveys in which the spectra of 245 591 objects have been obtained, with the scope of obtaining a representative picture of the large scale distribution of galaxies in the nearby universe (Colless et al. 2001, 2003). It is a magnitude-limited survey, with galaxies selected down to a limiting magnitude of $b_J \sim 19.45$ from the extended APM Galaxy Survey (Maddox 1990a, b, c). The galaxies were selected in three regions, which together cover an area of approximately 1500 squared degrees. These regions include two declination strips, each consisting of overlapping 2° fields, as well as a number of ‘randomly’ distributed 2° control fields. In Fig. 7.1 a map of the survey fields on the sky is shown. One strip (the SGP strip) is located in the southern Galactic hemisphere and covers about $80^\circ \times 15^\circ$ close to the South Galactic Pole ($21^h40^m < \alpha < 03^h40^m$, $-37.5^\circ < \delta < -22.5^\circ$). The other strip (the NGP strip) is located in the northern Galactic hemisphere and covers about $75^\circ \times 10^\circ$ ($09^h50^m < \alpha < 14^h50^m$, $-7.5^\circ < \delta < +2.5^\circ$). The redshifts of these galaxies have been measured using the 2dF multifibre spectrograph on the Anglo-Australian Telescope, which is capable of simultaneously observing 400 objects in a 2° diameter field. Reliable redshifts were obtained for 221 414 galaxies. These data have been made publically available in the form of the 2dFGRS final data release (available at <http://msowww.anu.edu.au/2dFGRS/>) and are plotted in Fig. 7.2.

7.3 DTFE: method

The DTFE is a procedure to reconstruct a continuous and volume-covering density field from a set of irregularly distributed points sampling this field. It has been extensively described in Chapter 3. Here we shortly describe the main workings and the most important properties of the DTFE. It consists of three main steps. These are illustrated in Fig. 7.3 for a set of points sampling a Gaussian peak. These three steps consist of:

Step 1: Construction of the Delaunay tessellation

The core of the DTFE is the Delaunay tessellation corresponding to the discrete point set. The first step is to construct this Delaunay tessellation for the point sample at hand (Fig. 7.3, top right-hand frame).

Step 2: Estimation of the density field at the locations of the sampling points

In the next step the density is estimated at the locations of the sampling points by making use of the observation that the area of the union of all Delaunay triangles surrounding a point, the *contiguous Voronoi cell* of that point, is inversely proportional to the local density of sampling points. This is illustrated in the bottom right-hand frame of Fig. 7.3. The contiguous Voronoi cells are shaded: one at the lower right-hand corner, in a low density region, with a relatively large area, and one near the center, in a high density region, with a relatively small area. The DTFE defines the density at the location of a sampling point i as the inverse of the volume V of its contiguous Voronoi cell \mathcal{W}_i times a normalization constant:

$$\hat{\rho}(\mathbf{x}_i) = \frac{(D+1)m_i}{V(\mathcal{W}_i)}. \quad (7.1)$$

Here m_i is the mass of sampling point i and D the dimension of space.

DTFE reconstruction procedure:

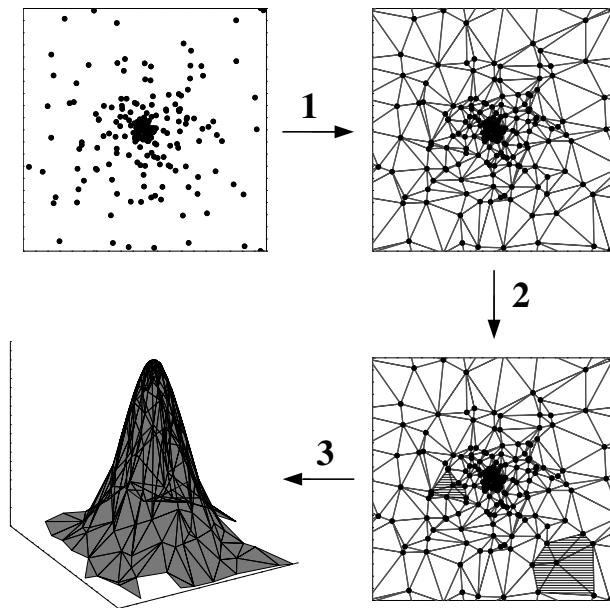


Figure 7.3 — Overview of the DTFE reconstruction procedure. Given a point distribution (top left), one constructs the corresponding Delaunay tessellation (top right), estimates the density at the position of the sampling points by taking the inverse of the area of their corresponding contiguous Voronoi cells (bottom right) and assumes that the density varies linearly within each Delaunay triangle, resulting in a volume-covering continuous density field (bottom left).

Step 3: Interpolation of the density field to all other points in space

In the final step the density field is defined at all other points in space by making use of the fact that the Delaunay triangles form the multi-dimensional equivalent of one-dimensional linear interpolation intervals. Given the $(D + 1)$ density estimates at the vertices of each Delaunay triangle, the value $\widehat{\rho}(\mathbf{x})$ at location \mathbf{x} inside triangle m is defined through a straightforward linear interpolation:

$$\widehat{\rho}(\mathbf{x}) = \widehat{\rho}(\mathbf{x}_i) + \left. \widehat{\nabla \rho} \right|_m \cdot (\mathbf{x} - \mathbf{x}_i), \quad (7.2)$$

where \mathbf{x}_i is the location of one of the Delaunay vertices of the triangle. It involves an evaluation of the (linear) density gradient $\widehat{\nabla \rho}|_m$ inside triangle m , which follows rather straightforwardly from the set of density estimates $\rho(\mathbf{x}_i)$ at the locations of the vertices of Delaunay triangle m . This step is illustrated in the bottom left-hand frame of Fig. 7.3.

Once a density field has been reconstructed it may subsequently be processed. Processing may involve various operations. These vary from relatively straightforward applications such as image reconstruction and smoothing to more complex and sophisticated applications such as feature detection and various statistical analyses.

7.4 DTFE: properties

The DTFE is a first-order reconstruction procedure, resulting in a linearly varying and continuous density field. The derivative of the reconstructed field is constant inside each Delaunay triangle, but discontinuous at their edges (Fig. 7.3, bottom left-hand frame).

7.4.1 Advantages

The DTFE has a number of properties which makes it ideally suited for the analysis of patterns in the large scale galaxy distribution. Firstly, it automatically adapts to the local density of the galaxy distribution under consideration. This finds its basis in the Delaunay tessellation of the corresponding galaxy distribution, which automatically adapts to the local density. This can be appreciated from Fig. 7.4. In the top frame the 2dFGRS galaxies are shown, while the subsequent frames zoom in on the corresponding Delaunay tessellation. The figure illustrates the adaptive nature of the Delaunay tessellation, which makes the DTFE better suited than conventional reconstruction procedures for resolving point distributions which are characterized by objects of widely varying densities (see Chapter 4).

Fig. 7.4 shows that DTFE does not only adapt to the density, but also to the geometry of the local galaxy distribution. Unlike most conventional reconstruction procedures the DTFE does not involve an intrinsically spherical smoothing kernel. This results in a much more accurate description of the real intrinsic anisotropies of reconstructed objects (see Chapter 4). A telling example of this property is present in the bottom right-hand frame of Fig. 7.4, where just to the right of the center a thin, anisotropic filamentary structure can be recognized, a finger of God.

A third virtue of the DTFE is its capability of an accurate description of low density regions without introducing the shot-noise effects by which conventional methods are beset. In Fig. 7.5 the DTFE reconstruction of a void-like region taken from the 2dFGRS survey is compared to several Triangular-Shaped Clouds (TSC) reconstructions at different resolutions. The TSC procedure is a conventional fixed grid-based reconstruction procedure in which the resolution has to be pre-specified by the user (e.g. Hockney & Eastwood 1981). In the top row of the figure the galaxy distribution and the DTFE reconstruction are shown, while in the bottom row three different TSC reconstructions at increasing resolution are shown. It is clear that at none of the resolutions all characteristics of the void are accurately described by TSC. At the lowest resolution the overall shape of the void is recovered, but the sharp edges visible in the galaxy distribution are smeared out into rather featureless blobs. To a lesser extent the same is true for the reconstruction at average resolution. Here, however, one starts to recognize the inner structure of the void breaking up into distinct density blobs, indicating that the grid resolution in the interior of the void is too high. This is in particular true for the high resolution reconstruction, in which the edges of the void seem to be described much better, but in which the interior of the void is dominated by shot-noise effects. The DTFE recovers the sharp edges of the void as such, while at the same time the interior of the void is reconstructed as a gently varying low density region. It is important to note that this void is just one of the many objects present in the large scale galaxy distribution. Any grid- or filter-based reconstruction method will be inappropriate for describing all but some of the relevant features.

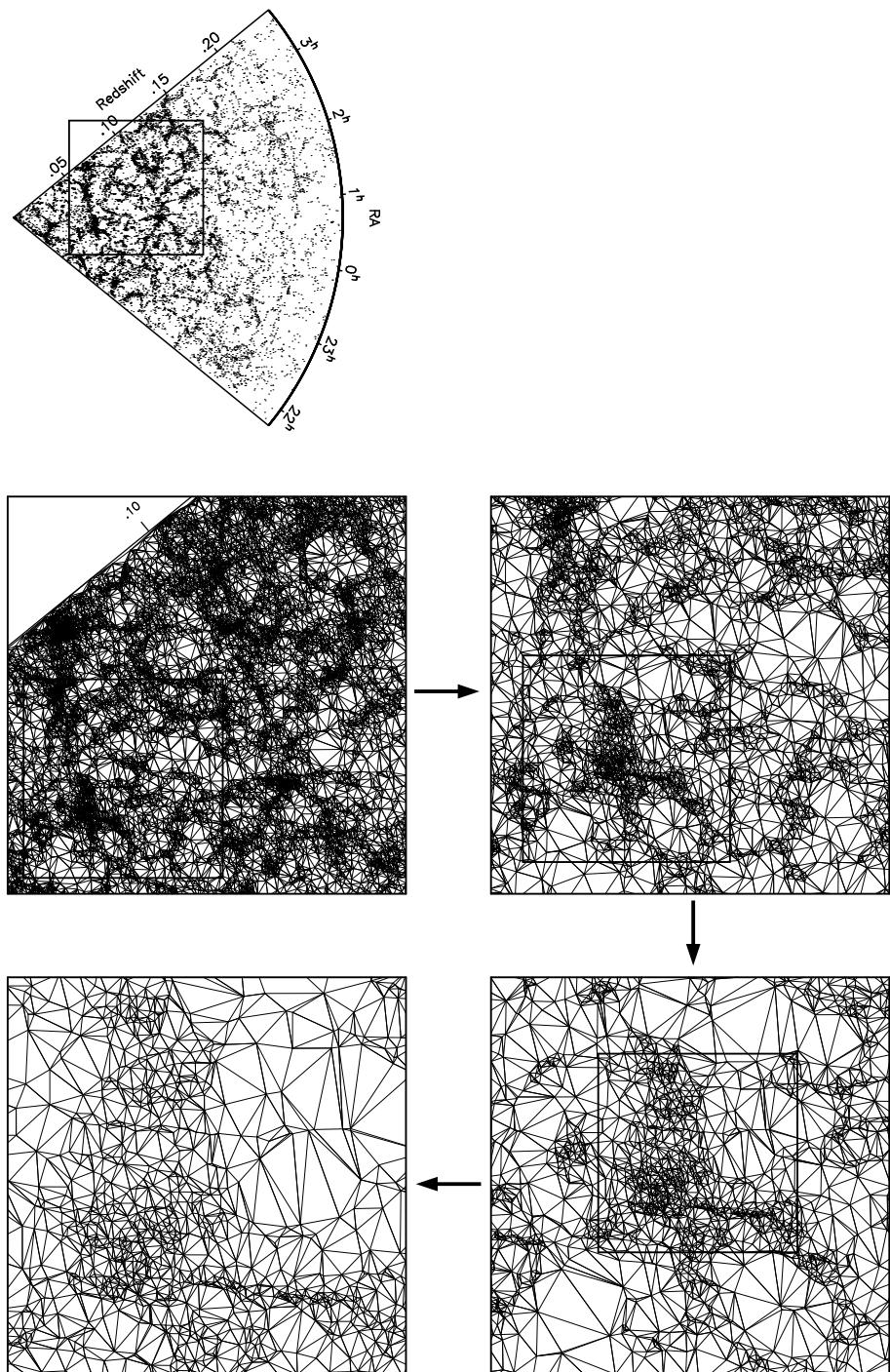


Figure 7.4 — Illustration of the adaptive properties of the DTFE. Top frame: the galaxy distribution of the 2dFGRS SGP sample. The subsequent frames zoom in on the Delaunay triangulation of the region indicated by the solid squares.

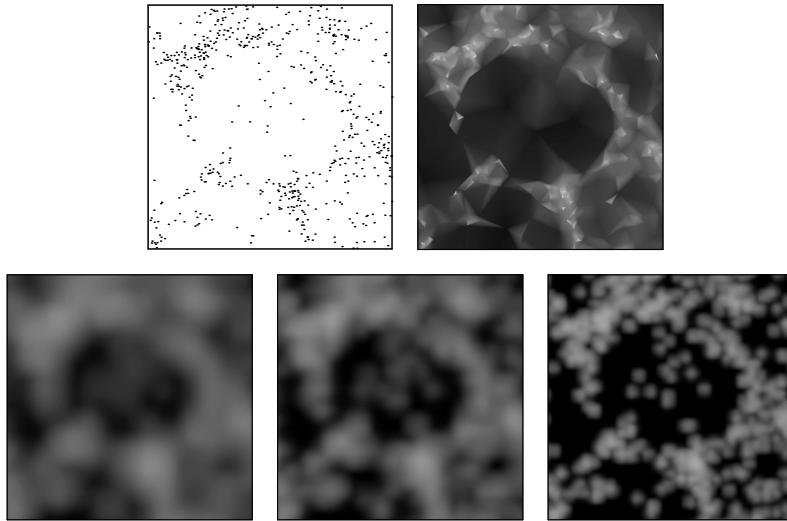


Figure 7.5 — Density field reconstructions of a void region. Top left: galaxy distribution. Top right: DTFE reconstruction. Bottom row: TSC reconstructions at different resolution. Going from left to right the resolution of the TSC grid increases from the standard choice of on average one galaxy per grid-cell for the full galaxy sample, via on average one particle per grid-cell in the depicted region, to on average one particle per four grid-cells in this region.

7.4.2 Limitations

Apart from these advantages the DTFE also involves a number of limitations and disadvantages. The first is that due to the very high resolution of its effective smoothing kernel, the DTFE is very sensitive to sampling noise in the point distribution. Any variations in the local density due to sampling noise will have repercussions on the reconstructed density field. In Chapter 8 the effects of sampling noise and the significance of reconstructed features are discussed in detail. Note that an advantage of the DTFE is that the effects of sampling noise are local. A second disadvantage of the DTFE is that at the scale of the local smoothing kernel, which is as large as the local contiguous Voronoi cell, the reconstructed density field contains triangular artefacts. These artefacts are the imprints of the linear interpolation procedure within each Delaunay triangle. Obviously, this issue is not specific to the DTFE as any reconstruction scheme contains the imprint of its effective smoothing kernel at small scales. Another limitation of the DTFE is that it is not capable of reconstructing regions of density zero. Any region with a finite size corresponds to Delaunay triangles with a finite size and therefore a non-zero density. The final limitation of the DTFE we mention involves the boundary conditions. In the case of non-periodic boundary conditions the outermost triangles extend outwards to infinity, inducing a density of zero inside their interior. The density reconstruction is therefore not accurate for the outermost triangles.

7.5 Galaxy surface density field reconstructions

We have reconstructed the galaxy surface density field in redshift space corresponding to the 2dFGRS galaxy distribution. The projected galaxy distribution and the resulting DTFE re-

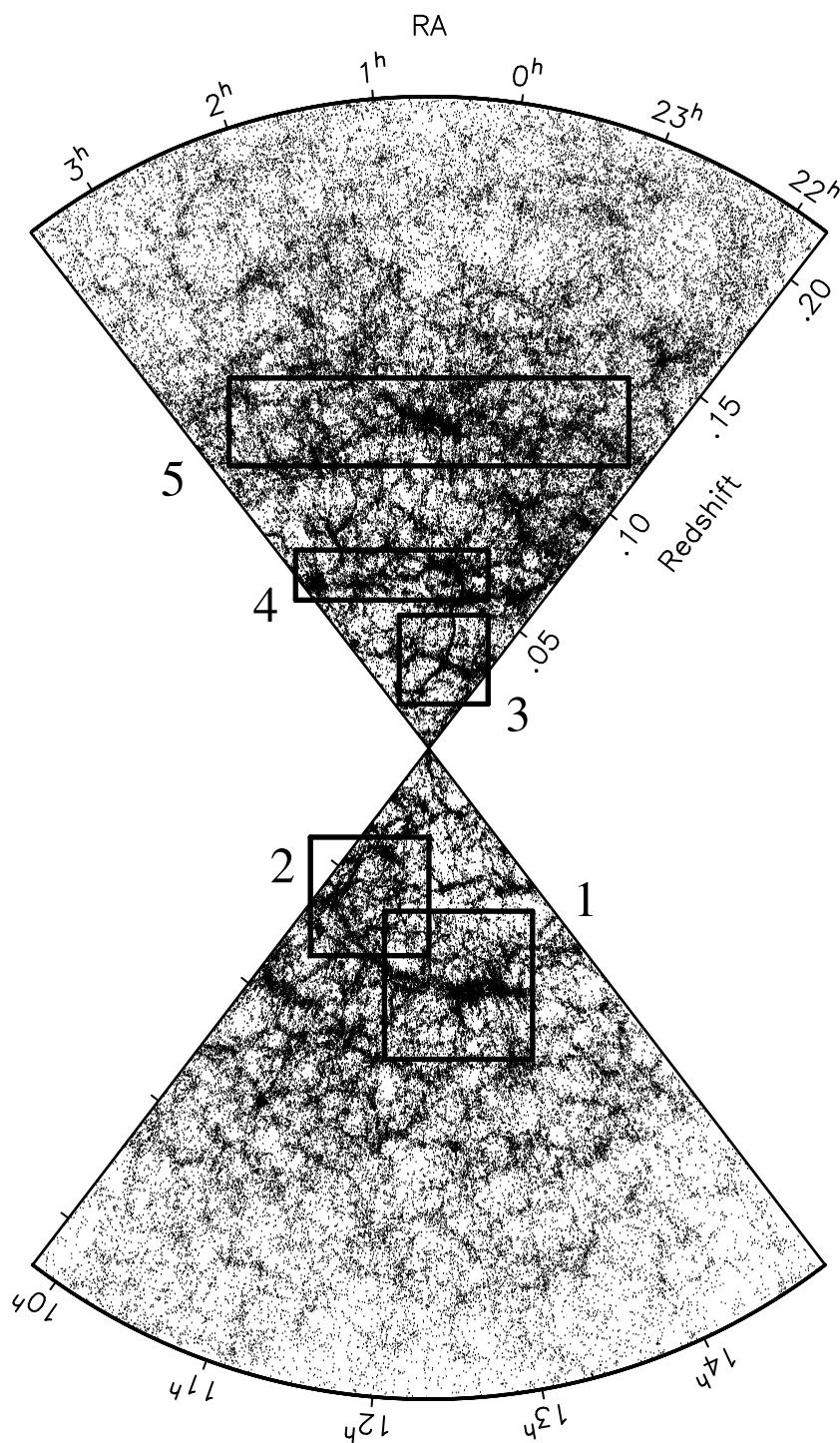


Figure 7.6 — The 2dFGRS galaxy distribution.

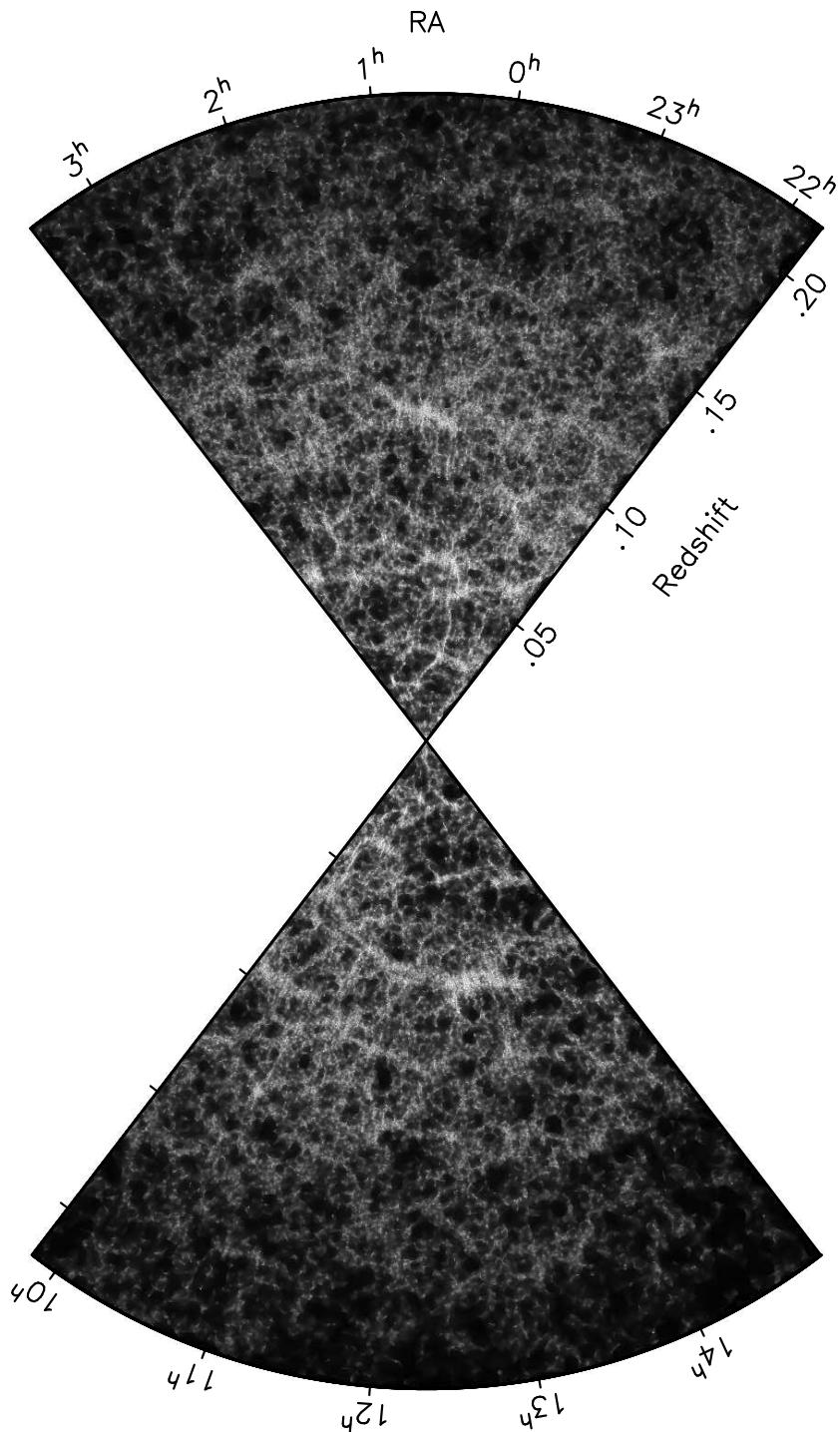


Figure 7.7 — The DTRE reconstructed galaxy surface density field.

constructed density field are shown in Figs. 7.6 and 7.7. In the galaxy distribution in Fig. 7.6 five regions are indicated, which are shown in more detail in Fig. 7.8. All density field reconstructions are DTFE reconstructions on the basis of the measured galaxy positions without correcting for any observational selection effects.

Fig. 7.7 shows the ability of the DTFE to reveal the strong contrasts in densities present in the large scale galaxy distribution at an automatically adapted resolution. The resolution of the DTFE is optimal in the sense that the smallest interpolation units it employs are also the smallest units defined by the data. At the same time the DTFE manages to bring out the fine structural detail of the intricate and often tenuous filamentary structures. Notice the sharp rendering of thin edges surrounding void-like regions.

To underline the capacity of the DTFE to dissect the internal structure of the various structural components of the cosmic web, in Fig. 7.8 we zoom in on several interesting regions. Region 1 focuses on the major mass concentration in the NGP region, the Sloan Great Wall (Gott et al. 2005). Various filamentary regions emanate from the high density core. In region 2 a void-like region is depicted. The DTFE renders the void as a low density area surrounded by various filamentary and wall-like features. Two fingers of God are visible in the upper right-hand corner of region 2, which show up as very sharply defined linear features. Many such features can be recognized in high density environments. It is interesting that the void is not reconstructed as a totally empty or featureless region. Substructures are present inside the void, which appears to contain a number of smaller ‘sub’-voids. This in accordance with current theories of the formation of voids (Dubinski et al. 1993, Sheth & van de Weygaert 2004). Region 3 is one of the most conspicuous structures in the 2dF field. This ‘cross’ consists of four tenuous filamentary structures emanating from a high density core located at the center of the region. Region 4 zooms in on some of the larger structures in the SGP region. In the bottom of this region part of the Pisces-Cetus supercluster is visible, while the concentration at the top of this region is the upper part of the Horologium-Reticulum supercluster. Finally, region 5 zooms in on the largest structure in the SGP region, the Sculptor supercluster (SCL9 in Einasto et al. 1997).

Even though the DTFE clearly offers a sharp image of the cosmic web, on the smallest scales (Fig. 7.8) the triangular imprint of the DTFE kernel is clearly visible. It is important to realize that any reconstruction procedure produces such artefacts whose shapes and scale are determined by the smoothing kernel. However, for most methods a smooth and spherically symmetric kernel is used, which tends to stand out less conspicuously.

A considerable amount of noise is visible in the reconstructions. This is a direct consequence of the high resolution of the DTFE reconstruction. Since no smoothing is applied, any noise present in the data will have a clear imprint on the reconstruction. Part of this noise is due to the statistical nature of the galaxy formation process. An extra source of noise is due to the fact that the galaxy positions have been projected on a two-dimensional plane before the density field reconstruction. This leads to additional Poisson noise as the DTFE algorithm connects galaxies which lie closely together in the projection, but may in reality be quite distant from each other. One way to circumvent this problem would be to reconstruct the three-dimensional galaxy density field and then do a two-dimensional projection, or by smoothing the field with a filter with a size equal to the local inter-galaxy separation.

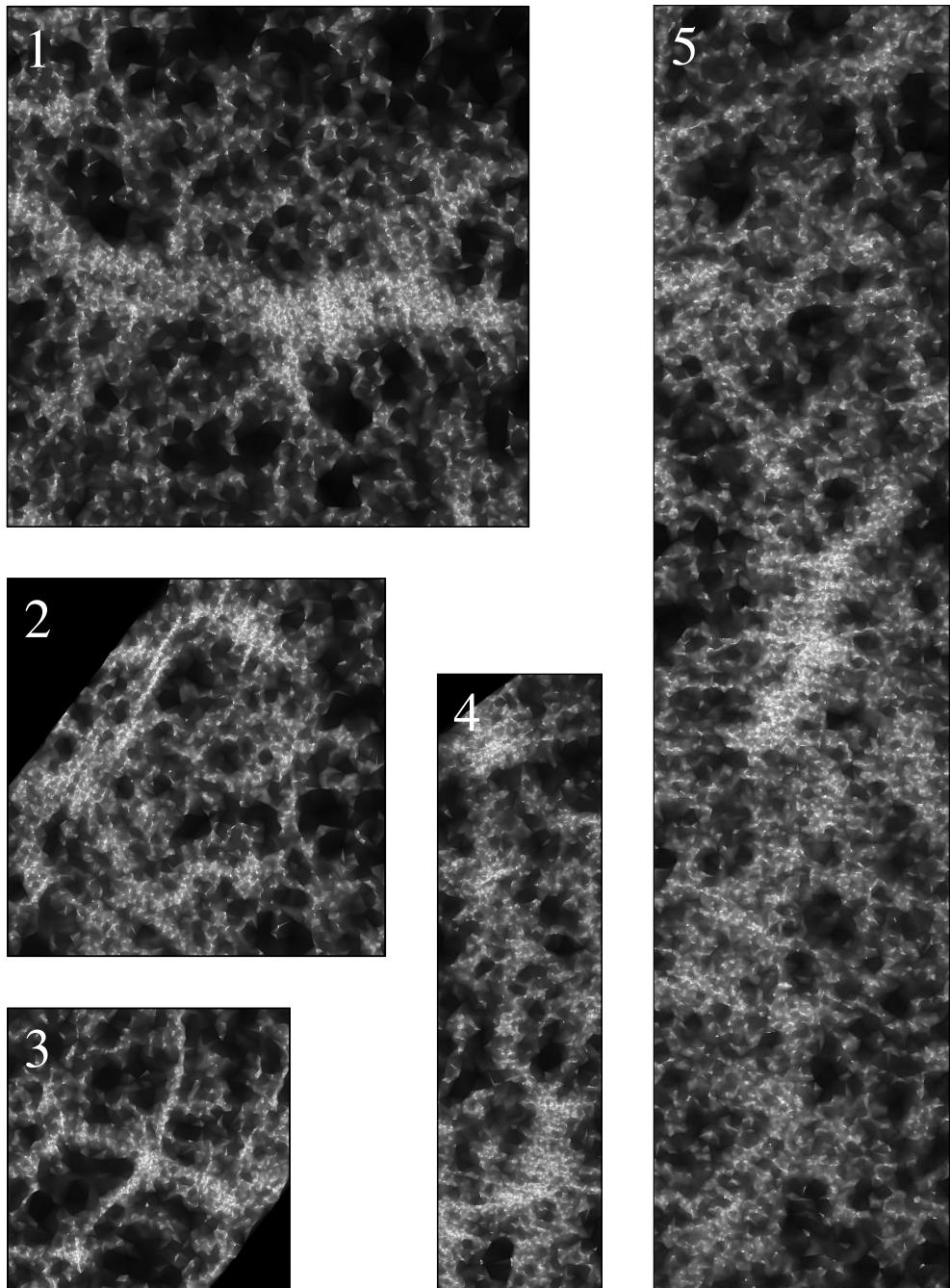


Figure 7.8 — Selected regions in the 2dFGRS galaxy surface density field.

7.6 Selection effects

The 2dFGRS is subject to a number of selection effects. In order to obtain a uniformly defined galaxy density field, one needs to take into account a number of observational selection effects which are present in the 2dFGRS sample:

- Non-uniform sampling;
- Varying redshift completeness;
- Magnitude limit variations;
- Varying radial selection function.

The scope of this subsection is to discuss these issues and the applied correction procedures.

7.6.1 Non-uniform sampling

The first effect one needs to take into account that in the 2dFGRS the observed 2° fields are not uniformly distributed. In Fig. 7.9 these fields are plotted as a function of position on the sky for the SGP and NGP strips (see also Colless et al. 2001, 2003). At certain values of the right ascension more 2° fields at different declinations have been observed than at other locations and the width of the observed area on the sky is larger. This is for example the case in the NGP strip for $9^h50^m < \alpha < 11^h45^m$ and in the SGP strip for $21^h50^m < \alpha < 23^h25^m$. At other places fields are missing, for example around $(\alpha, \delta) = (24^h0^m, -28^\circ)$ in the SGP strip. If not taken into account this will have a direct repercussion on the projected galaxy surface density field. Although in principle it is possible to correct for the extra galaxies by dividing by the amount of observed fields at fixed right ascension, this would not yield a fair reconstruction as different parts of the reconstructed surface density map would correspond to larger regions. Therefore we have selected several slices of the same thickness (2°) for the projected galaxy surface density field reconstructions.

Unfortunately it is not straightforward to correct for missing fields. These regions form a problem for the DTFE reconstruction procedure and should be excluded from the analysis. The reason for this is that in the DTFE procedure no a priori assumptions are made about the density field and an in principle arbitrary amount of galaxies could lie within a missing field. However, by imposing certain properties on the galaxy distribution, for example a particular power spectrum, one may fill in these blank regions. This can be done by using for example constrained field reconstruction techniques (Hoffman & Ribak 1991, Zaroubi et al. 1995, van de Weygaert & Bertschinger 1996). Such an analysis falls outside the scope of this chapter.

7.6.2 Varying redshift completeness

Another selection effect in the 2dFGRS concerns the fact that the fraction of redshifts which have been determined varies as a function of location on the sky. This redshift completeness, the local fraction of galaxies for which the redshift has been determined, is indicated in Fig. 7.9 by the color scaling (see Colless et al. 2001, 2003). The variations in redshift completeness occur for a number of reasons. Firstly, the redshift completeness is in general different for each observed 2° field, because of different observing conditions. Secondly, at many locations different fields overlap and contribute to the completeness. Finally, at some

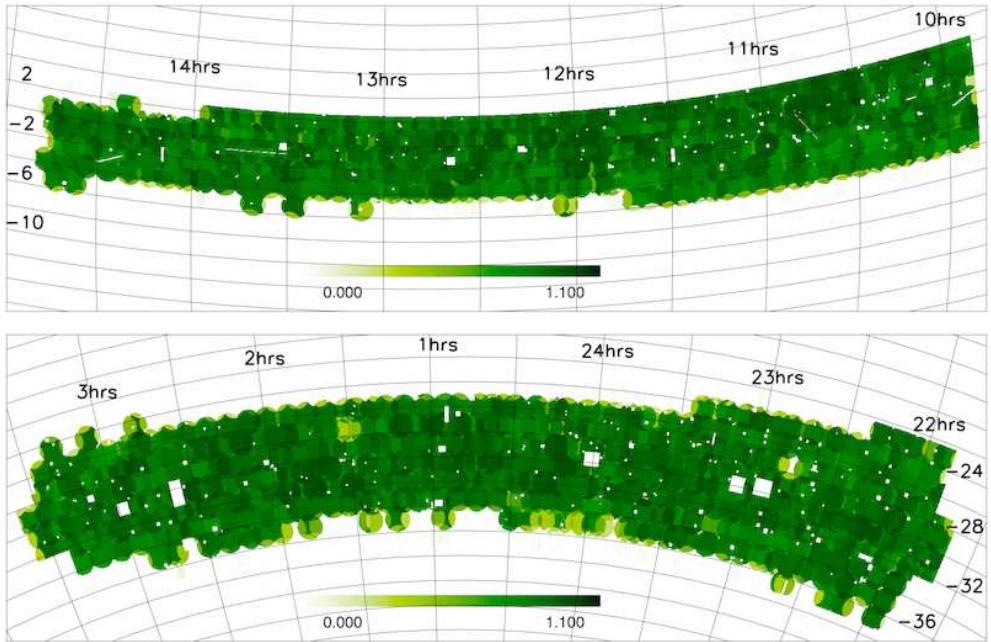


Figure 7.9 — The observed 2° fields in the 2dFGRS NGP (top) and SGP (bottom) strips. The grey scales denote the redshift completeness as a function of position on the sky, ranging from white (no redshifts have been determined) to black (all redshifts have been determined). Source: Colless et al. 2003.

locations the completeness equals zero. In such regions no galaxies have been observed, e.g. because of the presence of bright stars or due to satellite trails.

In order to obtain a uniformly defined galaxy surface density field it is important to correct for the variation of the redshift completeness and for the presence of holes where no galaxies have been observed. In the DTRE procedure the galaxy distribution is assumed to be a homogeneous Poisson point process of the underlying galaxy density field. If the point process is inhomogeneous Eqn. 7.1 is not applicable. Here it would underestimate the density by the inverse of the redshift completeness and we may therefore correct for the variations in redshift completeness by giving each galaxy a weight which is equal to the inverse of the redshift completeness. The resulting expression for the estimate of the density at the locations of the galaxies is then given by

$$\widehat{\rho}(\mathbf{x}_i) = \frac{(D+1)m_i}{\psi(\alpha, \delta) V(W_i)}. \quad (7.3)$$

In this expression $\psi(\alpha, \delta)$ denotes the redshift completeness at location (α, δ) and we have assumed that the sampling is uniform along redshift (see section 7.6.4). In the two-dimensional case of reconstructing the galaxy surface density field the redshift completeness $\psi(\alpha, \delta)$ should be replaced by the average redshift completeness $\psi(\alpha)$ at right ascension α . The latter is de-

fined as

$$\psi(\alpha) \equiv \frac{\int d\delta \psi(\alpha, \delta)}{\int d\delta}. \quad (7.4)$$

A complication occurs for locations at which the redshift completeness is very low or equal to zero. For these locations the above correction procedure may not be applied. Regions with redshift completeness zero will in fact be automatically excluded from the correction procedure as no galaxies have been observed at these locations. Since no data has been observed these locations should be treated in the same way as missing fields (see section 7.6.1). Because of the large uncertainty of the correction in regions with a very low redshift completeness, we have ignored regions with a redshift completeness smaller than 0.2 and set the density inside such regions equal to zero.

The correction for the varying redshift completeness comes at a price. Reconstructed structures inside regions with a lower redshift completeness have been sampled by a smaller amount of galaxies and therefore have a lower spatial resolution and a lower significance.

7.6.3 Magnitude limit variations

Another complication is that although the 2dFGRS was originally selected to have a uniform magnitude limit of $b_J = 19.45$, this limit in fact varies slightly with position on the sky (Colless et al. 2001). The reason for this is that the photometric calibrations on which the parent 2dFGRS catalogue has been based and the extinction corrections towards each galaxy have been improved since the beginning of the survey. In Colless et al. (2003) the magnitude limit is plotted as a function on the sky. In the NGP strip the magnitude limit is $b_J = 18.88$ at lowest, while in the SGP strip the magnitude limit is $b_J = 19.17$ at lowest. In order to get a proper magnitude-limited sample over these two strips, we selected only those galaxies with magnitudes smaller or equal to $b_J = 18.88$.

7.6.4 Varying radial selection function

The 2dFGRS is a magnitude-limited survey, which means that with increasing distance less and less galaxies are observed. This effect is clearly visible in both the galaxy distribution and in the DTFE galaxy surface density field reconstruction in Fig. 7.7. Here we describe how one may obtain a uniformly defined galaxy surface density field from a magnitude-limited sample.

One well-known way is to construct a volume-limited sample from the magnitude-limited 2dFGRS sample. Norberg et al. (2001, 2002) describe this procedure for the purpose of studying the dependence of galaxy clustering on luminosity and spectral type. However, as they point out, this procedure has the disadvantage that a large number of galaxies in the flux-limited sample does not satisfy the selection criteria for being included in the volume-limited sample. The reason for this is that the 2dFGRS sample has both a bright and a faint apparent magnitude limit. This means that galaxies are only included in the sample if their redshifts fall within a certain inclusion redshift range, which is dependent on their absolute magnitude. The latter means that a proper volume-limited sample can only be constructed in the redshift range where these inclusion ranges overlap. The major disadvantage of this approach is thus that one loses information and therefore structural resolution. Another disadvantage of this

approach is that the results are dependent on the assumed cosmological model, because the relation between apparent and absolute magnitude depends on cosmology.

Another approach is to model the radial selection function and to subsequently correct for it. Tegmark, Hamilton & Xu (2002) describe this procedure for determining the power spectrum of galaxies in the 2dFGRS 100k sample, as do Erdo\u0111du et al. (2004) in a Wiener filter analysis of the full 2dFGRS. The advantage of this approach is that all observed galaxies can be used for the analysis over the whole observed redshift range. Also, the results are not dependent on the adopted cosmology. However, the disadvantage of this approach is that the results critically depend on the accuracy by which the radial selection function has been modeled. Also, one implicitly assumes that galaxies of different luminosities have the same spatial distribution, which in reality is not the case.

Because one is able to use all the observed data over the whole observed redshift range, the approach of our choice is to model the radial selection function and to subsequently correct the reconstructed field for it.

7.6.4.1 Modeling of the radial selection function

The radial selection function denotes the fraction of galaxies which is observable as a function of redshift per volume unit. It is determined by the galaxy luminosity function and the magnitude limit of the survey. The varying radial selection function implies that the observed galaxy sample is not a homogeneous but an inhomogeneous Poisson process of the underlying galaxy density field. Eqn. 7.1 is therefore not applicable. Using this equation would underestimate the density by a factor which is equal to the inverse of the radial selection function and we may therefore correct for it by giving each galaxy a weight which is equal to the inverse of the radial selection function. The resulting expression for the estimate of the density at the locations of the galaxies is therefore given by

$$\widehat{\rho}(\mathbf{x}_i) = \frac{(D+1)m_i}{f(z)V(W_i)} \quad (3D). \quad (7.5)$$

Here $f(z)$ denotes the radial selection function at redshift z . When reconstructing the two-dimensional galaxy surface density field the radial selection function $f(z)$ should be replaced by $z f(z)$, where the extra geometrical factor z accounts for the linear increase of the thickness of the slice with redshift:

$$\widehat{\rho}(\mathbf{x}_i) = \frac{(D+1)m_i}{z f(z)V(W_i)} \quad (2D). \quad (7.6)$$

The selection function should be normalized such that it gives the fraction of observed galaxies at a particular redshift. Here we have normalized the selection function such that it specifies the observed amount of galaxies with a luminosity larger than L_0 , which corresponds to the lowest magnitude limit b_J of the 2dFGRS (see section 7.6.3). It is given by

$$f(z) = \frac{\int_{L(z)}^{\infty} \Phi(L)dL}{\int_{L_0}^{\infty} \Phi(L)dL}, \quad (7.7)$$

where $\Phi(L)$ is the galaxy luminosity function. We have modelled Φ by a Schechter galaxy luminosity function of the form

$$\Phi(L)dL = \Phi^* \left(\frac{L}{L^*} \right)^{\alpha} \exp \left(-\frac{L}{L^*} \right) d \left(\frac{L}{L^*} \right), \quad (7.8)$$

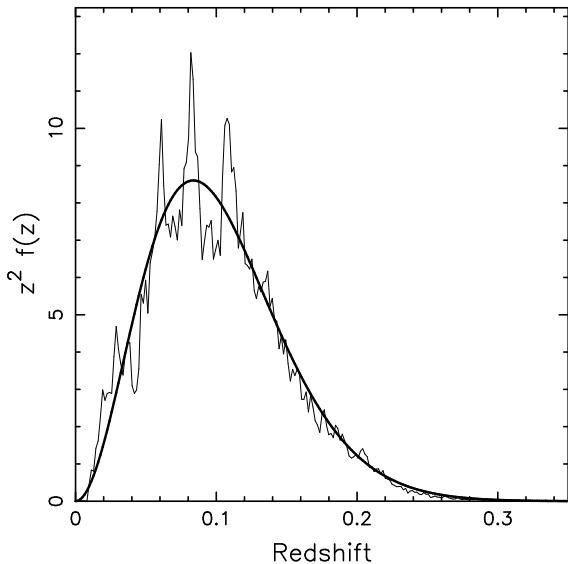


Figure 7.10 — The redshift distribution function of the galaxies in the NGP and SGP strips (thin solid line). An analytic prediction is also shown (thick solid line).

in which the parameters Φ^* , L^* and α have been measured by Norberg et al. (2002), who found $\alpha = -1.21$, $\Phi^* = 1.61 \cdot 10^{-2} h^3 \text{ Mpc}^{-3}$ and L^* following from $M^* - 5 \log_{10} h = -19.67$. These numbers yield a background density of galaxies of $n_b(L > L_0) = 1.39 \cdot 10^{-2} h^3 \text{ Mpc}^{-3}$.

In Fig. 7.10 we have plotted the observed redshift distribution of the galaxies in the NGP and SGP strips (the thin solid line). In a redshift slice the amount dN of galaxies in redshift bin $[z, z + dz]$ is related to the radial selection function by

$$dN \propto z^2 f(z) dz. \quad (7.9)$$

The factor z^2 is a geometric factor, accounting for the linear increase of both the width and the thickness of the slice with redshift. This analytic approach with the selection function $f(z)$ given by Eqn. 7.7 is also plotted in Fig 7.10 and is indicated by the smooth solid curve. This plot shows that the observed and predicted selection functions are in good agreement.

7.6.5 Itinerary

Below we have summarized the reconstruction procedure including the corrections for the different selection effects:

1. Select the galaxies down to a uniform magnitude limit;
2. Construct the corresponding Delaunay tessellation;
3. Estimate the densities at the galaxy locations according to the DTFE procedure;
4. Correct for the redshift completeness;
5. Correct for the radial selection function;
6. Interpolate the density field according to the DTFE procedure.

7.7 2-D galaxy surface density field reconstructions

We have reconstructed the galaxy surface density field in the 2dFGRS NGP and SGP regions. The results are shown in Figs. 7.11 and 7.13. Both regions have been partitioned into several slices with a width in declination of 2° . In this way one may obtain an impression of the three-dimensional structure of the 2dFGRS regions. In Figs. 7.12 and 7.14 the same slices are shown after smoothing with a Gaussian filter with an FWHM of $5h^{-1}$ Mpc, which may serve as a comparison with conventional density field reconstructions. The magnitude of the galaxy surface density is indicated in the color bar, in which the corresponding galaxy surface density is denoted in units of the background galaxy surface density Σ_b .

A visual inspection of the slices shows that the slices appear quite different. This illustrates the dangers involved in analyzing two-dimensional projections such as shown in Fig. 7.7. To a lesser extent the same holds for the slices shown in this section. Nevertheless we can make several observations. Structures at high redshifts appear more extended and blurry than structures at low redshifts. This is due to three effects. The first is that the effective resolution is varying across the map. The DTFE employs an effective smoothing kernel whose resolution is set by the local density (and geometry) of the galaxy distribution. This ensures that the resolution is higher in high density regions in which most information is present. Note that because the sampling is not uniform this also implies that similar structures at different locations are not resolved at the same resolution. Secondly, at higher redshifts the slice is thicker, effectively involving a smoothing over a larger declination range. Thirdly, the dilution of the sampling at higher redshifts means that structures are inherently less resolved. The latter two effects are not unique to the DTFE as may be appreciated in the Gaussian smoothed images in Figs. 7.11 and 7.13.

It is interesting to see how coherent features are over the different slices. The first thing which may be noted is that structures appear most coherent at low redshifts ($z < 0.05$), which is where the slices are relatively thin. At high redshift where the slices are thicker the structures in each slice appear uncorrelated. Most voids do seem to be present in more than one or in all slices. This is in particularly true for the largest voids.

In the NGP strip the most prominent structure is the large concentration at a redshift of about 0.08 and RA ranging from 12^h30^m to 13^h30^m . This structure corresponds to the Sloan Great Wall (Gott et al. 2005) and is most prominent in the slice with declination ranging from $\delta = -2^\circ$ to $\delta = 0^\circ$. In the slice with δ ranging from 0° to 2° it can also be clearly recognized. In this slice it visible that the Sloan Great Wall is indeed a huge structure, extending out to an RA of 10^h30^m . In the SGP strip the most prominent structure is located at a redshift of around 0.12 and an RA ranging from about 0^h to 1^h30^m , with the precise extent depending on the slice. This is the Sculptor Supercluster (SCL9 in Einasto et al. 1997).

7.8 3-D galaxy density field reconstructions

We have also reconstructed the full three-dimensional galaxy density field in the NGP and SGP regions of the 2dFGRS. The results are shown Figs. 7.15 and 7.16. These figures show a three-dimensional rendering of the NGP and SGP slices out to a redshift $z = 0.1$. To obtain a better impression of the three-dimensional structure inside these slices they are depicted at different viewing angles in Fig. 7.17. The iso-density contour in these figures corresponds to twice the average background galaxy density n_b .

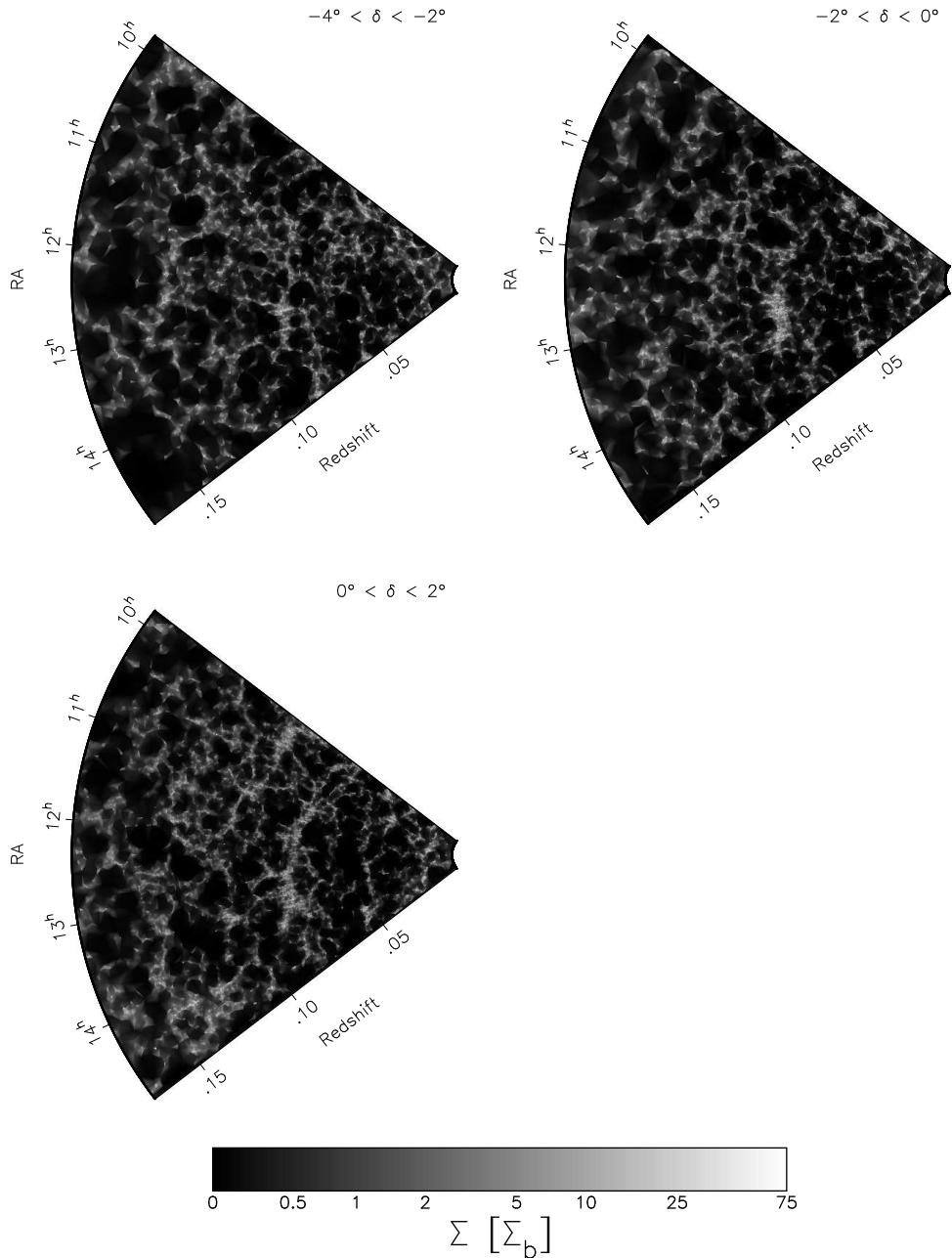


Figure 7.11 — The NGP 2dFGRS galaxy surface density field in several selected slices with a declination width of 2° . The color bar indicates the galaxy number density in units of the average background galaxy surface density (see text for a description).

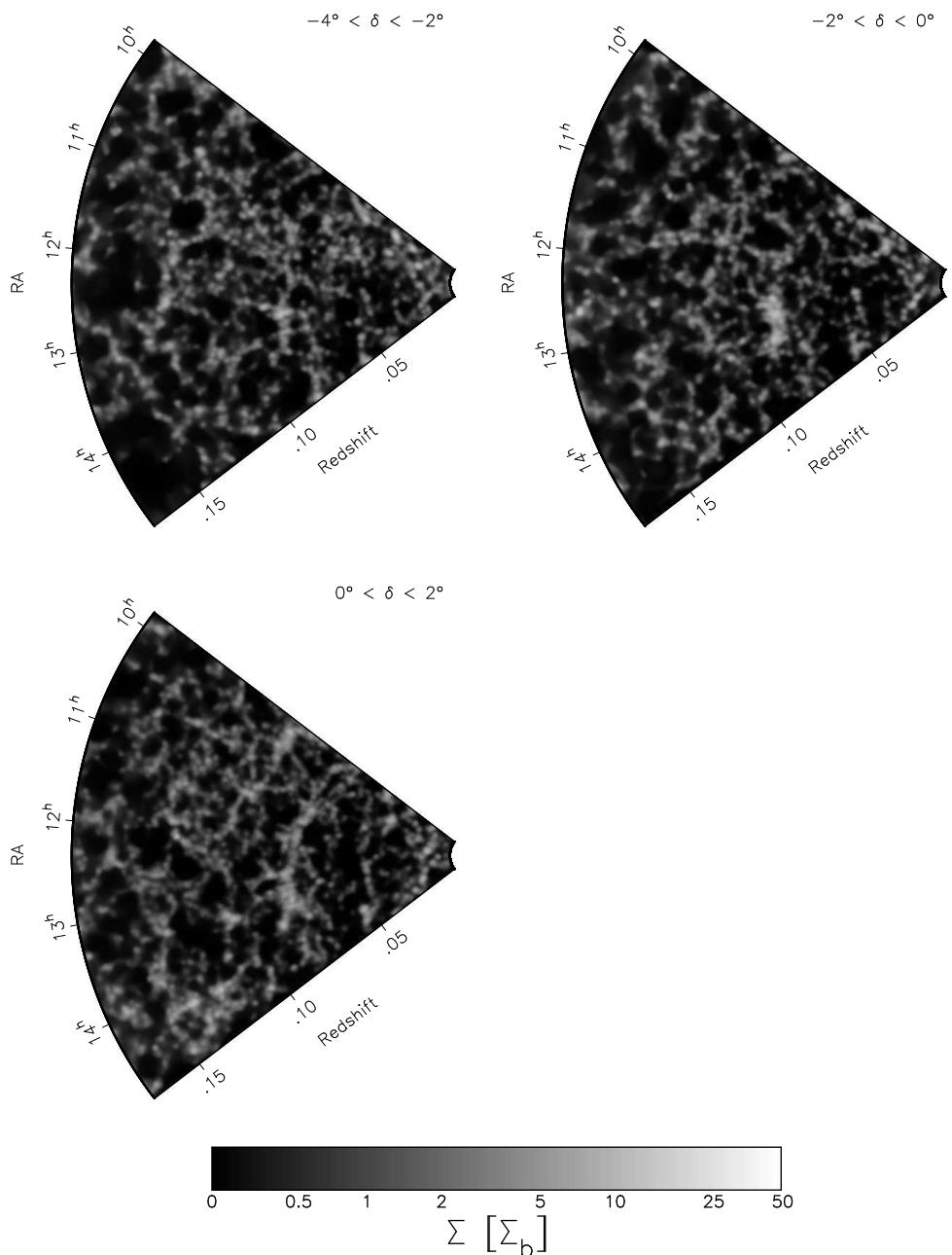


Figure 7.12 — Same as Fig. 7.11, smoothed with a Gaussian filter with an FWHM of $5h^{-1}$ Mpc.

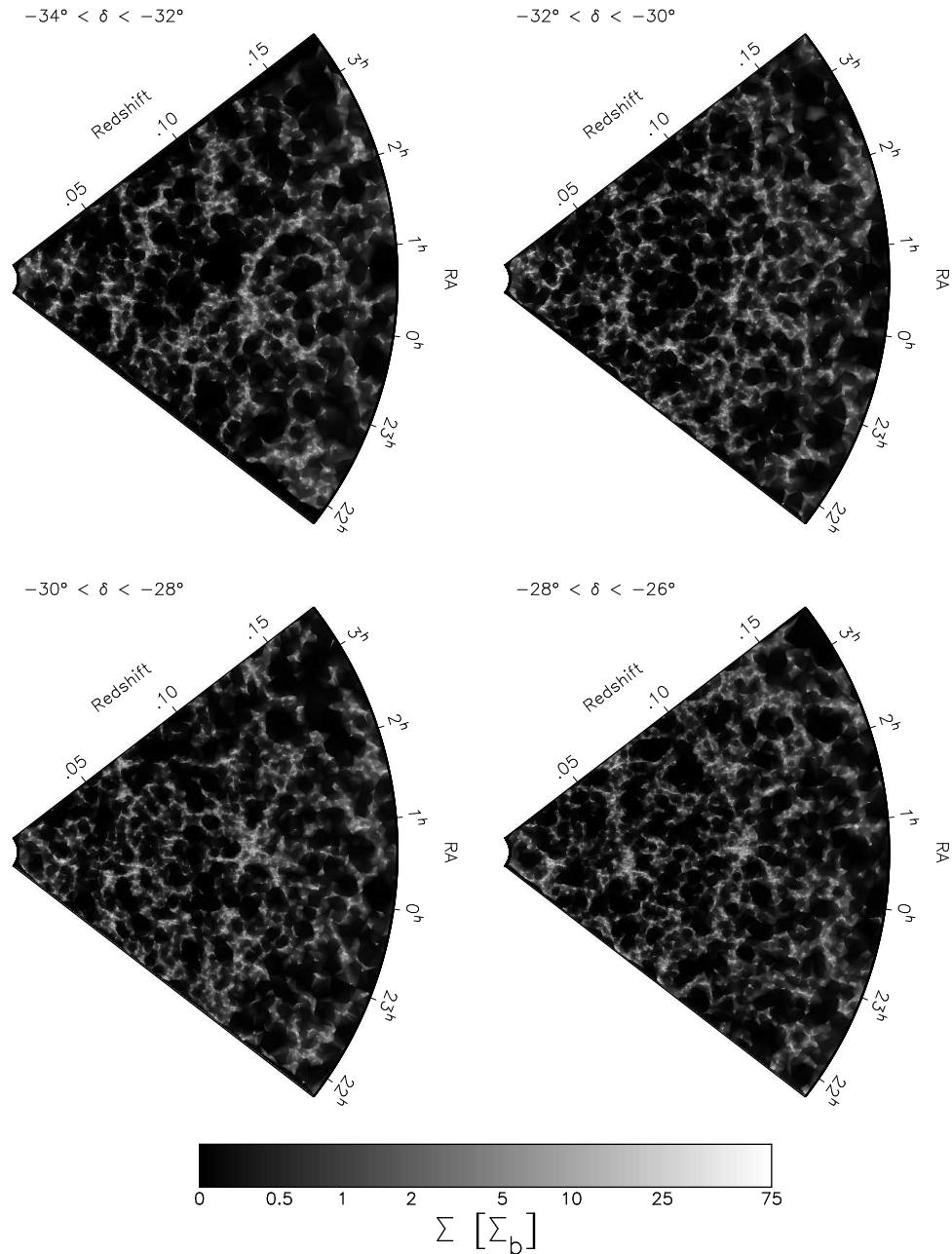


Figure 7.13 — The SGP 2dFGRS galaxy surface density field in several selected slices with a declination width of 2° . The color bar indicates the galaxy number density in units of the average background galaxy surface density (see text for a description).

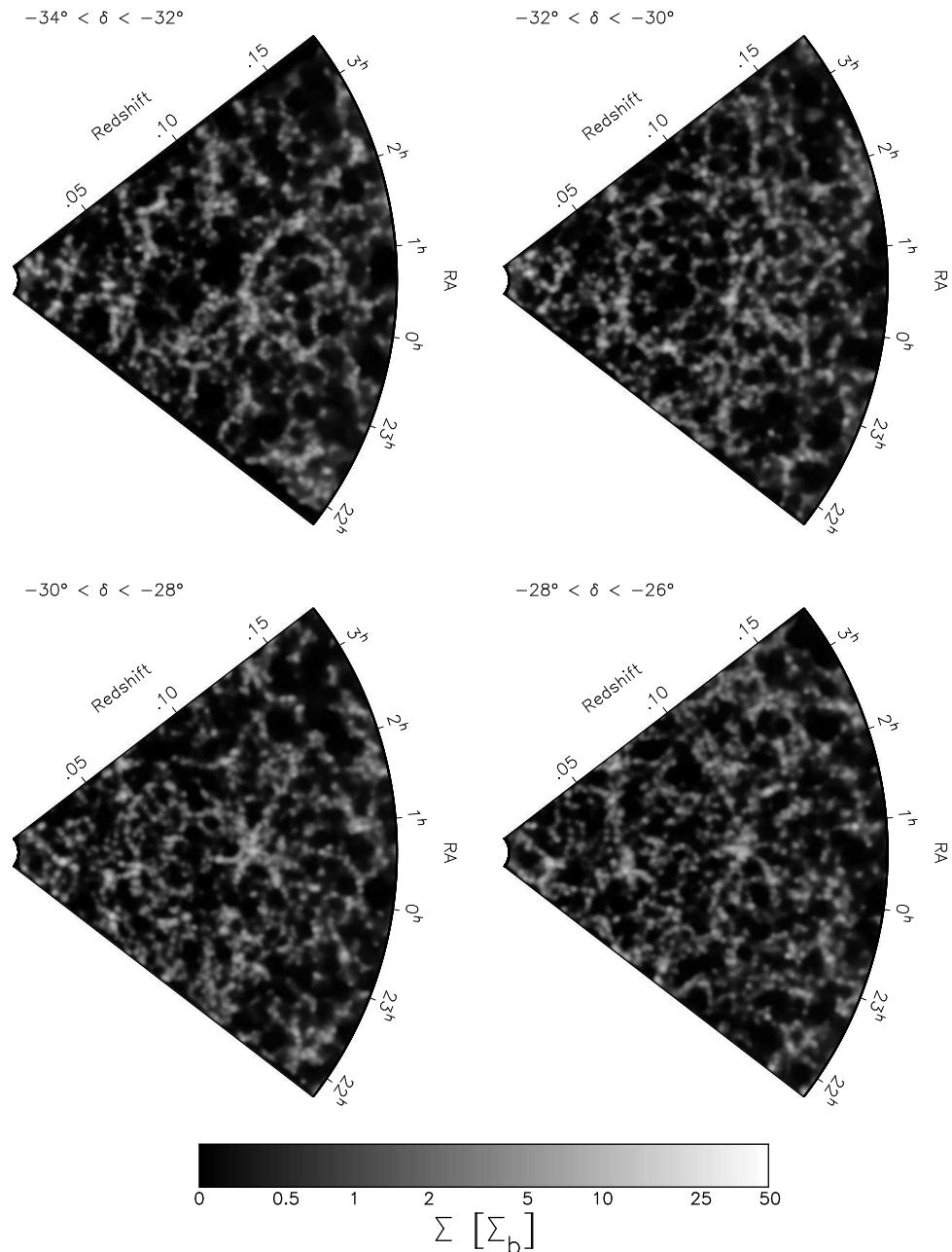


Figure 7.14 — Same as Fig. 7.13, smoothed with a Gaussian filter with an FWHM of $5h^{-1}$ Mpc.

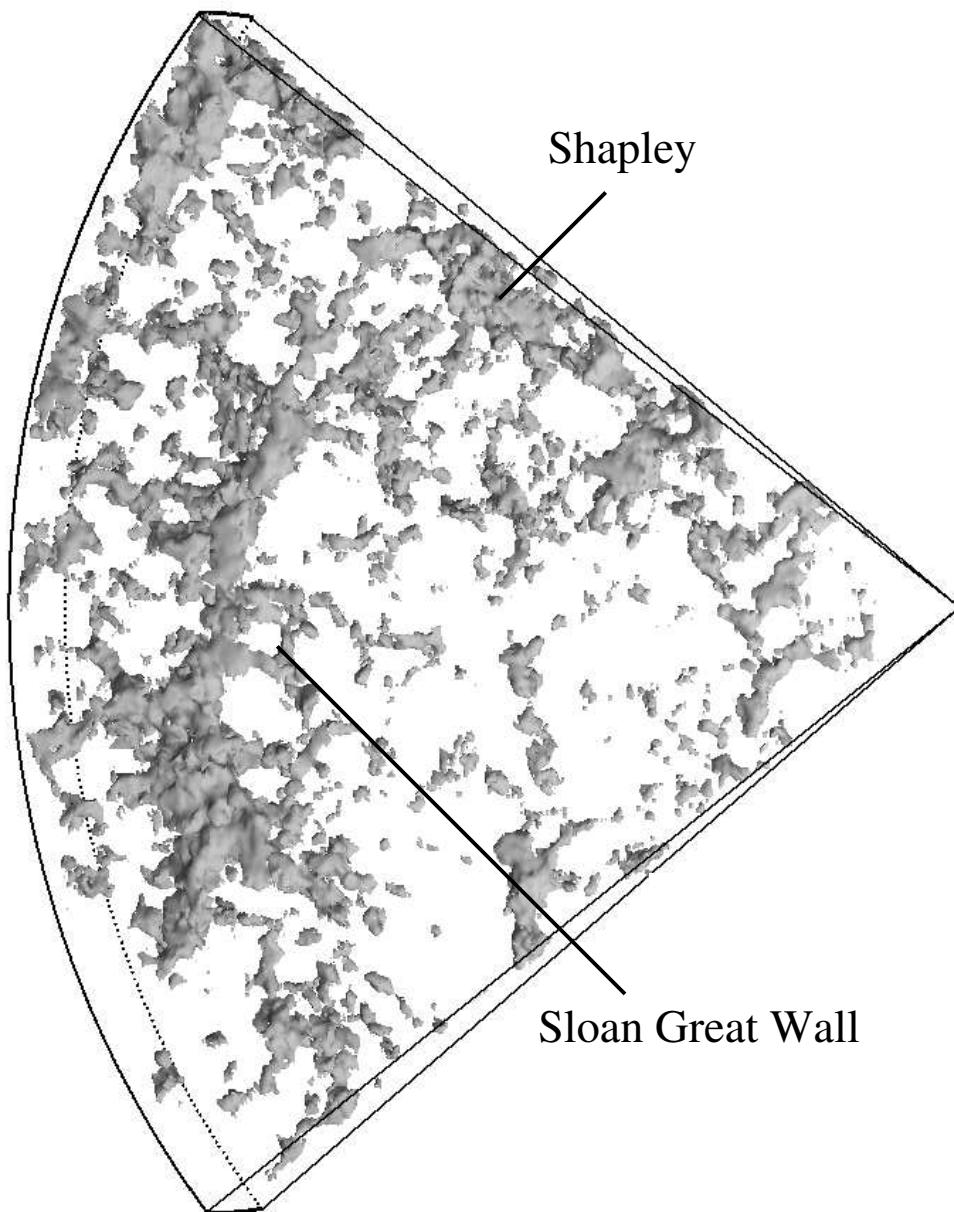


Figure 7.15 — The DTFE reconstructed NGP 2dFGRS galaxy density field. The iso-density contour shown corresponds to twice the background density. Several well-known structures are indicated.

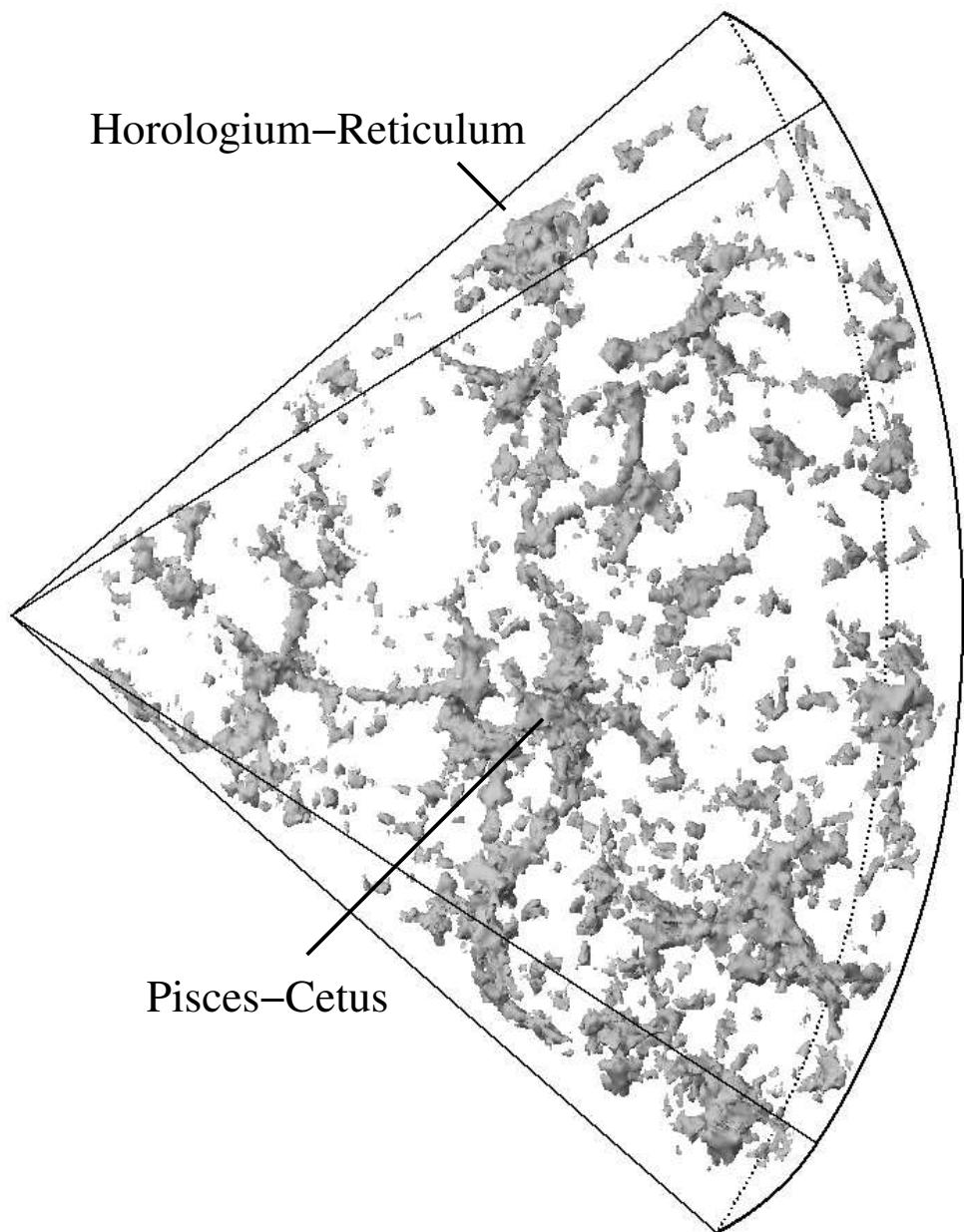


Figure 7.16 — The DTRE reconstructed SGP 2dFGRS galaxy density field. The iso-density contour shown corresponds to twice the background density. Several well-known structures are indicated.

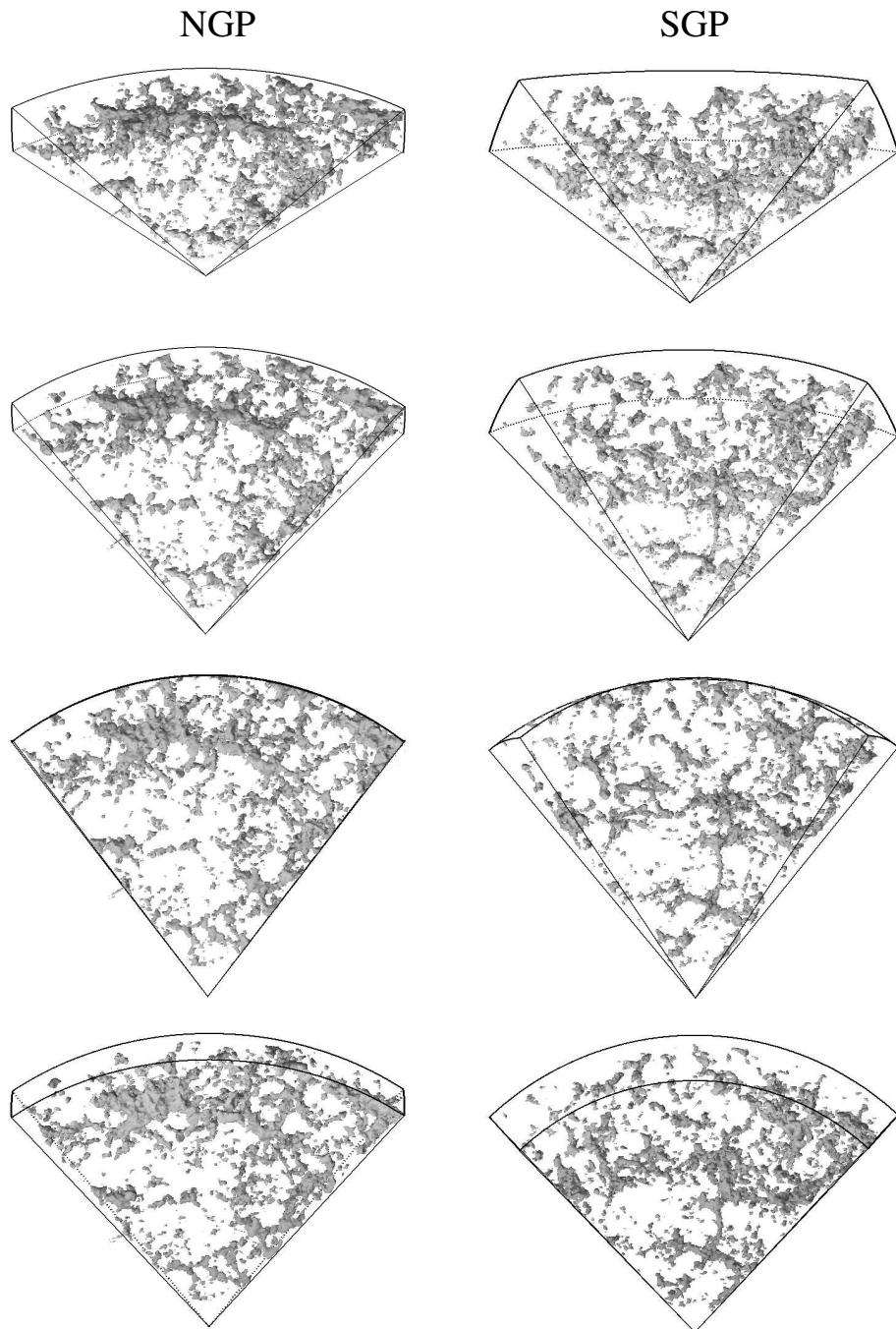


Figure 7.17 — The DT_E reconstructed NGP and SGP 2dFGRS galaxy density field at different viewing angles. The iso-density contour shown corresponds to twice the background density. The second row has the same viewing angle as Figs. 7.15 and 7.16.

The rendering of the large scale galaxy distribution in Figs. 7.15 and 7.16 show that the DTFE is able of recovering the three-dimensional structure of the cosmic web as well as its individual elements. The NGP region is dominated by the large supercluster at a redshift of about 0.8, the Sloan Great Wall (Gott et al. 2005). The structure near the upper edge in Fig. 7.15 at a redshift of 0.05 to 0.06 is part of the upper edge of the Shapley supercluster. In the SGP region several known superclusters can be recognized as well. The supercluster in the center of this region is part of the Pisces-Cetus supercluster. The huge concentration at the top in Fig. 7.16 at a redshift of about 0.07 is the upper part of the enormous Horologium-Reticulum supercluster.

Although less obvious than for the two-dimensional reconstructions, the effective resolution of the three-dimensional reconstructions is also varying across the map. Here the interpretation of the reconstructed maps is further complicated by the fact that the DTFE is not able of handling completely empty regions.

7.9 Summary and discussion

In this chapter we have described the DTFE reconstruction of both the two-dimensional galaxy surface density field and the three-dimensional galaxy density field corresponding to the 2dF galaxy redshift survey. We have argued that the DTFE is well suited for the reconstruction of the large scale galaxy distribution. It is capable of accurately describing the key characteristics of the large scale galaxy distribution. The detailed renderings of the large scale galaxy distribution shown in Fig. 7.7 and in Figs. 7.15 and 7.16 show that the DTFE manages to recover the overall three-dimensional foam-like structure of the cosmic web as well as its individual elements and their intricate interdependencies. Telling examples are the filamentary structures in frame 3 of Fig. 7.8.

The galaxy density maps presented in this chapter show that the high spatial resolution of the DTFE makes it rather sensitive to sampling noise. At the smallest scales this sampling noise is visible as the triangular imprint of the linear DTFE interpolation procedure. The sensitivity of the DTFE to sampling noise makes it of crucial importance to be able to determine the statistical significance of reconstructed structures. This is described in Chapter 8. A possible way of reducing the amount of sampling noise is by processing the DTFE reconstructed density fields, e.g. by applying a filter. In principle extensions of the DTFE procedure to higher-order interpolation are possible as well. A particularly promising example is natural neighbor interpolation (Sibson 1981) which produces smooth fields. Such higher-order interpolation schemes have already been successfully implemented for two-dimensional problems in the fields of geophysics (Sambridge et al. 1995, Braun & Sambridge 1995) and solid state physics (Sukumar 1998). However, for various reasons the implementation of such an algorithm in the context of reconstructing density fields is not trivial.

An obvious next step would be to reconstruct the galaxy density field in real space in stead of in redshift space. The latter is distorted by the peculiar velocity field (Kaiser 1987, Hamilton 1998). On small scales the high gravitationally induced peculiar velocities of galaxies in clusters cause a smearing along the line of sight (fingers of God). These cluster motions are highly non-linear and the redshifts of cluster galaxies therefore cannot be straightforwardly mapped to real space locations. Instead one has to assume a dynamical model for the galaxy clusters and assign galaxy locations accordingly. On larger scales infall motions at large

distances from clusters or around filamentary and wall-like structures induce a compression along the line of sight. Conversely, the expansive motions inside voids have the opposite effect and cause a stretching along the line of sight. To be able to transform from redshift to real space the mass distribution responsible for these motions has to be modelled. However, a severe limitation of redshift surveys is that these flows may as well be produced by structures lying outside the survey volume and one needs to incorporate data from other, more extended all-sky surveys. Such an analysis falls beyond the scope of this chapter.

Once the transformation to real space has been done, the high resolution DTFE reconstruction of the large scale galaxy density field may be used in a variety of subsequent analyses. As yet, a physical description of the properties of the cosmic foam has mostly been in qualitative terms. The quantitative studies that do exist have usually been expressed in terms of statistical and ensemble averaged quantities. For a proper understanding of the large scale structure of our universe it is crucial to study both the large scale topology as well as the local structure of the cosmic web and its characteristic constituent elements. Such studies will not only provide insight into the properties and formation of the cosmic web but may also provide important clues to where and how galaxies have formed.

The DTFE is ideally suited for a systematic analysis of the properties of individual elements of the cosmic web. As it is not beset by artificial smoothing and automatically adapts to the local density and geometry of the galaxy distribution, it provides an optimal description of local structure, be they high density clusters, anisotropic filaments and walls, or low density voids. Figs. 7.5 and 7.8 form a telling illustration of this point. Indeed, Aragón-Calvo & van de Weygaert (2006) have devised an advanced algorithm, the Multiscale Morphology Filter, with the purpose of identifying clusters, filaments and walls in galaxy redshift surveys and cosmological N -body simulations. Another technique, the Watershed algorithm (Platen & van de Weygaert 2006), can identify void regions. Both techniques are based on the DTFE density field reconstruction.

A particularly prominent characteristic of the large scale structure matter distribution is its interconnectedness. One of the ways in which the topological properties of the large scale galaxy distribution can be studied is by using the percolation properties of the galaxy density field (see e.g. Zel'dovich 1982, Shandarin & Zel'dovich 1989, Sahni et al. 1997, Sheth et al. 2003, Shandarin et al. 2004, Pandey & Bharadwaj 2005). These percolation properties will depend on non-linear effects such as the large scale coherence or the local sub-clumping of structure into smaller entities. The percolation properties will also depend on the structure of the characteristic building blocks of the cosmic web and on how these are connected with the cosmic web. Studies of the percolation properties of the large scale galaxy distribution will therefore provide important information on the properties of the cosmic web.

Previous studies of the percolation properties of the large scale galaxy distribution have been hampered by the use of inappropriate reconstruction techniques with very poor numerical resolution. For example, in Shandarin et al. (2004) a fixed grid-based reconstruction technique has been used. This results in a density field in which most characteristic structures have been smeared out. The anisotropic filaments and walls have been reconstructed as low level, extended and roundish structures. Such inappropriate reconstruction techniques will lead to significant distortions of the percolation properties of the resulting reconstructed density field. Since the DTFE reconstruction technique does not make use of artificial filtering, its use may yield a significant improvement over conventional studies.

The results of this chapter show that the DTFE reconstruction procedure is capable of reconstructing the foam-like morphology of the large scale galaxy distribution. With the advent of ever larger and deeper redshift surveys, such as the Sloan Digital Sky Survey (Stoughton et al. 2002) and the 6dF Galaxy Survey (Jones et al. 2004), the DTFE promises to become an important tool for the analysis of the properties of the large scale galaxy distribution. Ultimately the use of such powerful tools will lead to a better understanding of the properties, formation and evolution of the cosmic web.

References

- Adelman McCarthy J.K. et al., 2006, ApJS, 162, 38
 Aragón-Calvo M., van de Weygaert R., 2006, astro-ph/0610249
 Bernardeau F., van de Weygaert R., 1996, MNRAS, 279, 693
 Bond J.R., Kofman L., Pogosyan D., 1996, Nature, 380, 603
 Braun J., Cambridge M., 1995, Nature, 376, 655
 Colberg J.M. et al., 2006, MNRAS, 360, 1
 Cole S. et al., 2005, MNRAS, 362, 505
 Colless M. et al., 2001, MNRAS, 328, 1039
 Colless M. et al., 2003, astro-ph/0306581
 Delone B.N., 1934, Bull. Acad. Sci. USSR: Classe Sci. Mat, 7, 793
 Dodelson S. et al., 2002, ApJ, 572, 140
 Dubinski J., Da Costa L.N., Goldwirth D.S., Lecar M., Piran T., 1993, ApJ, 410, 458
 Einasto M., Tago E., Jaaniste J., Einasto J., Andernach H., 1997, A&AS, 123, 119
 Erdogdu P. et al., 2004, MNRAS, 352, 939
 Gott J.R. III et al., 2005, ApJ, 624, 463
 Hamilton A.J.S., 1998, in *The evolving universe*, Hamilton D. (ed.), Kluwer Academic Publishers, Dordrecht, NL
 Hockney R., Eastwood J., 1988, *Computer Simulations Using Particles*, Taylor & Francis, Bristol, PA, USA
 Hoffman Y., Ribak E., 1991, ApJ, 380, 5
 Hoyle F., Vogeley M., 2002, ApJ, 566, 641
 Hoyle F., Vogeley M., 2004, ApJ, 607, 751
 Huchra J.P., Geller M.J., de Lapparent V., Corwin H.G. Jr., 1990, ApJS, 72, 433
 Jones D.H. et al., 2004, MNRAS, 355, 747
 Kaiser N., 1987, MNRAS, 227, 1
 Lin H. et al., 1996, ApJ, 471, 617
 Maddox S.J., Efstathiou G., Sutherland W.J., Loveday J., 1990a, MNRAS, 242, 43
 Maddox S.J., Efstathiou G., Sutherland W.J., Loveday J., 1990b, MNRAS, 243, 692
 Maddox S.J., Efstathiou G., Sutherland W.J., 1990c, MNRAS, 246, 433
 Monaghan J.J., 1992, ARAA, 30, 543
 Norberg P. et al., 2001, MNRAS, 328, 64
 Norberg P. et al., 2002, MNRAS, 332, 827
 Pandey B., Bharadwaj S., 2005, MNRAS, 357, 3
 Park C., Vogeley M.S., Geller M.J., Huchra J.P., 1994, ApJ, 431, 569
 Patiri S.G. et al., 2006, MNRAS, 369, 335
 Percival W.J. et al., 2001, MNRAS, 327, 1297
 Platen E., van de Weygaert R., 2006, *private communication*
 Regos E., Geller M., 1991, ApJ, 377, 14
 Rosati P., Borgani S., Norman C., 2002, ARAA, 40, 593
 Rybicki G.B., Press W.H., 1992, ApJ, 398, 169

- Sahni V., Sathyaprakash B.S., Shandarin S.F., 1997, ApJ, 476, L1
Sambridge M., Braun J., McQueen H., 1995, Geophys. J. Int., 122, 837
Shandarin S.F., Zel'dovich Y.B., 1989, Rev. Mod. Phys., 61, 185
Shandarin S.F., Sheth J.V., Sahni V., 2004, MNRAS, 353, 162
Shectman S.A., Landy S.D., Oemler A., Tucker D.L., Lin H., Kirshner R.P., Schechter P.L., 1996, ApJ, 470, 172
Sheth J.V., Shani V., Shandarin S.F., Sathyaprakash B.S., 2003, MNRAS, 343, 22
Sheth R.K., van de Weygaert R., 2004, MNRAS, 350, 517
Sibson R., 1981, in *Interpreting Multi-Variate Data*, Barnet V. (ed.), Wiley, Chichester, UK
Stoughton C. et al., 2002, AJ, 123, 485
Sukumar N., 1998, PhD Thesis, Northwestern University, Evanston, IL, USA
Szomoru A., van Gorkom J., Gregg M.D., Strauss M.A., 1996, AJ, 111, 2150
Tegmark M., Hamilton A.J.S., Xu Y., 2002, MNRAS, 335, 887
Tegmark M. et al., 2004, ApJ, 606, 702
Tyson J.A., Kochanski G.P., Dell'Antonio I.P., 1998, ApJ, 498, L107
van de Weygaert R., 1991, PhD Thesis, University of Leiden, Leiden, The Netherlands
van de Weygaert R., van Kampen E., 1993, MNRAS, 263, 481
van de Weygaert R., Bertschinger E., 1996, MNRAS, 281, 84
van de Weygaert R., 2002, *Froth across the Universe*, in '2nd Hellenic Cosmology Meeting',
Plionis M., Cotsakis S. (eds.), Kluwer Academic Publishers, Dordrecht, NL
Zaroubi S., Hoffman Y., Fisher K.B., Lahav O., 1995, ApJ, 449, 446
Zel'dovich Y.B., 1982, Soviet Astron. Lett., 8, 102

8

Sampling Noise, Significance and Error Estimation

W.E. Schaap & R. van de Weygaert

ABSTRACT — We discuss the statistical properties of the DTFE and show how the significance of structures in DTFE reconstructed homogeneous, inhomogeneous and filtered density fields can be determined. We describe how different types of sampling errors affect DTFE reconstructed fields. We describe measurement errors in the location or weight of the sampling points as well as systematic errors due to inhomogeneous sampling or sampling distortions. Finally, we summarize and discuss our results.

8.1 Introduction

In this thesis we have described the Delaunay Tessellation Field Estimator (DTFE), a method for reconstructing volume-covering and continuous density fields which are sampled by a discrete set of points. The DTFE is based on a basic concept of stochastic and computational geometry, the Delaunay tessellation of the point sample (Delone 1934, for further references see also Okabe et al. 2000, Møller 1994 and van de Weygaert 1991). This spatial volume-covering division of space into mutually disjunct triangular (in two dimensions) or tetrahedral (in three dimensions) cells adapts to the local density and geometry of the point distribution. The DTFE exploits this virtue and adapts automatically to changes in the density or the geometry of the distribution of sampling points. The Delaunay tetrahedra are used to obtain local estimates of the spatial density and as multi-dimensional intervals for linear interpolation of the field values estimated at the location of the sampling points (this interpolation procedure was introduced by Bernardeau & van de Weygaert 1996).

The DTFE has a self-adaptive spatial resolution. In Chapters 3 and 4 we have shown that its effective smoothing kernel is more localized than that of other reconstruction methods. It is therefore less forgiving with respect to sampling noise and errors. If present these will have a direct impact on the reconstructed field. This makes it necessary to understand how reconstructed structures are affected and how the significance of DTFE reconstructed density fields can be determined.

In the first half of this chapter we discuss the significance of DTFE reconstructed density fields. We restrict ourselves to Poisson sampling noise, which one for example encounters in N -body simulations and galaxy redshift surveys. Other types of sampling noise usually depend on a specific application and fall beyond the scope of this work. We discuss a number of frequently occurring situations, such as the determination of the significance of a signal in a uniform background field and the fitting of a model to an observed data set. We also discuss the determination of the significance of filtered fields. Finally, we discuss how the significance of DTFE reconstructed fields is related to the sampling density.

In the second half of this chapter we discuss the effects of sampling errors on DTFE reconstructed density fields. First we look at measurement errors, which may be present for the locations and the weights of the sampling points. Subsequently we consider systematic sampling errors, such as inhomogeneous sampling procedures and sampling distortions.

8.2 Sampling noise and significance

In this section we work out the significance of DTFE reconstructed density fields. In order to understand how sampling noise affects reconstructed density fields it is important to realize that even uniform fields are affected by sampling noise, because due to the Poisson nature of the sampling procedure different regions will contain more sampling points than others. This is illustrated in Fig. 8.1, in which several uniform Poisson point samplings are shown together with the corresponding Voronoi and Delaunay tessellations as well as the zeroth-order VTFE and first-order DTFE reconstructed density fields. One may observe density fluctuations in both the point samples, the tessellations and the VTFE and DTFE density fields.

Clearly the magnitude and frequency of these fluctuations determine how significant a particular reconstructed density value is. Very high and very low density values occur very rarely and are therefore rather significant. We will base our measure of significance on a

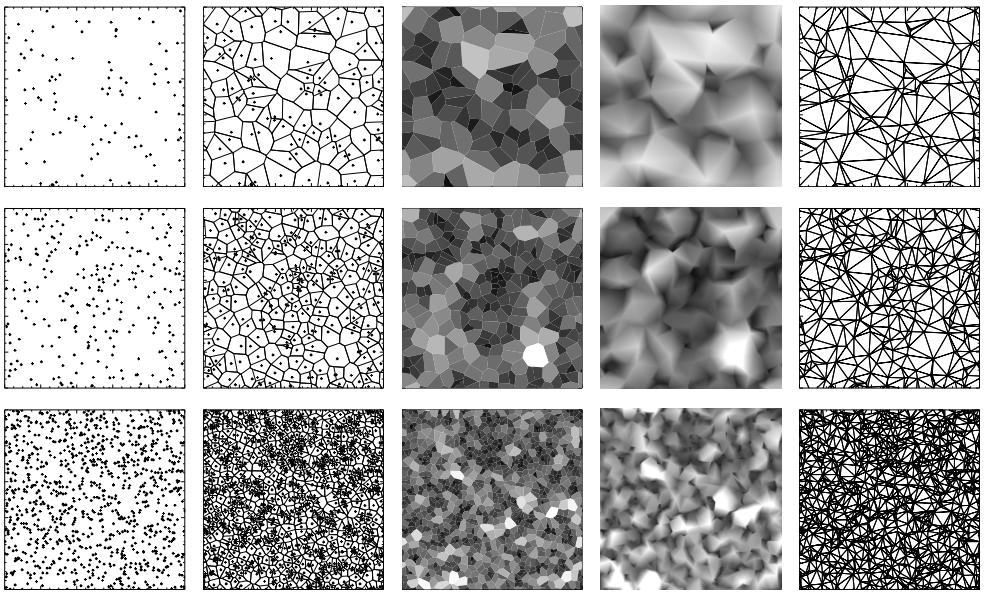


Figure 8.1 — Poisson sampling noise in uniform fields. The rows illustrate three Poisson point samplings of a uniform field with increasing sampling density (from top to bottom consisting of 100, 250 and 1000 points). From left to right the point distribution, the corresponding Voronoi tessellation, the zeroth-order VTFE reconstructed density field, the first-order DTFE reconstructed density field and the corresponding Delaunay tessellation are shown.

consideration of the one-point distribution function of the DTDFE density field reconstruction of a Poisson sampling of a uniform field, which describes the probability that a particular density value occurs due to Poisson fluctuations.

8.2.1 Statistical properties of Poisson sampling noise

Consider a Poisson point sampling of a uniform field. Kiang (1966) has shown that the distribution of sizes of Voronoi cells of such a point sampling may be approximated by

$$dp(\tilde{a}) = \frac{c}{\Gamma(c)} (c\tilde{a})^{c-1} e^{-c\tilde{a}} d\tilde{a}. \quad (8.1)$$

Here $\tilde{a} = a/\langle a \rangle$ is the size of the Voronoi cell in units of the average cell size. The value of c , a numerical constant, depends on the dimension of space. In two dimensions $c = 4$, in three dimensions $c = 6$. Note that there is some discussion on the precise functional form of $dp(\tilde{a})$ and the value of c (e.g. Tanemura 1988, Járai-Szabó & Néda 2004, see also Okabe et al. 2000 for the properties of Voronoi tessellations corresponding to Poisson point samplings), but these differences are not relevant for the range of interest of \tilde{a} for our purposes.

The probability for a random point to lie inside a cell of size \tilde{a} is equal to $p(\tilde{a}) \times \tilde{a}$. In the zeroth-order Voronoi scheme this is the likelihood for a random point to have density $\lambda = 1/\tilde{a}$. The one-point distribution function of the corresponding zeroth-order VTFE density

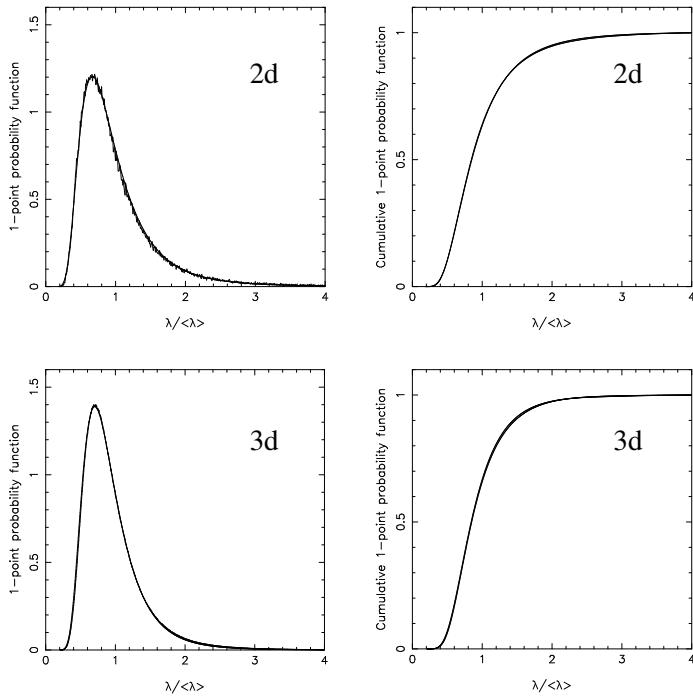


Figure 8.2 — One-point distribution functions of the DTFE reconstructed field and an analytic approximation (smooth curve) corresponding to a Poisson point process of 10 000 points in two (left-hand frame) and 100 000 in three (right-hand frame) dimensions.

field may therefore be approximated by

$$2D : \quad dp(\tilde{\lambda}) = \frac{128}{3} \tilde{\lambda}^{-6} e^{-4/\tilde{\lambda}} d\tilde{\lambda}; \quad (8.2)$$

$$3D : \quad dp(\tilde{\lambda}) = \frac{1944}{5} \tilde{\lambda}^{-8} e^{-6/\tilde{\lambda}} d\tilde{\lambda}. \quad (8.3)$$

It turns out that these equations also form an excellent approximation of the one-point distribution function of the first-order DTFE reconstructed field in two and in three dimensions. This can be seen in Fig. 8.2. It shows the one-point distribution function of the DTFE reconstructed density field corresponding to a homogeneous binomial random field of 100 000 points, with the analytic approximations superimposed. In this figure the cumulative distribution functions are also displayed.

Evidently, the two- and three-dimensional one-point distribution functions are non-Gaussian, with a tail extending toward high densities. Note that the distribution function falls off much more rapidly in three than in two dimensions. Using Eqns. 8.2 and 8.3 one may calculate various statistics corresponding to DTFE reconstructed homogeneous binomial random fields in two and three dimensions, some of which are listed in Table 8.1. The more rapid decline of the three-dimensional one-point distribution function is reflected by the lower value of its variance. The reason for the fact that extreme densities are less likely in three than in two

Table 8.1 — Statistical properties of DTFE reconstructed two- and three-dimensional homogeneous binomial random fields.

$\widehat{\lambda}_{\text{DTFE}}$	2D	3D
mean	1	1
median	0.856	0.900
variance	1/3	1/5
skewness	$2\sqrt{3}$	$\sqrt{5}$
kurtosis	42	12

dimensions is that in three dimensions one effectively puts a constraint on three coordinates for a high density to occur and in two dimensions on only two. The positive value of the skewness for both distributions indicates the presence of the tail extending toward high densities. The larger value for the two-dimensional distribution is again indicative of the slower decline compared to the three-dimensional distribution. Finally, the kurtosis is (strongly) positive for both distributions, showing that they are more strongly peaked than normal distributions. The larger value for the two-dimensional distribution indicates that it is more strongly peaked than the three-dimensional one, presumably due to its relatively pronounced tail.

8.2.2 Significance of uniform fields

The one-point distribution function has a local interpretation as well. Given an ensemble of Poisson samplings of a uniform density field, at any location the distribution of DTFE reconstructed density values is equal to the one-point distribution function. This function may thus be used to assign the probability that a particular reconstructed density value is in accordance with an underlying uniform density field of a particular density.

Here we consider the practical application where one observes a particular point distribution and assumes that this is a realization of a certain uniform field with some signal field superposed. In such a case one is interested in the significance of the signal field. To be able to assign a measure of significance to local estimates of the recovered signal field we proceed as follows. Given a reconstructed density field $\widehat{\lambda}_R(\mathbf{x})$ which is supposed to be the sum of an underlying uniform field λ and a signal field $\sigma(\mathbf{x})$ we evaluate at each location \mathbf{x} the two probabilities

$$P_1(\mathbf{x}) \equiv P(\widehat{\lambda} \leq \widehat{\lambda}_R(\mathbf{x})) = \int_0^{\widehat{\lambda}_R(\mathbf{x})/\lambda} dp(\widehat{\lambda}'); \quad (8.4)$$

$$P_2(\mathbf{x}) \equiv P(\widehat{\lambda} \geq \widehat{\lambda}_R(\mathbf{x})) = \int_{\widehat{\lambda}_R(\mathbf{x})/\lambda}^{\infty} dp(\widehat{\lambda}'). \quad (8.5)$$

$dp(\widehat{\lambda})$ is the one-point distribution function of a DTFE reconstructed uniform field (Eqns. 8.2 and 8.3). At each location P_1 and P_2 represent the probability to find a density value $\widehat{\lambda}$ lower (P_1) or higher (P_2) than $\widehat{\lambda}_R$ in a DTFE reconstruction of a Poisson sampling of a uniform field with density λ . P_1 and P_2 are therefore local measures of the significance of the signal field.

Equations 8.4 and 8.5 can be solved analytically, yielding

$$P_1 = 1 - f\left(\frac{\widehat{\lambda}_R}{\lambda}\right); \quad (8.6)$$

$$P_2 = 1 - P(\widehat{\lambda} \leq \widehat{\lambda}_R) = f\left(\frac{\widehat{\lambda}_R}{\lambda}\right), \quad (8.7)$$

in which the function f is defined as

$$f(x) = \int_x^\infty dp(x') = \begin{cases} 2D: & e^{-4/x} \left[\frac{32}{3x^4} + \frac{32}{3x^3} + \frac{8}{x^2} + \frac{4}{x} + 1 \right] \\ 3D: & e^{-6/x} \left[\frac{324}{5x^6} + \frac{324}{5x^5} + \frac{54}{x^4} + \frac{36}{x^3} + \frac{18}{x^2} + \frac{6}{x} + 1 \right]. \end{cases} \quad (8.8)$$

In Fig. 8.3 the probabilities P_1 and P_2 are plotted as a function of density for two and three dimensions. It is interesting to note that the probability to find a density higher than the average density is smaller than 0.5 in both two and three dimensions, or, equivalently, that the median density is smaller than the average density. This reflects the fact that the probability distribution is asymmetric: low density regions are more extended than high density regions. Also visible is that extreme densities are more likely to occur in two than in three dimensions. Finally, from the figure it is clear that low densities are much more unlikely to occur than high densities, both in two and in three dimensions.

We now proceed to define the significance of the signal field. From Eqns. 8.6 and 8.7 one may observe that P_1 and P_2 are equal when $P_1 = P_2 = 0.5$. Given an ensemble of DTFE reconstructions of Poisson point samplings of a given uniform density field, at any particular location half the reconstructions will have a smaller density and half will have a higher density than the value of the reconstructed density field for which P_1 and P_2 are equal. This value is the median density $\widehat{\lambda}_{\text{median}}$ of a DTFE reconstruction of a Poisson sampling of a uniform field with density λ . We define the significance S of the signal corresponding to the median density to be zero, while smaller and larger signals have a positive significance. The median density corresponds to a signal $\sigma = \widehat{\lambda}_{\text{median}} - \lambda$. Since $\widehat{\lambda}_{\text{median}} < \lambda$ (see Table 8.1) a signal $\sigma = 0$ (for which the reconstructed field has the same value as the underlying uniform field) has a positive significance. This counterintuitive result is the consequence of the asymmetric form of the one-point distribution function (Eqns. 8.2 and 8.3): low density regions are more extended than high density regions.

Next consider reconstructed density values higher than the median density. We define the significance of such densities (or equivalently, of the corresponding signals) to be equal to P_2 , the probability to find in a DTFE reconstructed uniform Poisson field densities which are higher than $\widehat{\lambda}_R$. Similarly we define the significance of reconstructed density values (or equivalently, of the corresponding signals) smaller than the median to be equal to P_1 , the probability to find in a DTFE reconstructed uniform Poisson field densities which are smaller than $\widehat{\lambda}_R$.

It is convenient to express the probabilities P_1 and P_2 in terms of Gaussian standard deviations (σ 's). We convert the significance S of a particular reconstructed density field value to the distance in units of σ from the median density to the density for which the probability to find a point in a Gaussian distribution with a value equal to or more extreme than this density

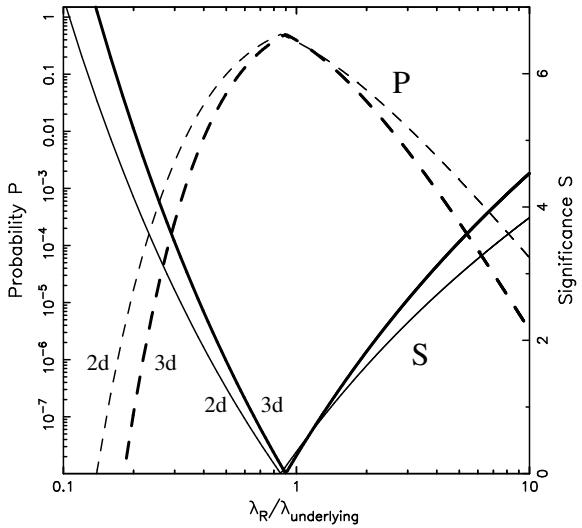


Figure 8.3 — The probability P (dashed lines) and significance S (solid lines, see description in text) to find a more extreme density than a particular density value $\widehat{\lambda}_R/\lambda$ in a DTFE reconstruction of a homogeneous binomial random field. Results are shown for both two-dimensional (thin lines) and three-dimensional (thick lines) fields.

is equal to the probability calculated using Eqns. 8.6 and 8.7. Mathematically,

$$\int_{S\sigma}^{\infty} dt G(t) = \begin{cases} P(\widehat{\lambda} \leq \widehat{\lambda}_R), & \text{if } \widehat{\lambda}_R < \widehat{\lambda}_{\text{median}} \\ 0, & \text{if } \widehat{\lambda}_R = 0 \\ P(\widehat{\lambda} \geq \widehat{\lambda}_R), & \text{if } \widehat{\lambda}_R > \widehat{\lambda}_{\text{median}} \end{cases}; \quad G(t) = \frac{1}{\sigma \sqrt{2\pi}} e^{-t^2/2\sigma^2}. \quad (8.9)$$

Solving this equation yields the following expressions for the significance S :

$$S(\widehat{\lambda}_R) = \begin{cases} \sqrt{2} \operatorname{erfc}^{-1} \left[2f\left(\frac{\widehat{\lambda}_R}{\lambda}\right) - 1 \right], & \text{if } \widehat{\lambda}_R < \widehat{\lambda}_{\text{median}} \\ 0, & \text{if } \widehat{\lambda}_R = \widehat{\lambda}_{\text{median}} \\ \sqrt{2} \operatorname{erfc}^{-1} \left[1 - 2f\left(\frac{\widehat{\lambda}_R}{\lambda}\right) \right], & \text{if } \widehat{\lambda}_R > \widehat{\lambda}_{\text{median}}. \end{cases} \quad (8.10)$$

For general values of $\widehat{\lambda}_R$ this equation has to be solved numerically. In Fig. 8.3 the significance is plotted as a function of density for two and three dimensions.

Interestingly, Eqn. 8.10 does not contain any dependence on the sampling density. Naively one might think that the influence of sampling noise diminishes for a larger number of sampling points. However, since the DTFE is a local procedure, the magnitude of the density fluctuations due to sampling noise is independent of the sampling density, only their spatial scale changes. The effects of sampling noise may be suppressed by filtering the reconstructed field. The effects of filtering on the significance of reconstructed features is discussed in section 8.2.4.

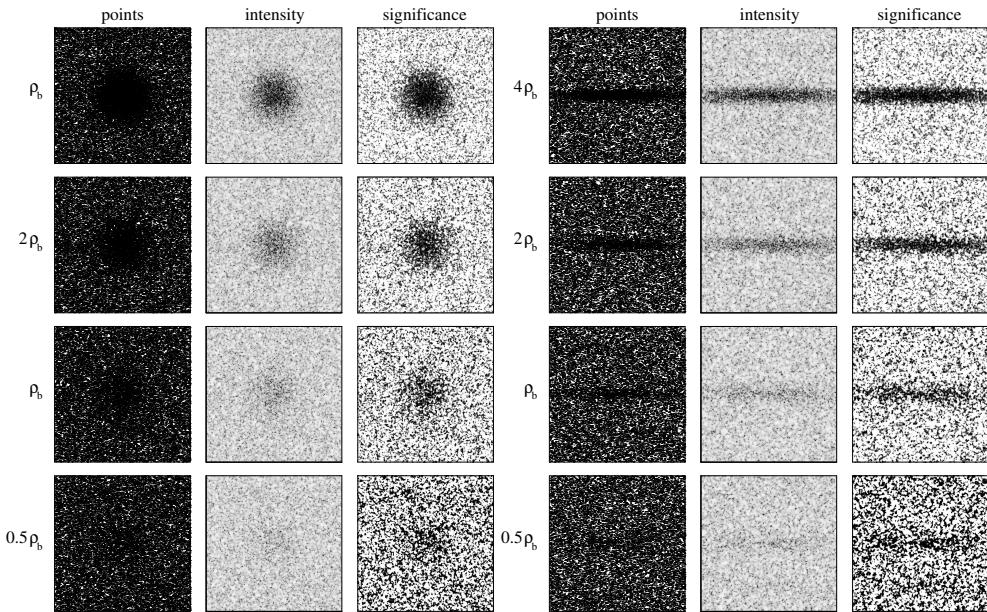


Figure 8.4 — Simulation of the observation of a Gaussian (left-hand panel) and an 10 : 1 ellipsoidal (right-hand panel) object in a Poisson background. Left-hand column in each panel: simulated observed point distributions. Going from top to bottom the Gaussian structure has peak-to-background ratio of 4, 2, 1 and 0.5. Central column in each panel: DTFE reconstructed density fields. Right-hand column in each panel: significance maps of the reconstructed density fields. Only the significance of positive deviations has been plotted.

8.2.2.1 Example: significance of a detected signal in a uniform background

As an example of how the above formalism works in practice we have simulated the observation of a number of Gaussian and ellipsoidal objects in a Poisson background, with varying ratios between the density of the background and the peak density of the object. The simulated observed point distributions are shown in Fig. 8.4. In the left-hand panel of Fig. 8.4 a Gaussian object is shown with, going from top to bottom, a peak-to-background ratio (P/B) of 4, 2, 1 and 0.5. In the right-hand panel the same is done for an ellipsoidal object with an axis ratio of 10 : 1 and the same peak-to-background ratios as for the Gaussian object. In Table 8.2 we have listed a number of parameters for these simulations, including the relative amount of points belonging to the Gaussian and ellipsoidal structures with respect to the background, for the frames shown in Fig. 8.4. In the central columns the DTFE reconstructed density fields are displayed. In the right-hand columns the significance of the reconstructed field values is plotted. As we are interested in a signal on top of the background, we have only displayed the significance values for positive deviations from the background.

In Fig. 8.4 the objects with P/B = 4 and P/B = 2 are conspicuously visible, both in the point distributions as well as in the field reconstructions and even more so in the significance maps. The same is true to a lesser extent for the object with P/B = 1. In this case the significance maps are the most clear indicator of the presence of these objects. The objects

Table 8.2 — Parameters for the simulations of Gaussian and ellipsoidal objects in a Poisson background. Listed is the number of points belonging to these objects as well as their relative amount with respect to the total number of points.

peak-to-background	Gauss	Percentage	Ellipsoid	Percentage
4.0	6283	20,4%	4734	20,8%
2.0	3141	11,4%	2370	11,6%
1.0	1570	6,0%	1186	6,2%
0.5	785	3,1%	603	3,2%

with $P/B = 0.5$ are not visible in the point distribution and only a hint of them in the field reconstructions. However, the significance maps reveal a stronger indication of the presence of an object. The last example shows that the DTDE manages to pick up objects below the level of the background. Nevertheless, in all cases it is clear that the reconstructed fields are strongly influenced by the background noise, which is the result of the high spatial resolution of the DTDE procedure. In practical applications one would often proceed by fitting a model to the observed signal (see section 8.2.3) or by filtering the reconstructed density field (see section 8.2.4).

8.2.3 Significance of inhomogeneous fields

In the previous section we have analyzed the significance of a detected signal on top of a uniform background field. Another frequently occurring practical situation is the fitting of a model to a particular point sampling. In such a case one determines the significance of the deviations of the model and evaluates whether these are likely to be caused by sampling noise or whether these are statistically significant. It is important to realize that it is not possible to directly determine the significance of the model itself, one can only rule out models to a particular degree by evaluating the significance of the deviations. The previously discussed example of a signal in a uniform background is in fact a special case in which the model is a uniform field and one evaluates the significance of the deviation of the signal from the uniform background (see section 8.2.2).

For a proper analytical description of inhomogeneous fields one should take into account the multi-point distribution functions or moments. However, for practical purposes a proper assessment on the basis of a multi-point distribution is not feasible within the scope of this work. To take into account spatial correlations within the field would quickly lead to a cumbersome mathematical treatment, not necessarily going along with a clearly defined significance (see e.g. the discussion in Gaztañaga 1989, 1992, 1994, Lemson 1995).

To be able to assign a measure of significance to local estimates of a recovered density field, we simply proceed by assuming that the field is locally uniform. In such a case one may determine at each location the significance using the formalism described in section 8.2.2 after replacing the uniform underlying density field by an inhomogeneous density field.

The assumption that the field is locally uniform clearly is valid when the sampling density is high enough for the density field not to fluctuate significantly over the extent of a Delaunay

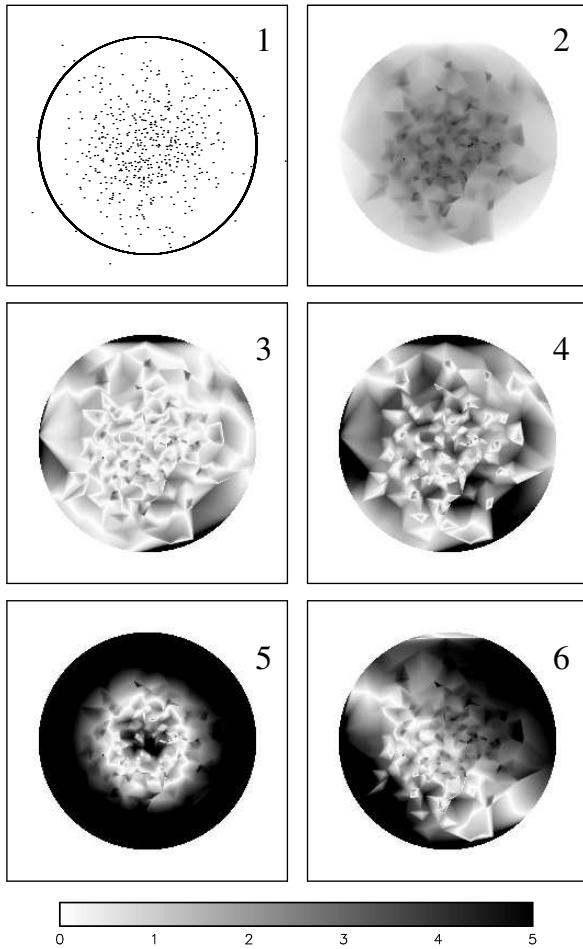


Figure 8.5 — Gaussian model fitting. Frame 1: Poisson sampling of a Gaussian density profile by 500 points. The circle denotes the outermost radius R_0 at which the Gaussian is sampled relatively well. Frame 2: DTNE reconstruction of the region inside R_0 . Frames 3 to 6: significance maps of the deviations of the DTNE reconstruction with four different models. Frame 3: model equals the underlying profile. Frame 4: model with a too large mass. Frame 5: Model with a too small spatial extent. Frame 6: model with an offset central position. The amplitude of the significance is indicated by the color bar.

triangle or tetrahedron:

$$\frac{|\nabla \lambda| \cdot \lambda^{-D}}{\lambda} \ll 1, \quad (8.11)$$

in which D is the dimension of space. If this condition is not valid then the calculated significance will be an upper limit and Monte Carlo simulations may be used to get an estimate of the true significance. An example of how Monte Carlo simulations can be used for determining the significance of a DTNE reconstructed density field is discussed in section 8.3.1.

8.2.3.1 Example: Gaussian model fitting

As an example we have fitted a number of models to the DTNE reconstruction of a point sampling following a Gaussian profile:

$$\lambda(\mathbf{x}) = \frac{M}{\pi\sigma^2} \exp \left\{ -\left(\frac{\mathbf{x} - \mathbf{x}_0}{\sigma} \right)^2 \right\}. \quad (8.12)$$

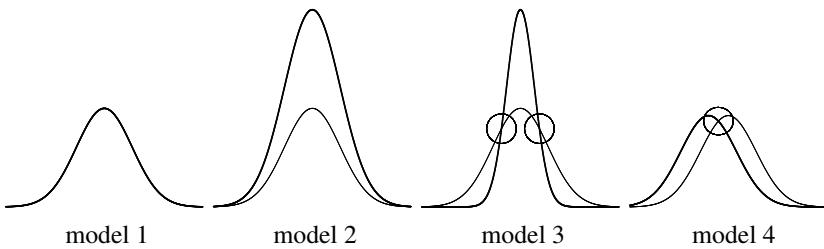


Figure 8.6 — The four models used for the Gaussian model fitting. Model 1 equals the underlying distribution. The thin line in models 2 to 4 indicates the underlying distribution. The circles plotted in model 3 and model 4 indicate where model and underlying distribution have a comparable amplitude.

The realization consists of 500 points and is shown in frame 1 of Fig. 8.5. The circle corresponds to the outermost ring which is still properly sampled. It can be shown that this radius is in fact somewhat larger than appropriate for DTFE reconstructions, but it will serve to illustrate the deviations between the reconstruction and the underlying model which arise due to the finite number of sampling points. The DTFE reconstruction is shown in frame 2.

We have fitted four models to this reconstruction. All consist of Gaussian profiles, given by Eqn. 8.12. The models are graphically illustrated in Fig. 8.6, while their parameters are listed in Table 8.3. Model 1 corresponds to the underlying profile, model 2 has a larger mass, model 3 has smaller spatial extent and in model 4 the position of the center is offset. For each of these models we have calculated the significance of the deviations of the DTFE reconstruction with respect to the model. The resulting significance maps are shown in Fig. 8.5, frames 3 to 6.

The first thing to note is that there are strong differences between the four models. Naturally we have adopted model parameters which result in clearly distinct profiles, so that the differences between the models stand out. Model 1, the underlying profile, fits best. Most deviations have a small significance (< 2) and only few locations have a large significance (> 3). These correspond to small areas, usually confined to a single triangle. An exception to this are the outer edges, which deviate from the model with a large significance. This is due to the poor sampling in this area. Also note that the scale of the variations is varying, from relatively small in the central regions to larger variations in the outer regions. This is obviously due to the variation of the sampling density, which decreases as a function of distance to the center, resulting in a lower effective resolution towards the edge.

Table 8.3 — Overview of the properties of the models used for fitting a DTFE reconstruction of a inhomogeneous binomial random point sampling of a Gaussian density profile.

	M	σ	x_0
model 1	1	1	0
model 2	2	1	0
model 3	1	0.5	0
model 4	1	1	-0.5σ

The significance map of model 2 looks very similar to the one of model 1, except that the significance of the deviations is much larger everywhere. This is because model 2 is a scaled up version of model 1, with only the amplitude different (see Fig. 8.6). The map of model 3 shows a very different pattern, with a systematically very high significance at the center and the outer regions, while the significance is low in a ring around the center. This ring corresponds to regions where the model happens to have an amplitude which is comparable to that of the underlying field (see Fig. 8.6). The map of model 4 shows a different pattern, with a diagonal band with low significance and a systematically very high significance in the other regions. This band corresponds to regions where the model happens to have an amplitude which is comparable to that of the underlying field (see Fig. 8.6).

In conclusion we have seen that it is possible by calculating the significance maps to discern between different models. In this way one could do a parameter fit for a particular generic type of model (such as the Gaussian profile of Eqn. 8.12) or compare completely different types of models.

8.2.4 Significance of filtered fields

So far we have only discussed the uncertainty due to sampling noise in unprocessed DTFE density field reconstructions. However, in practical applications reconstructed fields are often filtered in order to increase the signal-to-noise ratio. In such cases it becomes rather complex to calculate the significance of reconstructed density values. The reasons are similar to those encountered when determining the significance of inhomogeneous fields: one has to take into account the multi-point distribution functions of the reconstructed fields. The results also depend on the particular shape of the smoothing filter, which may depend on the application.

This problem is rather familiar from the analysis of for example radio observations, in which the observed signal is convolved with an observational beam. There it is customary to define the significance of an observed signal in units of the variance of the signal in a homogeneous part of the sky which does not contain distinct radio sources. This variance is a measure of the probability that a certain signal occurs due to background noise. Note that due to the particular parameters specific for the observations, this variance is not usable for other observation runs or other telescopes. Moreover, due to the spatial extent of the beam the variance is not uniquely related to the probability for such a signal to occur due to the background noise. This makes the interpretation of the significance measured in units of the variance somewhat subjective. Nevertheless, the straightforwardness and conceptual simplicity of the variance has made it a reliable and standard discriminator of which regions are statistically significant and which are not.

In order to get an impression of the effect filtering has on the significance of DTFE reconstructed density fields, we have filtered the DTFE reconstruction of a 10 000 point Poisson sampling of a uniform field and measured the variance for a number of different filtering radii. On the left-hand side of Fig. 8.7 the unprocessed DTFE reconstruction is shown (upper left-hand frame) as well as the filtered density fields (subsequent frames). The fields have been filtered by a fixed Gaussian kernel whose FWHM is plotted in the upper left-hand section of each density field. The FWHM is specified in units of the average scale of the DTFE kernel, which has been defined as the square root of the volume of the contiguous Delaunay cell corresponding to a sampling point. On the right-hand side of Fig. 8.7 the measured variance of the filtered fields is plotted as a function of the FWHM of the Gaussian filter.

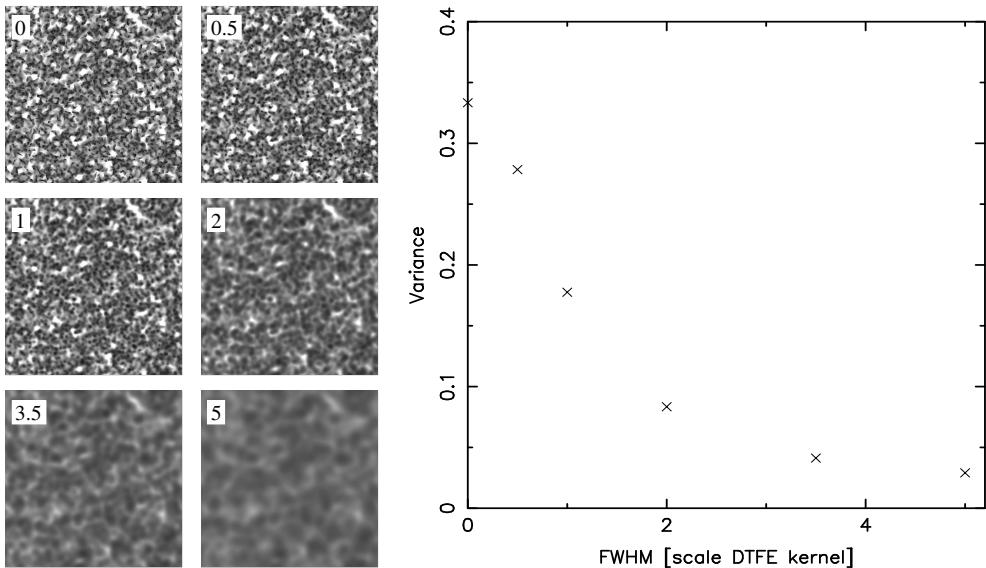


Figure 8.7 — The effect of filtering on the sampling noise in DTFE reconstructed density fields. In the upper left-hand frame a DTFE reconstructed density field corresponding to a 10 000 point Poisson sampling is shown. In the subsequent frames the same field is shown after filtering with a Gaussian filter whose FWHM is plotted in the upper left-hand section of each density field. The FWHM is specified in units of the scale of the DTFE kernel (see text for a description). On the right-hand side the measured variance in these density fields is plotted as a function of the FWHM of the Gaussian filter.

It can be seen that filtering the DTFE reconstructed field by a Gaussian filter with a FWHM up to twice the average size of the DTFE kernel reduces the variance quite substantially, while for larger filters the variance decreases relatively more slowly. This result suggests that one may suppress the sampling noise of a DTFE reconstructed field substantially by filtering it with a Gaussian with a FWHM up to twice the average size of the DTFE kernel. Filtering with a larger kernel is less profitable in terms of reducing the variance of the sampling noise. The reason for this different behaviour is that only for filters which are larger than the average scale of the DTFE kernel the density field is averaged over different contiguous Voronoi cells, while for smaller filters one merely changes the shape, but not the scale of the DTFE kernel.

Note that in general fields are inhomogeneous, which introduces a complication. For such fields using a fixed filter will reduce the amount of sampling noise by a factor which depends on the local density. To circumvent this one may use an adaptive filter.

8.2.4.1 Example: Filtering of a signal in a uniform background

As an example we have filtered the DTFE reconstructed density maps corresponding to the Gaussian and ellipsoidal objects in a Poisson background with P/B = 1 (see section 8.2.2). The resulting filtered maps are plotted in Fig. 8.8. In the left-hand column the unsmoothed fields are plotted, while the subsequent frames show the same images, but smoothed with a

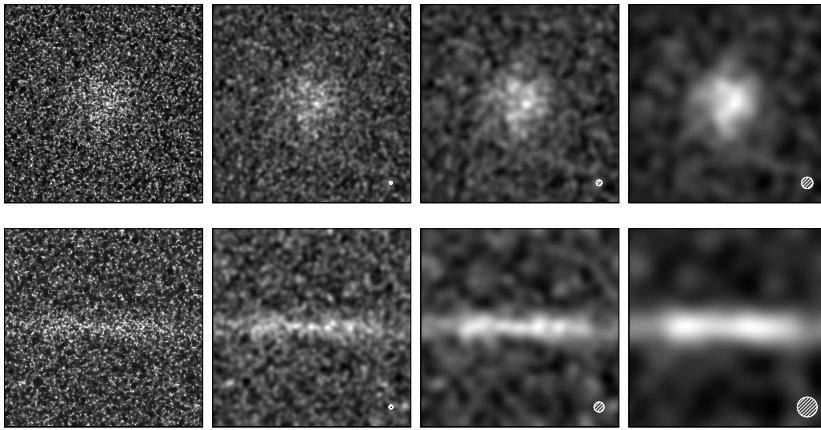


Figure 8.8 — Smoothing of DTRE reconstructed images. In the left-hand column the DTRE reconstructed images of a Gaussian and ellipsoidal object in a Poisson background are shown (for both cases $P/B = 1$). These images have been smoothed with Gaussian filters of varying size. The resulting images are shown in the subsequent columns. The FWHM of the Gaussian filter is denoted by the shaded circle in the lower right-hand corner of each smoothed image. The color scaling of each image has been optimized in each individual frame for an optimal contrast between object and background.

Gaussian filter of increasing size. The diameter of the shaded circle shown in the lower right-hand corner of these frames indicates the FWHM of the filter. The significance of the signal may be expressed in units of the variance due to the background. This variance could either be measured in the filtered fields or be deduced from Fig 8.7.

This example clearly shows how such filtering succeeds in increasing the signal-to-noise levels, but it also shows that the properties of the objects are affected by the size and shape of the smoothing kernel. In particular, the anisotropy of the object seen in the bottom right-hand frame is less pronounced than the true anisotropy of the underlying object. The advantage of using the DTRE as reconstruction tool guarantees that no extra smoothing is introduced by the reconstruction procedure itself and the interpretation of these smoothed maps is therefore relatively straightforward.

8.3 Sampling errors

Apart from being affected by sampling noise, data in practical applications are usually affected by one or more types of sampling errors. An important example concerns measurement errors in the position or the mass of the sampling points. Systematic errors may be introduced by an inhomogenous sampling procedure or by sampling distortions. In this section we work out the effect of these errors on the DTRE reconstruction.

8.3.1 Measurement errors: the location of the sampling points

A common type of error in many types of applications is an uncertainty in the determination of the position of the sampling points. In Fig. 8.9 this is illustrated for the observation of a Voronoi wall model of the large scale galaxy distribution (see Chapter 4 for a description), in which the location of the sampling points has not been measured with absolute certainty,

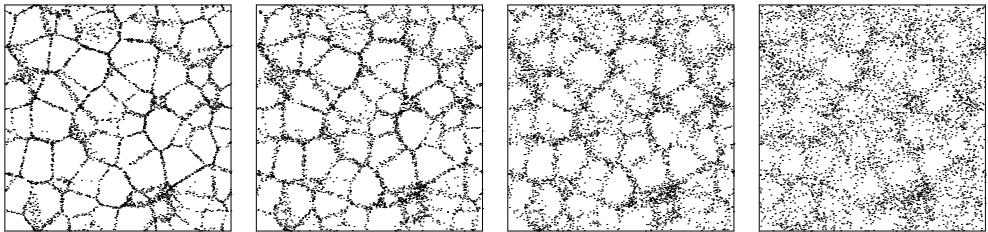


Figure 8.9 — Measurement errors in the position of the sampling points. Shown is a slice through a Voronoi wall model of the large scale galaxy distribution (see Chapter 4 for a description). The position of the sampling points is measured with an uncertainty whose magnitude increases from no error (left-hand frame) to a significant error (right-hand frame).

but with some measurement error. In the figure the error increases from no error (left-hand frame) to a rather large error (right-hand frame). It is obvious that the point distribution does not change significantly when the measurement errors are small, while it appears very differently for large measurement errors.

One might expect that the Delaunay tessellation corresponding to a set of sampling points does not change when measurement errors are small, while it does for large measurement errors. These two cases are illustrated in Fig. 8.10. In the left-hand panel the errors are so small that although the position of each point has changed somewhat, the effects are not large enough for the triangle identities of the tessellation to change. In other words, the vertices of the triangles or tetrahedra in the Delaunay tessellations will still correspond to the same sampling points. If the triangle identities do not change, a formal expression may be written

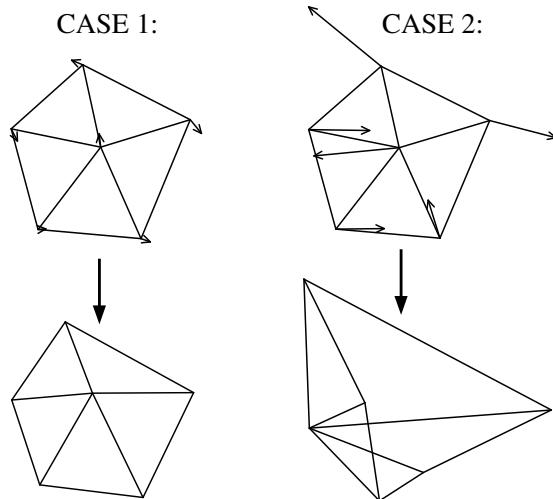


Figure 8.10 — Consequences of changes in the position of vertices for the triangle identities in a Delaunay tessellation. Case 1: the uncertainties in the position of the sampling points are not large enough for the triangle identities to change. Case 2: the uncertainties are so large that the tessellation must have changed.

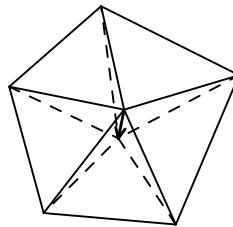


Figure 8.11 — Demonstration that the contiguous Voronoi cell of a point does not change due to a change in its position as long as the triangle identities remain unchanged.

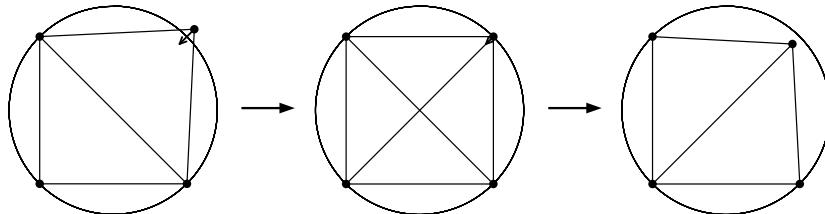


Figure 8.12 — Change of triangle identities due to changes in the position of the sampling points. Shown is a series of three frames depicting a set of four sampling points. The upper right-hand point is moving downwards to the left as denoted by the arrow. In the central frame the tessellation has become degenerate, as the four points have equal distance to each other. In the right-hand frame the triangle identities of the tessellation have changed with respect to the left-hand frame.

down for the uncertainty in the density estimates at the location of the sampling points due to the uncertainty in the position of the sampling points errors. As a side remark, it is interesting to note that for an unchanged tessellation an uncertainty in the position of a point does not lead to an uncertainty in the density estimate at the location of that point, but instead to an uncertainty in the density estimates at the location of its surrounding points. The reason for this is that as a particular point moves around, its contiguous Voronoi cell remains unchanged as long as the triangle identities stay the same, which is illustrated in Fig. 8.11.

In the right-hand panel of Fig. 8.10 the errors are so large that the triangle identities of the tessellation have changed. There it can be seen that the errors lead to a configuration which is not a Delaunay tessellation, so that the triangle identities must have changed. In such cases it is not possible to write down a formal expression for the error for the density estimation at the location of the sampling points. Instead, one should perform Monte Carlo simulations in order to get a handle on the uncertainty in the reconstructed density field values.

Even though these two cases may be discerned in principle, in practice it is usually impossible to tell beforehand if uncertainties in the position determination may lead to identity changes. The reason for this is that in certain configurations it is possible that an infinitesimal change in position leads to a change of triangle identities. One may see this from Fig. 8.12, in which the four sampling points in the left-hand and right-hand frame are close to a degenerate configuration, which is shown in the central frame. In practice one should therefore always perform Monte Carlo simulations in order to get an estimate of the uncertainty in reconstructed density values.

8.3.1.1 Example: Monte Carlo simulations of the observation of anisotropic structures

As an example of how this works in practice, we have simulated the observation of an ellipsoidal structure with axis ratio of 10 : 1, in which the positions of the sampling points are measured with a Gaussian error. For this we have simulated an ellipsoidal structure whose density follows the profile

$$\rho(x, y) = \frac{M}{\pi(\sigma_x^2 + \sigma_y^2)} \exp\left[-\left[\left(\frac{x - x_0}{\sigma_x}\right)^2 + \left(\frac{y - y_0}{\sigma_y}\right)^2\right]\right]. \quad (8.13)$$

The resulting point distribution is shown in the top left-hand frame of Fig. 8.13. We have assumed that the positions of these sampling points have been measured with a Gaussian error with a standard deviation σ_F which is five times larger than the standard deviation σ_y of the ellipsoidal profile along its minor axis. The resulting observed point distribution is shown in the top right-hand frame of the figure. This distribution is equivalent to another ellipsoidal distribution, but with the σ_x and σ_y replaced by

$$\sigma_x \rightarrow \sigma'_x = \sqrt{\sigma_x^2 + \sigma_F^2} \approx 1.1\sigma_x; \quad (8.14)$$

$$\sigma_y \rightarrow \sigma'_y = \sqrt{\sigma_y^2 + \sigma_F^2} \approx 5.1\sigma_y. \quad (8.15)$$

The observed distribution forms an ellipsoidal object with axis ratio of about 2.2 : 1.

We have checked if the DTFE reconstruction follows this theoretically expected behaviour by measuring at four characteristic locations, denoted by the thick dots in the top right-hand frame of Fig. 8.13, the uncertainty in the DTFE reconstructed density field. For this we have constructed 1000 realizations of possible underlying point distributions, based on the measured locations and the Gaussian distribution of measurement errors, and reconstructed the corresponding density field. For each of the four positions the resulting error distributions are shown in the bottom four frames of the figure. In these frames the one-point probability distribution function of the reconstructed density at that position is shown, normalized by the expected observed density at that position (as given by an ellipsoidal with parameters given by Eqns. 8.14 and 8.15). The smooth curve depicts the theoretical probability distribution function given by Eqn. 8.2. The thick vertical line denotes the observed average reconstructed density, which should equal unity, because we have normalized the observed densities by the expected observed density.

In the figure it can be seen that at all four positions the average reconstructed density is close to unity and that at positions 1, 2 and 3 the observed probability distribution function fits the theoretical distribution very well. At position 4 this is not the case: an overabundance of very low densities is present at this location. This happens because the number of sampling points is too small for the density field to be properly sampled at that position. We conclude that the DTFE reconstruction behaves in accordance with the theoretical expectation.

8.3.2 Measurement errors: the weight of the sampling points

The second type of measurement errors we discuss is an uncertainty in the weight of the sampling points. An example of this is the uncertainty in the mass of a galaxy in deriving the density field corresponding to an observed redshift survey, or the uncertainty in the flux

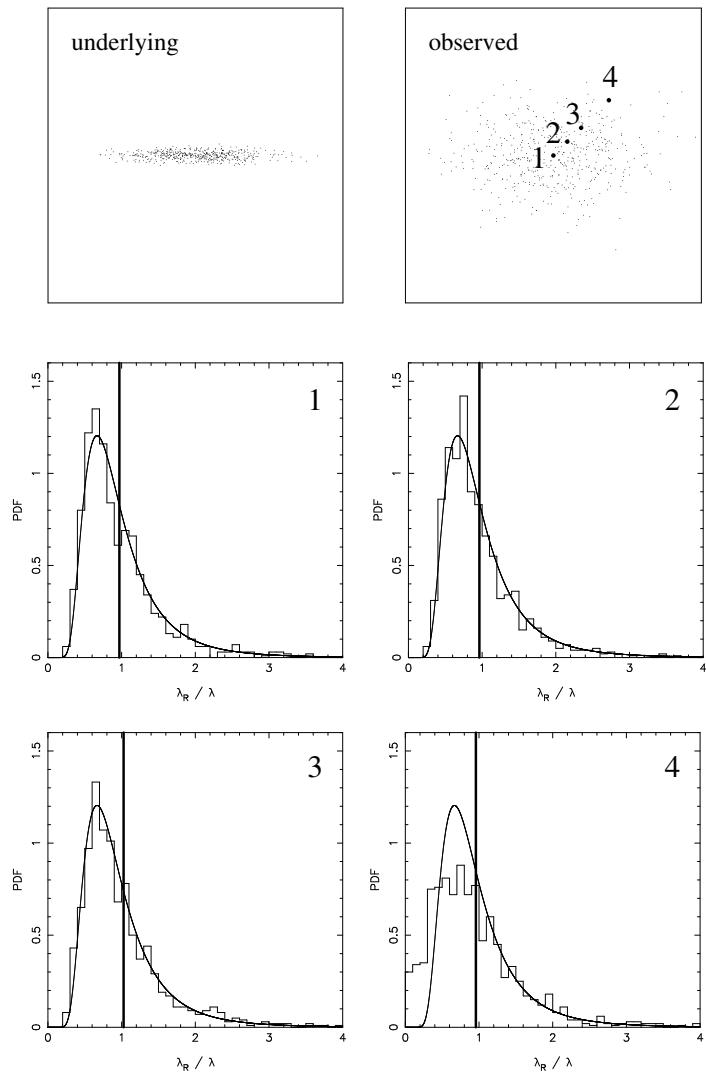


Figure 8.13 — Monte Carlo simulations of the observation of an ellipsoidal structure. In the top row the real underlying point distribution is shown which consists of 500 points (left-hand frame) as well as the observed point distribution (right-hand frame), in which the observed positions have measurement errors. The numbers 1, 2, 3 and 4 denote the locations at which the one-point probability distribution function (PDFs) of the reconstructed density has been determined on the basis of 1000 Monte Carlo simulations. These PDFs are shown in the bottom four frames. In these frames the smooth curve denotes the expected PDF, which corresponds to the ellipsoidal profile smoothed with a Gaussian filter. The thick line denotes the observed average reconstructed density.

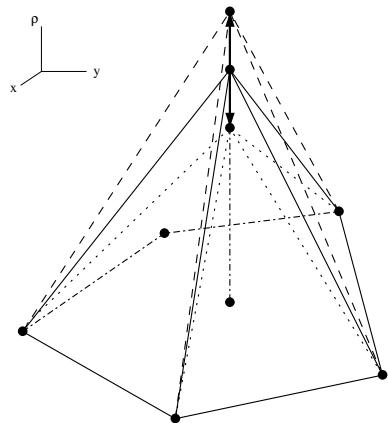


Figure 8.14 — Demonstration that the uncertainty due to the weight of a sampling point is confined to the contiguous Voronoi cell of this point.

of a photon when deriving an density field. Since an uncertainty in the mass of a particle i only leads to an uncertainty in the mass estimate at this location and to an uncertainty in the gradient of the triangles (tetrahedra) constituting its contiguous Voronoi cell, the effects of this uncertainty are confined to this contiguous Voronoi cell, unless other types of errors are present as well. This is illustrated in Fig. 8.14, in which the uncertainty in the effective DTDE kernel corresponding to a particular sampling point is shown in (\mathbf{r}, ρ) -space.

It is straightforward to write down the expression for the uncertainty in the density estimate at the location of this point and also, although a bit more cumbersome, for the uncertainty in the density gradient in its surrounding Delaunay triangles. Consider a sampling point i with mass $m_i \pm \Delta m_i$ and a contiguous Voronoi cell with volume $V(\mathcal{W}_i)$. The uncertainty $\Delta \widehat{\lambda}_i$ in the density estimate $\widehat{\lambda}_i$ at this location due to the uncertainty Δm_i in its mass is then given by

$$\Delta \widehat{\lambda}_i = \frac{(D+1)}{V(\mathcal{W}_i)} \Delta m_i. \quad (8.16)$$

Once the densities $\widehat{\lambda}_i$ at the locations of the sampling points have been estimated, the components of the density gradient in each triangle j may be found by solving the following set of equations:

$$\begin{aligned} \widehat{\lambda}(\mathbf{r}_1) &= \widehat{\lambda}(\mathbf{r}_0) + \left. \widehat{\nabla \lambda} \right|_j \cdot (\mathbf{r}_1 - \mathbf{r}_0); \\ &\vdots \\ \widehat{\lambda}(\mathbf{r}_D) &= \widehat{\lambda}(\mathbf{r}_0) + \left. \widehat{\nabla \lambda} \right|_j \cdot (\mathbf{r}_D - \mathbf{r}_0). \end{aligned} \quad (8.17)$$

Here D is the dimension of space and $\mathbf{r}_0, \dots, \mathbf{r}_D$ are the D vertices of triangle j . The uncertainty in the density gradient in its surrounding triangles due to the uncertainty in the mass of particle i can be found by solving this set of equations and applying standard error analysis. Note that these equations may only be used for an error determination if no other types of errors, such as errors in the position of the sampling points are present. If that is not the case, Monte Carlo simulations can be used to determine the error distribution.

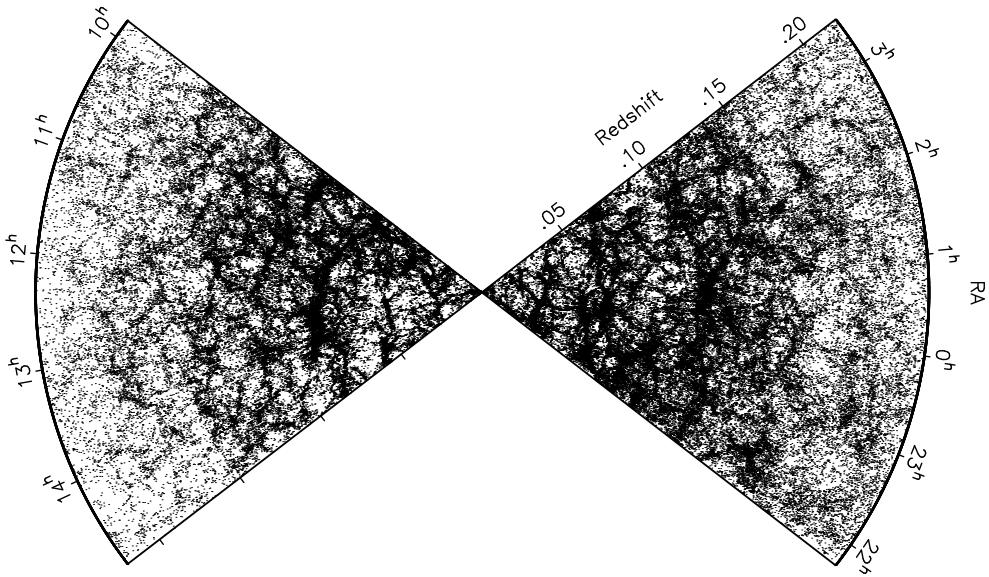


Figure 8.15 — Inhomogeneous sampling in the 2dF galaxy redshift survey. Clearly visible is that at larger redshifts a smaller fraction of galaxies has been observed. Courtesy: the 2dF Galaxy Redshift Survey team.

8.3.3 Systematic errors: inhomogeneous sampling

An important issue for many practical applications occurs when the number density of sampling points is not just a function of the underlying field, but also of some other variable(s). In such a case the sampling is inhomogeneous over the region of space one is looking at. An example of an application in which one is dealing with a selection function are magnitude-limited cosmological redshift surveys, in which one is measuring the positions and redshifts of galaxies in some part of the sky. Because the amount of the light observed from these galaxies decreases with their redshift, a smaller fraction of galaxies is observed at larger redshifts. This is clearly visible in Fig. 8.15, in which the galaxy distribution is shown as has been measured by the 2dF Galaxy Redshift Survey (Colless et al. 2001).

The number density of sampling points $n(\mathbf{x})$ may be written as proportional to the product of the underlying field and a sampling function $\phi(\mathbf{x})$, which is often called the selection function,

$$n(\mathbf{x}) \propto \lambda(\mathbf{x}) \phi(\mathbf{x}). \quad (8.18)$$

Here we have defined the selection function as the fraction of sampling points relative to some normalization. The selection function includes the dependency on all variables other than the underlying field.

In case the selection function is known one may correct for the smaller fraction of observed galaxies by simply multiplying the weight of each sampling point by the inverse of the selection function. Mathematically, the corrected density estimate at the location of sampling

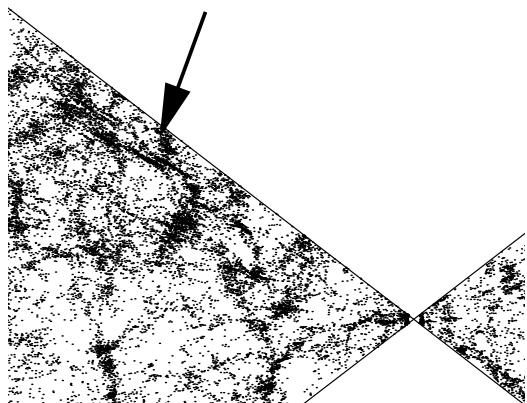


Figure 8.16 — Fingers of God in the 2dF galaxy redshift survey. Courtesy: the 2dF Galaxy Redshift Survey team.

point i is given by

$$\widehat{\lambda}_i = \frac{1}{\phi(\mathbf{x}_i)} \frac{(D+1)m_i}{\mathcal{W}_i}. \quad (8.19)$$

Clearly an uneven sampling will have repercussions on the resolution of the reconstructed field, which have to be taken into account when interpreting or analyzing a reconstructed field. In particular, the effective resolution of features with the same density is different at different locations.

8.3.4 Systematic errors: sampling distortions

A more complicated effect occurs when the sampling of an underlying field is not measured as a function of position, but instead as a function of some other variable(s) $\alpha(\mathbf{x})$. In this case a directly reconstructed density field will yield a distorted view of the proper density field. Such distortions occur for example in cosmological redshift surveys, where the distance to observed galaxies is not measured directly, but via their redshifts. The redshift of a galaxy depends for a large part on the distance through Hubble's law (Hubble 1929), which contains a dependency on the local dynamical environment, the so-called proper motion of a galaxy. The proper motion of a galaxy induces a non-linear relation between distance and redshift, which makes the transformation between the two very hard. For example, large-scale infalling regions appear compressed in redshift-space, the Kaiser effect (Kaiser 1987). Conversely, large underdense regions which are growing are elongated in redshift space. Perhaps the most striking and readily visible distortions are the well-known fingers of God (see Fig. 8.16). Due to the large virial motions inside a galaxy cluster these do not appear as compact regions in redshift space, but instead as very elongated finger-like structures which appear to be pointing toward us. A distorted point sampling should be transformed to real space before a proper DTFE field reconstruction can be done. This is only straightforward when the mapping between the variables α and the proper positions \mathbf{x} is known and invertible. In practical applications this is often not the case and the mapping is strongly dependent on the specific application. In the case of the galaxy redshift surveys the mapping is for example determined by the local mass distribution.

8.4 Summary and discussion

In this chapter we have discussed how the significance of DTFE reconstructed density fields may be determined. For this purpose we have defined the significance as a measure of the probability to find a point in a reconstructed homogeneous Poisson field of a particular intrinsic density λ with a density more extreme than the reconstructed value $\widehat{\lambda}_R$. This complicates the interpretation of derived significance maps, as nearby points may lie within the same triangle and are therefore not independent from each other. This is similar to the problem one encounters in the analysis of many types of astronomical observations, in which the observing instrument has a particular point spread function which affects the statistics of the observed signal. An extra complication is that the sizes and shapes of the Delaunay triangles are varying over the map. One way to avoid the problem of dependent image pixels would be to define the significance not at each location but for each individual triangle. This calculation is in principle not different from the calculation in section 8.2.2. However, the significance maps calculated along these lines would also be complicated to interpret as in this case the size of the independent regions (the triangles) would vary over the map. In particular, a visual comparison would be strongly biased towards low density regions, where the triangles are much larger than in high density regions.

We have discussed two types of systematic sampling effects, inhomogeneous sampling and sampling distortions. Depending on the application other systematic sampling effects may be present as well. An important example of these which occurs in the analysis of cosmological redshift surveys is the (inhomogeneous) Malmquist bias. Since these effects depend on the application, we will not discuss them any further here.

References

- Bernardeau F., van de Weygaert R., 1996, MNRAS, 279, 693
 Colless M. et al., 2001, MNRAS, 328, 1039
 Delone B.N., 1934, Bull. Acad. Sci. USSR: Classe Sci. Mat, 7, 793
 Gaztañaga E., 1989, PhD Thesis, University of Barcelona, Barcelona, Spain
 Gaztañaga E., 1992, ApJ, 398, 17
 Gaztañaga E., 1994, MNRAS, 268, 913
 Hubble E., 1929, Proc NAS., 15, 168
 Járai-Szabó F., Néda Z., 2004, cond-mat/0406116
 Kaiser N., 1987, MNRAS, 277, 1
 Kalman R.E., 1960, Journal of Basic Engineering, 82D, 35
 Kiang T., 1966, Z. Astrophys., 64, 433
 Lahav O., Fisher K.B., Hoffman Y., Scharf C.A., Zaroubi S., 1994, ApJ, 423, 93
 Lemson G., 1995, PhD Thesis, University of Groningen, Groningen, The Netherlands
 Maybeck P.S., 1979, *Stochastic Models, Estimation, and Control*, Vol. 1, Academic Press,
 New York, USA
 Møller J., 1994, in *Lecture Notes in Statistics 87*, Springer-Verlag, New York, USA
 Okabe A., Boots B., Sugihara K., Nok Chiu S., 2000, *Spatial Tessellations, Concepts and
 Applications of Voronoi Diagrams*, 2nd edition, Wiley, Chichester, UK
 Sibson R., 1981, in *Interpreting Multi-Variate Data*, Barnet V. (ed.), Wiley, Chichester, UK
 Tanemura M., 1988, J. Micr., 151, 247
 van de Weygaert R., 1991, PhD Thesis, Leiden University, Leiden, The Netherlands
 Zaroubi S., Hoffman Y., Dekel A., 1999, ApJ, 520, 413

9

Summary and Outlook

9.1 Summary

In this thesis we have described a new tool for reconstructing a volume-covering and continuous density field from a discrete point sampling. We have shown that this tool, the *Delaunay Tessellation Field Estimator* (DTFE), is very well suited for the analysis of highly complex point distributions, such as the large scale galaxy distribution and cosmological N -body simulations.

9.1.1 The Delaunay Tessellation Field Estimator

In Chapters 2 and 3 we have extensively described the DTFE. In essence the DTFE is a stochastic-geometrical algorithm for transforming a given discrete point sampling to its corresponding continuous density field. It involves an extension of the interpolation procedure described by Bernardeau & van de Weygaert (1996), who introduced the Delaunay tessellation (Delone 1934, see also Okabe 2000 and references therein) of a point set as a natural and self-adaptive frame for multi-dimensional interpolation.

The DTFE consists of three main steps, which are illustrated in Fig. 9.1. The starting point is a given discrete point distribution. In the upper left-hand frame of Fig. 9.1 a point distribution is plotted in which at the center of the frame an object is located whose density diminishes radially outwards. In the first step of the DTFE the Delaunay tessellation of the point distribution is constructed. This concerns a volume-covering division of space into triangles (tetrahedra in three dimensions), whose vertices are formed by the point distribution (Fig. 9.1, upper right-hand frame). The Delaunay tessellation is defined such that inside the interior of the circumcircle of each Delaunay triangle no other points from the defining point distribution are present.

The Delaunay tessellation forms the heart of the DTFE. In Fig. 9.1 it is clearly visible that the tessellation automatically adapts to the both the local density and geometry of the point distribution: where the density is high, the triangles are small and vice versa. The size of the triangles is therefore a measure of the local density of the point distribution. This property of the Delaunay tessellation is exploited in step 2 of the DTFE, in which the local density is estimated at the locations of the sampling points. For this purpose the density is defined at the location of each sampling point as the inverse of the area of its surrounding Delaunay

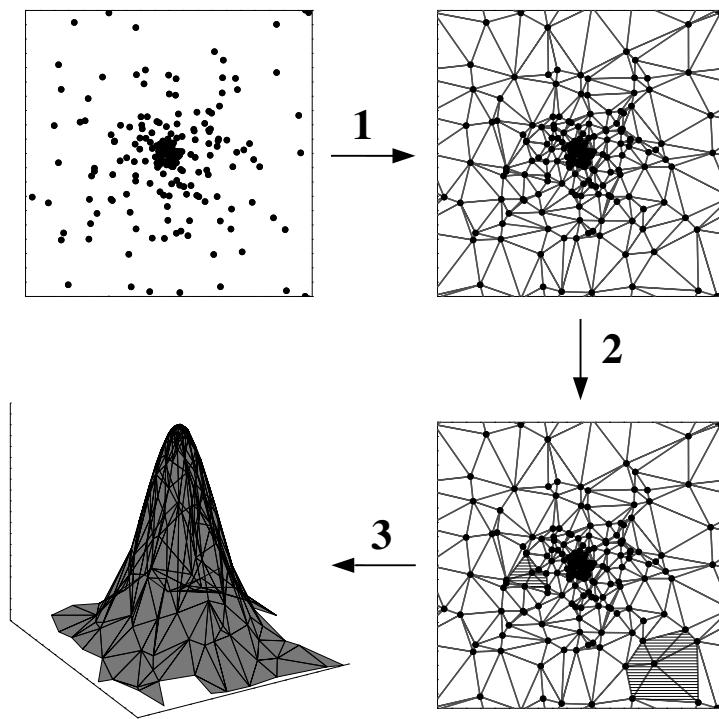


Figure 9.1 — Overview of the DTFE reconstruction procedure. Given a point distribution (top left), one has to construct its corresponding Delaunay tessellation (top right), estimate the density at the position of the sampling points by taking the inverse of the area of their corresponding contiguous Voronoi cells (bottom right) and to assume that the density varies linearly within each Delaunay triangle, resulting in a volume-covering continuous density field (bottom left).

triangles (times a normalization constant, see Fig. 9.1, lower right-hand frame).

In step 3 these density estimates are interpolated to any other point, by assuming that inside each Delaunay triangle the density field varies linearly (Fig. 9.1, lower left-hand frame).

In essence the DTFE is a first-order multi-dimensional interpolation scheme which is closely related to the more generic higher-order natural neighbor procedure (Sibson 1981, see also Watson 1992 and Sukumar 1998). This smooth and local spatial interpolation technique is the most general and robust method of interpolation available to date. However, the implementation of natural neighbour interpolation has been hampered by the fact that an efficient implementation in three dimensions is not trivial. Moreover, it is not clear how a density field reconstruction scheme based on natural neighbour interpolation would conserve mass.

The DTFE allows us to follow the same geometrical and structural adaptive properties of the higher-order natural-neighbour method, while allowing the analysis of truly large data sets and making sure that mass is conserved. The most important property of the DTFE is that it automatically adapts to both the local density and geometry of the point distribution. This results in important advantages with respect to conventional reconstruction procedures. In this thesis we have extensively tested the DTFE and compared its performance with conventional

reconstruction methods.

In Chapter 3 we have shown that of all studied reconstruction procedures the DTFE attains the highest spatial resolution. In Chapters 4 and 5 we have specifically studied the performance of the DTFE with respect to three key aspects of the large scale galaxy distribution:

- hierarchical clustering;
- anisotropic collapse;
- a complex cellular geometry with extended empty regions.

We have shown that with respect to all these aspects the DTFE significantly outperforms existing reconstruction procedures.

The very high spatial resolution of the DTFE makes it also sensitive to sampling noise and/or measurement errors. In Chapter 8 we have described and quantified the statistical properties of the DTFE. We have discussed how the significance of reconstructed density fields can be determined and what the influence is of several types of errors and uncertainties on the reconstructed field.

9.1.2 An atlas of the nearby universe

In this thesis we have described a number of applications for which the use of the DTFE may lead to substantial improvements. A rather straightforward application is the reconstruction of the cosmic density field in the nearby universe. In Chapter 7 we have reconstructed the density field corresponding to the 2dF galaxy redshift survey. The resulting two- and three-dimensional maps reveal an impressive view on the cosmic structures in the nearby universe.

9.1.3 Numerical simulations of structure formation

In Chapter 5 we have described how the DTFE can be incorporated in particle hydrodynamics codes. We have shown that the improved density estimates of the DTFE will yield a major improvement for simulations incorporating feedback processes, which play a major role in galaxy and star formation. In this chapter we have also shown that the DTFE estimate has convenient properties that make the implementation of viscous forces better defined. The presented results form an encouraging step towards the insertion of the DTFE in astrophysical particle hydrodynamics codes.

9.1.4 The cosmic velocity field

The DTFE has been designed for reconstructing density or intensity fields from a discrete set of irregularly distributed points sampling this field. In Chapter 6 we have shown that it can also be used to reconstruct other continuous fields which have been sampled at the locations of these points. The use of the DTFE for this purpose has the same advantages as it has for reconstructing density fields. The fields are reconstructed locally without the application of an artificial or user-dependent smoothing procedure, resulting in an optimal resolution and the suppression of shot-noise effects. The estimated quantities are volume-covering and allow for a direct comparison with theoretical predictions. In this chapter we have focused on the simultaneous reconstruction of the density and velocity fields corresponding to cosmological

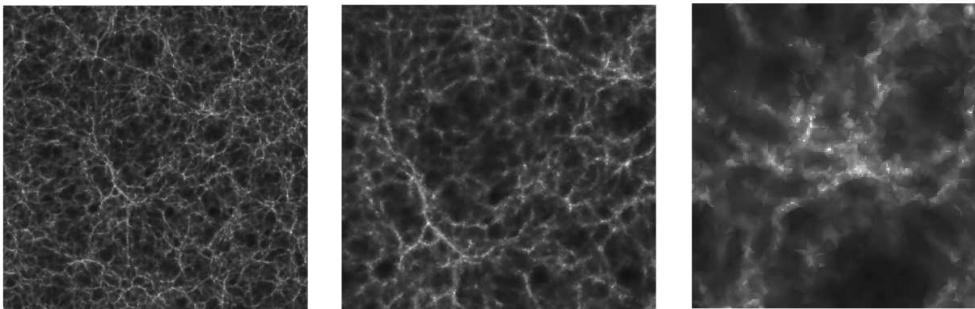


Figure 9.2 — Hierarchical clustering. Depicted is a slice through a cosmological N -body simulation (left-hand frame), together with a number of zoom-ins which focus on a low (central frame) and a high density (right-hand frame) region. On all depicted scales structures are present.

N -body simulations. The resulting fields closely adhere to the continuity equation. The DTFE reconstruction results in realistic density and velocity profiles without having to resolve to artificial smoothing procedures. The results show that the DTFE represents a major step forward for the analysis of cosmic velocity fields at both small and large scales.

9.1.5 Evolution and dynamics of the cosmic web

The DTFE has been specifically designed for describing the complex properties of the cosmic web. In Chapter 6 we have analyzed a number of simulations of cosmic structure formation. By means of a simultaneous DTFE reconstruction of the cosmic density and velocity field we have analyzed the dynamics of characteristic elements of the large scale galaxy distribution. We have described a number of analytic models of voids and shown that the DTFE reconstructions closely adhere to these models. We have also discussed how voids can be used to constrain the value of the cosmological constant. Finally, we have shown that the DTFE reconstructed density and velocity field near superclusters reproduces the theoretically expected infall patterns.

9.2 Virtues and limitations

In this section we shortly discuss the most important virtues and limitations of the DTFE with respect to conventional reconstruction procedures.

9.2.1 Virtues

9.2.1.1 *Virtue: resolving hierarchical substructure*

The first and foremost virtue of the DTFE is its self-adaptive nature. It automatically adapts to the local density of the sampling point distribution without depending on pre-fixed smoothing kernels or on other user-defined procedures. The key to the adaptiveness of the DTFE is the Delaunay tessellation of the sampling point distribution, which forms the heart of the DTFE reconstruction procedure. By definition the Delaunay tessellation is fully adaptive to the sampling point distribution. This property allows the DTFE to simultaneously resolve both high and low density regions. This is illustrated Fig. 9.2, in which an image of the DTFE

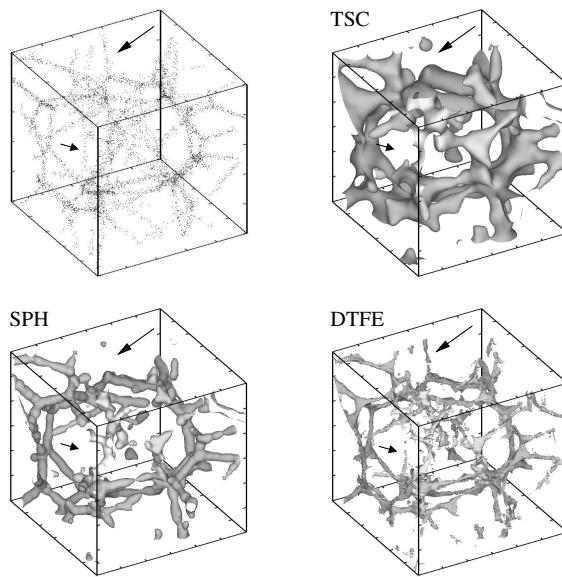


Figure 9.3 — Voronoi model of filamentary structure and the corresponding TSC, SPH and DTFE density field reconstructions. The density contours have been chosen such that 65% of the mass is enclosed. The arrows indicate two structures which are visible in both the galaxy distribution and the DTFE reconstruction, but which the conventional TSC and SPH reconstruction procedures do not resolve.

density field reconstruction of a slice through a cosmological N -body simulation is shown (left-hand frame). Clearly visible is that the slice contains all kinds of structures on different scales and of different density. The subsequent frames zoom in on a low and a high density region. In the central frame, where a low density void is depicted, one may observe that small-scale substructure is present within the void, while it is surrounded by flattened filamentary structures of relatively high densities. In the right-hand frame a high density clump of matter is depicted, which contains even smaller clumps of even higher density. The figure shows that the DTFE is able of simultaneously resolving both the large scale structure as well as the substructures contained within smaller structures.

9.2.1.2 *Virtue: resolving anisotropy*

The next important virtue of the DTFE is that it does not only adapt to the local density of the sampling point distribution, but also to its local geometry. In particular, the DTFE is able of resolving highly anisotropic structures, whereas conventional reconstruction schemes tend to smear such structures over a large volume, effectively making them more spherical. This is illustrated in Fig. 9.3, in which a three-dimensional Voronoi model of a filamentary network is shown, together with the corresponding TSC, SPH and DTFE density field reconstructions. The TSC reconstruction appears very different from the galaxy distribution. Individual structures do not resemble their counterparts in the galaxy distribution, but obtain a significantly larger volume. The resulting filamentary network is substantially more roundish and less

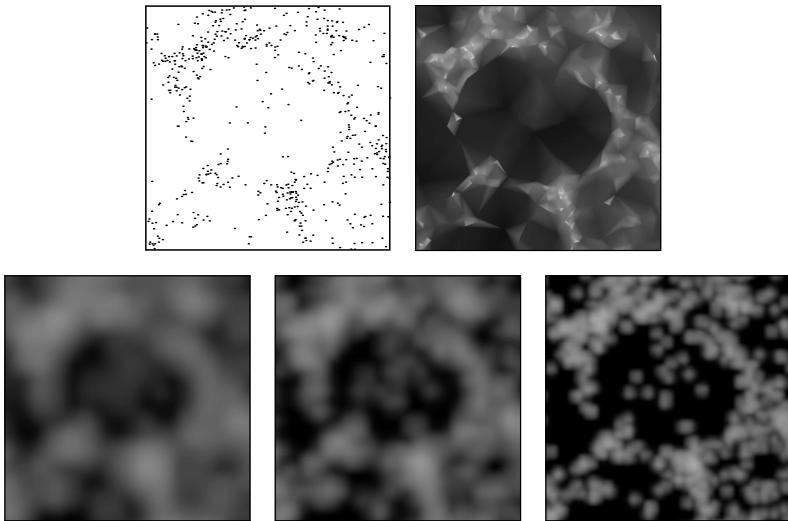


Figure 9.4 — Several density field reconstructions of a void region. Top left: galaxy distribution. Top right: DTFE reconstruction. Bottom row: TSC reconstructions at different resolution.

tenuous than the galaxy distribution. To a lesser extent the same is true for the SPH reconstruction. The DTFE reconstruction appears to resemble the galaxy distribution best. Note for example that at several locations structures are visible in the galaxy distribution which are also present in the DTFE reconstruction, but not in the TSC and SPH reconstructions (two examples are indicated by the arrows in Fig. 9.3).

9.2.1.3 *Virtue: resolving low density regions*

A third virtue of the DTFE is its capability of an accurate description of low density regions without introducing the shot-noise effects by which conventional methods are beset. This is illustrated in Fig. 9.4, in which the DTFE reconstruction of a void-like region is compared with several fixed grid-based TSC reconstructions at different resolutions. First look at the TSC reconstructions. It is clear that none of the adopted resolutions accurately describes the void. At the lowest resolution the overall shape of the void is recovered, but the sharp edges in the galaxy distribution are smeared into featureless blobs. To a lesser extent the same is true for the reconstruction at average resolution. Here, however, one starts to recognize the inner structure of the void breaking up into distinct density blobs, indicating that the resolution of the grid in the interior of the void is too high. This is very clearly visible in the high resolution reconstruction, in which the edges of the void seem to be described much better, but in which the interior of the void is dominated by shot-noise effects. Compare this with the DTFE reconstruction in the top right-hand frame. Here the sharp edges of the void are recovered as such, while at the same time the interior of the void is reconstructed as a gently varying low density region.

Although not shown in Fig. 9.4 SPH-like procedures represent a significant improvement over grid-based methods for low density regions. In Chapter 5 we have shown that the outer rims of filamentary and wall-like structures in SPH reconstructions are smeared into their

neighbouring voids, overestimating the local density. DTFE reconstructed fields are not affected by this problem.

9.2.1.4 *Virtue: reconstructing other dynamical fields*

Apart from reconstructing density or intensity fields the DTFE may also and if desired simultaneously be used to reconstruct other continuous fields. The use of the DTFE for this purpose has the same advantages as for the reconstruction of density fields. This is illustrated in Fig. 9.5, in which a typical void-like region is shown together with the DTFE density and velocity field reconstructions. The thick line running from the bottom to the top of these fields indicates the one-dimensional section along which the density and velocity field are plotted in the bottom right-hand frame of the figure. The DTFE procedure clearly manages to render the void as a realistically slowly varying region of low density. Notice the clear distinction between the empty (dark) interior regions of the void and its edges. The velocity field shows that the void is expanding. This expansion is rather uniform as can be observed in the one-dimensional sections through the density and velocity reconstruction shown in the bottom right-hand frame of Fig. 9.5. The linear ‘super-Hubble’ expansion of voids is well understood in terms of gravitational dynamics. According to Birkhoff’s theorem voids can be approximated as expanding, isolated universes unto themselves that do not accrete matter from the universe at large (e.g. van de Weygaert & van Kampen 1993, Goldberg & Vogeley 2004). Because voids are emptier than the rest of the universe they expand faster than the rest of the universe.

9.2.1.5 *Virtue: dimensional independence*

A final virtue we wish to mention is the fact that the DTFE is independent of the dimension of space. In this thesis we have applied it to both two- and three-dimensional fields. Recently, Arad et al. (2004, 2005) generalized the DTFE for computing the six-dimensional phase-space density $f(\mathbf{x}, \mathbf{v})$ and its PDF in an N -body system. They showed that f is a sensitive tool for studying the evolution of subhaloes during the hierarchical build-up of haloes.

9.2.2 Limitations

9.2.2.1 *Limitation: sensitivity to sampling noise*

The DTFE method has a self-adaptive spatial resolution and an effective smoothing kernel which is more localized than that of other reconstruction methods. It is therefore less forgiving with respect to sampling noise present in the data. Such noise will have a direct impact on the reconstructed field. In principle one may get rid of this noise by filtering the reconstructed density field in the post-processing stage. However, filtering destroys the adaptive capabilities of the DTFE which may be undesirable for many applications. It is therefore essential to understand how reconstructed structures are affected by sampling noise and how one may determine the significance of a reconstructed density field. This issue has been discussed in detail in Chapter 8.

9.2.2.2 *Limitation: boundary conditions*

An important practical issue for the DTFE is that of the boundary conditions. The problem resides in calculating density estimates at the border of the data point sample. If we

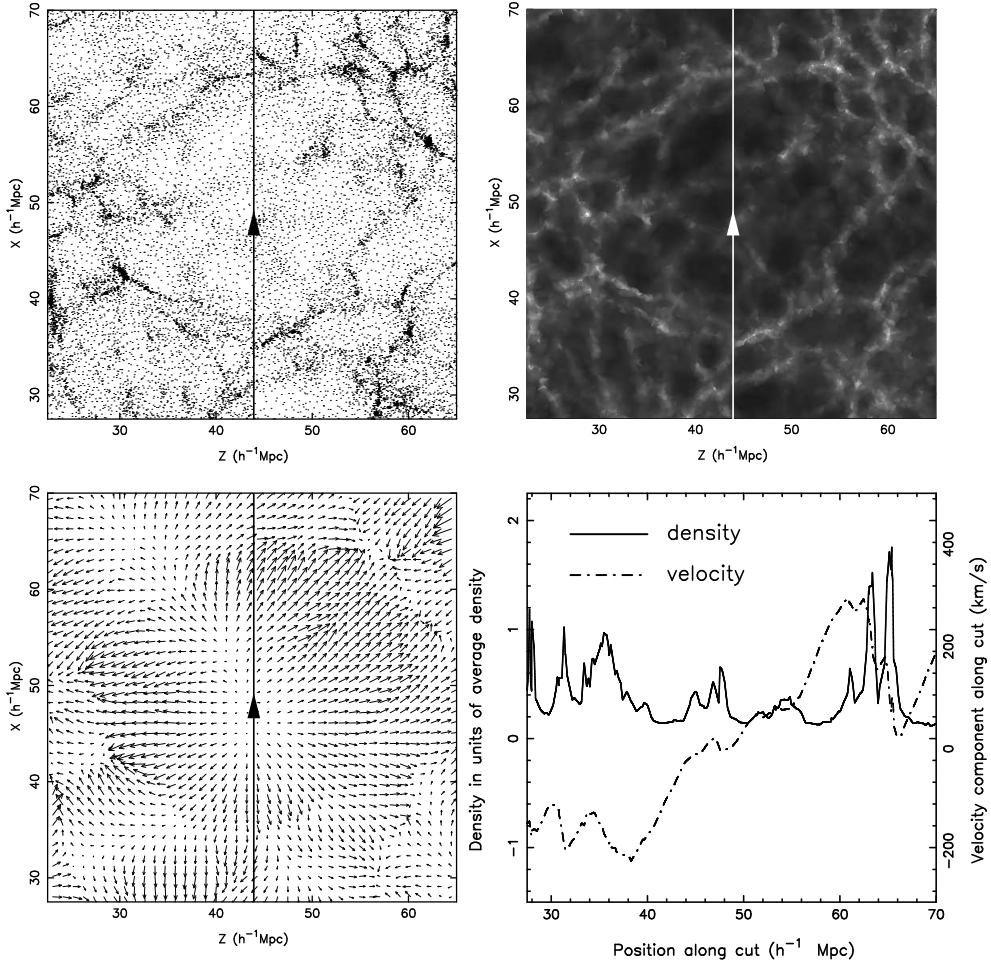


Figure 9.5 — Combined density and velocity field reconstructions. Top left-hand frame: particle distribution in a thin slice through the simulation box. Top right-hand frame: slice through the three-dimensional DTLE density field reconstruction. Bottom left-hand frame: slice through the three-dimensional DTLE velocity field reconstruction. Bottom right-hand frame: density and velocity reconstructions along the one-dimensional section indicated by the solid line shown in the other frames.

confine the Delaunay tessellation only within ‘natural’ edges imposed by the data set, the reconstructed density field will not be correct around the edges. This is because the outermost Delaunay cells stretch out into infinity. The volume of the outermost Delaunay cells is therefore infinitely large and the corresponding estimated densities at the location of the outermost sampling points are equal to zero.

Possible solutions to overcome this edge problem are: boundary padding, imposing periodic boundary conditions or using constrained realizations. Boundary padding refers to adding points beyond the edge of the sampling point distribution. The practical use mainly depends on the knowledge one has on the properties of the density field beyond the sampled region. For instance, for an object placed in a homogeneous background radiation field, the observed field may be extended by adding randomly distributed points.

A special case of boundary padding involves periodic boundary conditions, in which case the sampling point distribution is repeated beyond its edges. The practical use depends on the application. It is well suited for the analysis of cosmological N -body simulations, which usually involve periodic boundary conditions. The use of periodic boundary conditions for the analysis of real data is however rather limited.

A more advanced and consistent case of boundary padding would be to make use of the existing correlations in the field. Constrained field realizations (Bertschinger 1987, Hoffman & Ribak 1991, van de Weygaert & Bertschinger 1996) offer a natural solution for this strategy.

9.2.2.3 Limitation: linear artefacts

A third limitation of the DTFE is its linear nature. The density field is reconstructed as linearly varying with a gradient which is discontinuous at the edges of the Delaunay tetrahedra. At the scale of the local smoothing kernel, which is as large as the local contiguous Voronoi cell, this results in triangular artefacts, which form the imprint of the linear interpolation procedure. Physical density fields vary smoothly and are continuously differentiable. In principle it is possible to generalize the first-order DTFE reconstruction procedure to higher orders such as natural neighbor interpolation.

9.2.2.4 Limitation: empty regions

The final limitation of the DTFE we address here is that it is not capable of reconstructing regions of zero density. The reason for this is that any region with a finite size will correspond to Delaunay tetrahedra with a finite size and therefore a non-zero density.

9.3 Conclusions and outlook

In this thesis we have developed a new method for analyzing the large scale galaxy distribution. We have shown that this new method, the DTFE, performs significantly better than existing techniques. The applications we have described show that it can contribute to a variety of subjects in present-day cosmological research.

Several other astronomers have recognized the advantages of the DTFE and used it in their research. Bradac et al. (2004) used DTFE reconstructed surface density maps to compute the gravitational lensing pattern around galaxies, upon which Li et al. (2006) evaluated the method in its ability to trace higher-order singularities. Shandarin et al. (2004) advocate the DTFE for systematic studies of the size, shape and topology of the cosmic web by means

of Minkowski functionals. Neyrinck et al. (2005) have used the DTFE to identify haloes in cosmological N -body simulations. Arad, Dekel & Klypin (2004) used the DTFE to assess the six-dimensional phase-space density distribution of dark haloes in cosmological N -body simulations (see also Arad & Johansson 2005).

In Groningen the cosmology group has applied the DTFE to several aspects of the large scale galaxy distribution. Romano-Díaz (2004) has constructed velocity maps of the nearby universe using the DTFE. Using the DTFE he has also characterized the thermal state of the Local Group and Local Supercluster by measuring volume-weighted cosmic Mach number statistics. He was able to show that the coldness of the local flow is due to the specific mass configuration beyond the Local Group. Besides studies concerning the cosmic velocity field the DTFE has also been used as the basis of two advanced algorithms for detecting characteristic structures in galaxy redshift surveys and cosmological N -body simulations. One of these techniques, the *Multiscale Morphology Filter* (Aragón-Calvo et al. 2006), is able of identifying clusters, filaments and walls, and is presently used for a systematic study of the properties of these structures. The other algorithm, the *Cosmic Watershed Algorithm*, is able of identifying voids (Platen & van de Weygaert 2006). Sophisticated applications like these, tuned towards uncovering the characteristics of the reconstructed large scale galaxy density field, yield the real potential of the DTFE.

We conclude that the DTFE forms a major step forwards for studies of the large scale galaxy distribution. This field of research is still advancing rapidly from forthcoming new data and bigger and more advanced numerical simulations. With better data and new theoretical and computational tools, such as the DTFE, cosmologists have the prospect of answering some of the fundamental open questions related to our cosmic world view.

References

- Arad I., Dekel A., Klypin A., 2004, MNRAS, 353, 15
- Arad I., Johansson P.H., 2005, MNRAS, 362, 252
- Aragón-Calvo M., van de Weygaert R., Jones B.J.T., van der Hulst J.M., 2006, astro-ph/0610249
- Bernardeau F., van de Weygaert R., 1996, MNRAS, 279, 693
- Bertschinger E., 1987, ApJ, 323, 103
- Bradac M. et al., 2004, A&A, 423, 13
- Delone B.N., 1934, Bull. Acad. Sci. USSR: Classe Sci. Mat., 7, 793
- Goldberg D.M., Vogeley M.S., 2004, ApJ, 605, 1
- Hoffman Y., Ribak E., 1991, ApJ, 380, 5
- Li G.-L. et al., 2006, astroph/0603557
- Neyrinck M.C., Hamilton A.J.S., Gnedin N.Y., 2005, MNRAS, 362, 337
- Okabe A., Boots B., Sugihara K., Nok Chiu S., 2000, *Spatial Tessellations, Concepts and Applications of Voronoi Diagrams*, 2nd edition, Wiley, Chichester, UK
- Platen E., van de Weygaert R., 2006, *private communication*
- Shandarin S.F., Sheth J.V., Sahni V., 2004, MNRAS, 353, 162
- Sibson R., 1981, in *Interpreting Multi-Variate Data*, Barnett V. (ed.), Wiley, Chichester, UK
- Sukumar N., 1998, PhD Thesis, Northwestern University, Evanston, IL, USA
- Watson D.F., 1992, *Contouring: a Guide to the Analysis and Display of Spatial Data*, Pergamon Press, Oxford, UK
- van de Weygaert R., Bertschinger E., 1996, MNRAS, 281, 84
- van de Weygaert R., van Kampen E., 1993, MNRAS, 263, 481

Nederlandse samenvatting

Sinds de vroegste beschavingen zijn er voortdurend mensen geweest die 's nachts in verwondering omhoog hebben gekeken naar de sterrenhemel. Zij vroegen zich onder andere af hoe ons heelal eruit ziet en hoe het is ontstaan. Door de eeuwen heen zijn er allerlei verschillende, vaak religieus geïnspireerde kosmogonieën ontwikkeld om zo onze wereld in een kosmisch perspectief te kunnen plaatsen. Sinds een paar honderd jaar is er ook vooruitgang geboekt in wetenschappelijke zin, deels met behulp van waarnemingen en deels door theoretische modelvorming. Pas in de vorige eeuw is er een coherent model ontwikkeld dat daadwerkelijk gefalsificeerd kan worden door middel van waarnemingen. In dit model, de *oerknaltheorie*, is er een beginpunt, ongeveer 13,7 miljard jaar geleden, waarin ruimte, tijd en alle energie in het heelal hun oorsprong vinden. Wat er gebeurde op het precieze moment van de oerknal en in de eerste 10^{-43} seconde daarna (de *Plancktijd*) begrijpen we op dit moment nog niet goed. Wel weten we dat het heelal na afloop van de Plancktijd een extreem compact, dicht en heet plasma van straling en materie was. Vanaf de Plancktijd wordt de ontwikkeling van het heelal zeer nauwkeurig beschreven door de oerknaltheorie.

Hoewel de oerknaltheorie een groot aantal waargenomen verschijnselen in het heelal met succes heeft voorspeld en verklaard, is er nog wel een aantal onopgeloste fundamentele vraagstukken. De eerste betreft de zogenoamde *donkere materie*. Waarnemingen laten zien dat slechts 15% van de materie in het heelal uit 'normale' baryonische materie bestaat, waar ook de aarde, zon en planeten uit bestaan. Dit betekent dat de overige 85% van de materie in het heelal van een andere, niet-baryonische aard moet zijn. We weten dat deze materie bestaat door de werking van de zwaartekracht, maar we weten niet waaruit deze bestaat. We kunnen deze materie ook niet zien en daarom wordt vaak over donkere materie gesproken.

Een volgend raadsel vormt de mysterieuze *donkere energie* die overal in het heelal aanwezig is en die maar liefst 73% van de totale hoeveelheid energie in het heelal voor zijn rekening neemt. Hoewel er genoeg speculaties zijn over de aard van deze donkere energie is deze feitelijk een mysterie. Dit blijkt alleen al uit het feit dat de theoretisch voorspelde waarde 10^{118} – een 1 met 118 nullen – keer groter is dan de gemeten waarde. Merkwaardigerwijs wijzen waarnemingen uit dat er in het heelal precies zoveel donkere energie en materie aanwezig is dat de geometrie van het heelal vlak is. We begrijpen niet waarom dit het geval is.

Op dit moment bevat het heelal een ware rijkdom aan structuren van allerlei groottes. Op voor sterrenkundigen interessante schalen zijn dit bijvoorbeeld de aarde, de zon, de sterren, de Melkweg en andere sterrenstelsels. Het bestaan van structuur in het heelal kan niet verklaard worden binnen de oerknaltheorie, omdat daarin aangenomen wordt dat het heelal perfect homogeen en isotroop is. Er is daarom een uitbreiding van de oerknaltheorie nodig, waarin

aangenomen wordt dat er in het zeer jonge heelal hele kleine dichtheidsfluctuaties aanwezig waren: op sommige plaatsen in het heelal was net iets meer materie aanwezig dan op andere plaatsen. In de loop van de tijd zijn deze dichtheidsfluctuaties door de werking van de zwaartekracht uitgegroeid tot de structuren die we vandaag de dag waarnemen. Deze uitbreiding van de oerknaltheorie die de vorming en ontwikkeling van structuur in het heelal beschrijft, wordt *gravitatiele instabiliteitstheorie* genoemd.

Hoewel op basis van gravitatiele instabiliteitstheorie grote computersimulaties van kosmische structuurvorming zijn uitgevoerd die de algemene aanblik van het heelal op grote schaal reproduceren, wordt de ontwikkeling van structuur in het heelal nog niet goed begrepen. Dit komt doordat een groot aantal natuurkundige processen een rol speelt, die elk over heel verschillende tijd- en afstandschaalen werken. Hoewel er allerlei slimme trucs en technieken zijn bedacht om met deze inherente complexiteit om te kunnen gaan, zijn de tegenwoordige analyse- en simulatietechnieken niet in staat om alle relevante natuurkundige processen nauwkeurig te volgen en te beschrijven.

In dit proefschrift wordt een nieuwe methode beschreven waarmee complexe puntverdelingen – zoals de verdeling van sterrenstelsels op grote schaal – kunnen worden bestudeerd. Met deze nieuwe methode willen we een bijdrage leveren aan de oplossing van een aantal van de vele nog openstaande problemen binnen de hedendaagse kosmologie. De vragen die in dit proefschrift behandeld worden, hebben voornamelijk betrekking op de verdeling van sterrenstelsels op grote schaal, maar leveren daarmee tegelijkertijd ook een bijdrage aan de beantwoording van een aantal van de meer fundamentele vragen over ons kosmologische wereldbeeld.

De oerknaltheorie

Tot aan het begin van de twintigste eeuw werd algemeen aangenomen dat het heelal statisch is, dat wil zeggen eeuwigdurend en onveranderlijk. In 1915 formuleerde Albert Einstein de algemene relativiteitstheorie. Hij ontdekte dat volgens deze theorie het heelal niet statisch maar dynamisch is: het heelal dijt uit of krimpt. In 1922 slaagde Alexander Friedmann erin om de vergelijkingen die volgens de algemene relativiteitstheorie een homogeen en isotroop heelal beschrijven op te lossen. Onafhankelijk hiervan werden deze vergelijkingen in 1927 ook door de Belgische priester Georges Lemaître opgelost. Lemaître realiseerde zich ook wat de natuurkundige implicaties van de door hem gevonden oplossingen waren. Terugrekenend in de tijd zag hij in dat een uitdijend heelal een extreem heet en dicht beginpunt in de tijd gehad moet hebben, dat hij het *oeratoom* noemde. In de hieropvolgende jaren is er door voor- en tegenstanders stevig gedebatteerd over de vraag of het heelal echt een beginpunt in de tijd heeft gehad. Het heeft tot in de jaren zestig van de vorige eeuw geduurd voordat er zoveel bewijzen gevonden waren voor de oerknaltheorie dat deze vanaf toen het referentiekader vormt waarbinnen sterrenkundigen werken.

Op dit moment wordt de oerknaltheorie bekrachtigd door een grote hoeveelheid waarnemingen. Vier daarvan vormen de belangrijkste pilaren onder de oerknaltheorie. Wellicht de meest directe is het feit dat de oerknaltheorie een verklaring geeft voor een verrassend voor de hand liggende waarneming van Heinrich Olbers in de vroege 18^e eeuw: 's nacht is de hemel donker. Dit simpele feit kan alleen worden verklaard door aan te nemen dat het heelal een eindige leeftijd heeft in combinatie met een eindige lichtsnelheid. Het tweede belangrijke be-

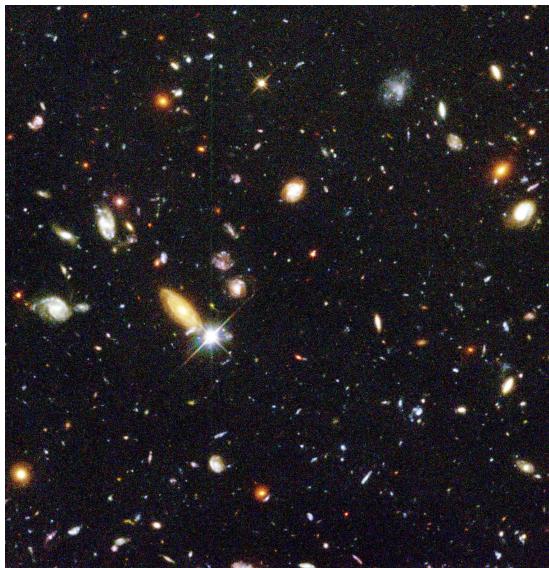
wijssstuk is de ontdekking in 1929 door Edwin Hubble dat sterrenstelsels van ons af bewegen met een snelheid die groter wordt naarmate ze verder van ons af staan, de *wet van Hubble*. Hiermee werd het eerste overtuigende bewijs geleverd dat ons heelal niet statisch is, maar uitdijt. In de afgelopen jaren hebben satellietwaarnemingen de uitdijingssnelheid van het heelal nauwkeurig bepaald op 71 kilometer per seconde per Megaparsec[†]. Dit betekent bijvoorbeeld dat een object dat op een afstand van 10 Megaparsec van ons af staat elke seconde 710 kilometer verder van ons af beweegt. De andere twee belangrijke bewijsstukken grijpen terug naar de tijd waarin ons heelal nog maar kort bestond. Tijdens de eerste drie minuten van het heelal was het heelal zo heet (enkele miljarden graden) dat de elementen Deuterium, ³Helium, ⁴Helium en Lithium door kernfusie gevormd konden worden. Omdat de oerklanttheorie nauwkeurig beschrijft hoe de afkoeling van het heelal verloopt, kan heel precies uitgerekend worden hoeveel van deze elementen er uiteindelijk gevormd worden. De voorspelde hoeveelheden blijken heel nauwkeurig in overeenstemming met de waargenomen hoeveelheden. Het laatste bewijs dateert van het moment waarop het heelal ongeveer 400 000 jaar oud was. Op dat moment was de temperatuur van het heelal afgekoeld tot ongeveer 3000 graden en konden de tot dan toe vrij in het heelal rondzwevende protonen en elektronen samen waterstofatomen vormen. Straling die tot dan toe voortdurend door de vrij rondzwevende elektronen werd afgebogen, kon nu ongehinderd door het heelal blijven bewegen. Deze *kosmische achtergrondstraling* werd in 1965 voor het eerst waargenomen door Arno Penzias en Robert Wilson, waarmee het definitieve bewijs voor de oerklanttheorie geleverd was.

De structuur van het heelal op grote schaal

De oerklanttheorie beschrijft de ontwikkeling van het heelal als geheel, waarbij aangegeven wordt dat het heelal isotroop en homogeen is (het *kosmologische principe*). Dit is echter alleen een adequate beschrijving als we kijken op een schaal groter dan enkele honderden miljoen lichtjaren. Op kleinere schaal bevat het heelal een ware rijkdom aan structuur in allerlei soorten en maten. Dit is al duidelijk door 's nachts naar de hemel te kijken. Duidelijk zichtbaar (tenminste buiten de grote stad) is een heldere band van sterren die dwars over de hemel loopt. De sterren in deze band horen allemaal bij de Melkweg, een verzameling van ongeveer 200 miljard sterren waartoe ook de zon behoort. Vrijwel alle sterren in het heelal bevinden zich in soortgelijke sterrenstelsels, groepen van enkele honderden miljarden sterren die door hun onderlinge zwaartekracht bijeen gehouden worden. In Fig. 1 is het *Hubble Deep Field* afgebeeld. Dit is een foto die door de Hubble ruimtetelescoop gemaakt is van een erg klein gedeelte van de hemel, ongeveer zo groot als een stuiver op een afstand van 25 meter. Doordat de belichtingstijd van deze foto maar liefst 10 dagen lang was, konden ook heel lichtzwakke stelsels, die erg ver van ons af staan, waargenomen worden. In totaal zijn aan dit kleine stukje van de hemel zeker 1500 sterrenstelsels te zien. Dit betekent dat het gehele zichtbare heelal ongeveer 100 miljard sterrenstelsels bevat, die elk bestaan uit enkele tientallen tot honderden miljarden sterren.

Sterrenstelsels zijn geen geïsoleerde objecten, maar vormen de bouwstenen van structuren op nog grotere schaal. Zo is de Lokale Groep een systeem van enkele tientallen sterrenstelsels, waarvan de Melkweg en Andromeda de grootste zijn. Deze twee spiraalvormige stelsels

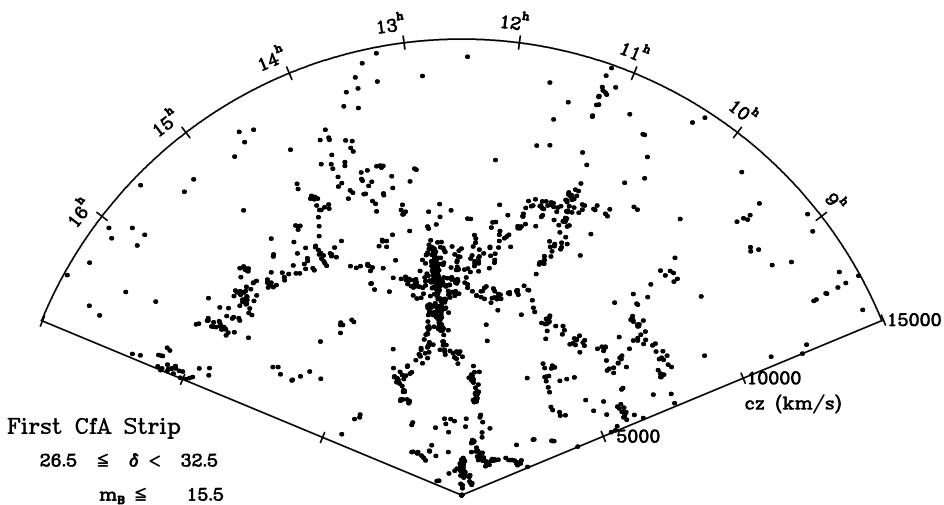
[†]Een Megaparsec is een sterrenkundige afstandsmaat en komt overeen met de afstand die het licht in 3,26 miljoen jaar aflegt, ongeveer 30 800 000 000 000 000 km.



Figuur 1 — Het Hubble Deep Field. De afbeelding is gemaakt door de Hubble ruimtetelescoop gedurende 10 dagen een foto te laten maken van een heel klein gebiedje aan de hemel. Daardoor zijn ook zeer lichtzwakke en ver weg staande sterrenstelsels waargenomen. Duidelijk te zien is dat het heelal zeer veel sterrenstelsels bevat. De grote sterrenstelsels staan relatief dichtbij, terwijl de kleinst zichtbare puntjes op een zeer grote afstand van ons staan.

worden door de zwaartekracht naar elkaar toegetrokken, terwijl ze beiden worden omringd door een zwerm van kleinere dwergstelsels en onregelmatige stelsels. Soortgelijke *groepen van sterrenstelsels* komen op veel plaatsen voor in het heelal. In sommige gevallen zijn er echter veel meer sterrenstelsels opeengepakt in hele dichte en massieve *clusters van sterrenstelsels*. De meest massieve van deze clusters bevatten duizenden sterrenstelsels binnen een relatief klein gebied met een doorsnede van enkele miljoenen lichtjaren. In de dichtbijstaande Virgo en Coma clusters liggen bijvoorbeeld meer dan duizend sterrenstelsels binnen een afstand van maar 5 miljoen lichtjaar van hun kern. Ter vergelijking, de afstand van de Melkweg tot Andromeda is al zo'n 2,5 miljoen lichtjaar. Door hun enorme helderheid zijn clusters makkelijk waarneembaar tot op grote afstanden in het heelal.

Een jaar of twintig geleden zijn astronomen begonnen om het heelal op nog grotere schaal in kaart te brengen. Ze hebben daarvoor de afstand tot een groot aantal sterrenstelsels bepaald. Helaas is het niet mogelijk om de afstand direct te meten, maar wel de snelheid waarmee een sterrenstelsel van ons af beweegt. De wet van Hubble beschrijft het verband tussen deze snelheid en de afstand tot het sterrenstelsel en kan dus gebruikt worden om de gemeten snelheid om te zetten in de afstand. In de sterrenkunde wordt de snelheid waarmee een stelsel van ons afbeweegt vaak uitgedrukt in *roodverschuiving*. Dit komt omdat licht uit golfjes bestaat. De lengte van deze golfjes wordt groter naarmate een voorwerp sneller van ons af beweegt. Het licht krijgt hierdoor een rodere kleur. Sterrenstelsels zien er dus roder uit naarmate ze sneller van ons af bewegen (en dus verder van ons afstaan). De eerste kaart van ons nabije heelal was de CfA roodverschuivingskaart uit 1986 (zie Fig. 2). Hierin is de roodverschuiving van



Figuur 2 — De CfA roodverschuivingskaart. De gegevens van ruim 1000 sterrenstelsels waren al voldoende om in te zien dat hun verdeling in het heelal alles behalve willekeurig is. Er is een duidelijke band van sterrenstelsels te zien die van links naar rechts door de figuur loopt. Deze supercluster wordt de Grote Muur genoemd. De opeenhoping van stelsels in het midden van de figuur is de Coma cluster.

alle waargenomen sterrenstelsels in een dunne strip aan de hemel bepaald. Elke stip in deze figuur correspondeert met een sterrenstelsel. De Melkweg bevindt zich helemaal onderaan de figuur. Elke rechte lijn die door de Melkweg gaat, correspondeert met één bepaalde positie aan de hemel. Sterrenstelsels die op zo'n lijn door de Melkweg liggen, zien we dus op dezelfde positie aan de hemel, maar ze liggen wel op een verschillende afstand van ons. In de figuur is duidelijk zichtbaar dat sterrenstelsels niet willekeurig over de ruimte verdeeld zijn, maar grote en langgerekte structuren vormen, die *superclusters* worden genoemd. Superclusters zijn meestal opgebouwd uit enkele rijke clusters en een groot aantal kleinere groepen van sterrenstelsels. Een indrukwekkend voorbeeld is de *Grote Muur*, de uitgestrekte band van sterrenstelsels die door het midden van Fig. 2 van links naar rechts loopt. Deze enorme, afgeplatte verzameling van sterrenstelsels heeft een afmeting van ongeveer 200 bij 230 bij 15 miljoen lichtjaar. De Coma cluster is midden in de Grote Muur zichtbaar. In de figuur is ook goed te zien dat er grote gebieden zijn waarbinnen zich vrijwel geen sterrenstelsels bevinden. Zulke gebieden worden *leegten* genoemd. Dit zijn uitgestrekte, ruwweg bolvormige gebieden met afmetingen in de orde van 75 tot 150 miljoen lichtjaar. Het bekendste voorbeeld is de Boötes leegte met een diameter van maar liefst 200 miljoen lichtjaar.

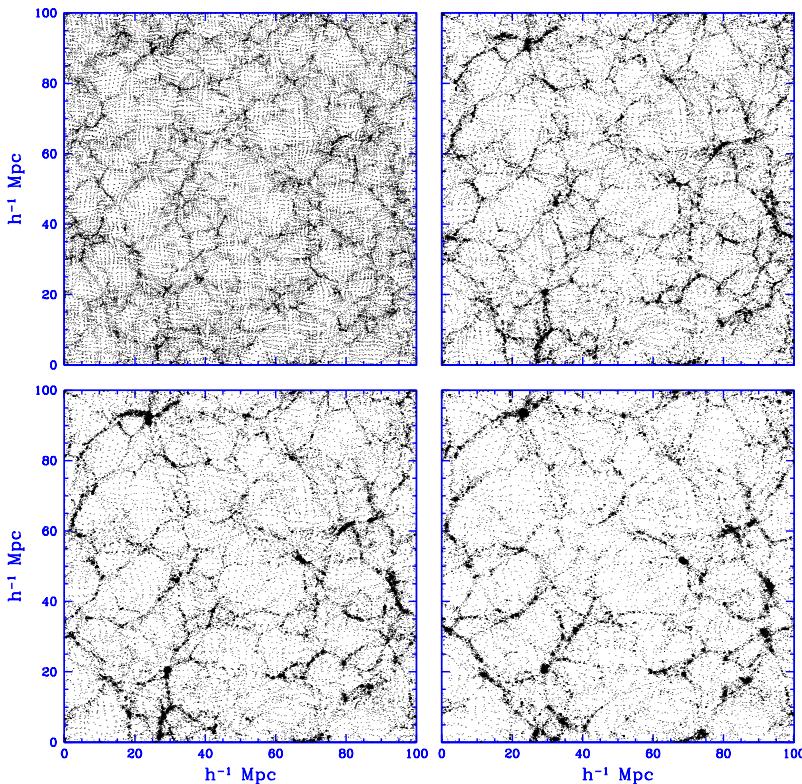
De laatste jaren hebben steeds grotere en diepere roodverschuivingskaarten ons beeld van het heelal op grote schaal verfijnd. Eén van de meest uitgebreide kaarten is de 2dF roodverschuivingskaart waarin de afstand tot ruim 220 000 sterrenstelsels is bepaald in twee dunne strips aan de hemel. Dergelijke kaarten laten zien dat superclusters een intrigerend netwerk vormen dat zich uitstrekkt over het hele waarneembare heelal. In dit netwerk bevinden zich uitgestrekte leegten tussen wand- en filamentvormige superclusters. Clusters van sterrenstelsels bevinden zich op plekken waar superclusters elkaar kruisen. Het resulterende netwerk wordt vanwege zijn vorm wel het *kosmische web* of het *kosmische schuim* genoemd.

De vorming en ontwikkeling van kosmische structuren

De oerknaltheorie beschrijft de ontwikkeling van het heelal als geheel, waarbij aangeomen wordt dat het heelal isotroop en homogeen is. Hoewel op voldoende grote schaal het heelal inderdaad als isotroop en homogeen beschouwd kan worden, bevat het heelal op kleinere schaal allerlei structuren. Dit roept de vraag op hoe deze structuren zijn gevormd in een heelal dat, volgens het kosmologische principe, perfect isotroop en homogeen is. Om dit te begrijpen is een uitbreiding van de oerknaltheorie nodig, waarin aangenomen wordt dat op vroege tijdstippen het heelal niet perfect homogeen was, maar dat er overal kleine dichtheidsverschillen waren. Dit wil zeggen dat sommige gebiedjes in het heelal meer materie bevatten dan andere. De kleine dichtheidsfluctuaties vinden hun oorsprong in quantumfluctuaties toen het heelal nog zeer jong en dus zeer compact was. We denken dat kort na de oerknal het heelal een fase-overgang heeft doorgemaakt waarin sprake was van een exponentieel snelle uitdijing. Deze periode van extreem snelle uitdijing, die *inflatie* wordt genoemd, heeft de quantumfluctuaties opgeblazen tot macroscopische proporties. De resulterende dichtheidsfluctuaties zijn onder invloed van de zwaartekracht uitgegroeid tot de rijkdom van structuren die we vandaag de dag waarnemen. De theorie die de vorming en ontwikkeling van deze structuren beschrijft, wordt *gravitationele instabiliteitstheorie* genoemd.

In Fig. 3 wordt het proces van structuurvorming onder invloed van gravitationele instabiliteit geïllustreerd. Hierin is op een aantal verschillende tijdstippen een computersimulatie van een gebied met een afmeting van 300 bij 300 miljoen lichtjaar afgebeeld. Het paneel linksboven laat de structuur van het heelal op een vroeg tijdstip zien, terwijl de daaropvolgende panelen met een later tijdstip corresponderen. Duidelijk zichtbaar is dat het vroege heelal relatief homogeen was, terwijl op latere tijdstippen de structuren steeds meer in het oog springen. Wat er gebeurt is dat gebieden die wat meer materie dan hun omgeving bevatten een wat grotere zwaartekracht op hun omgeving uitoefenen dan andersom het geval is. Hierdoor stroomt er materie naar deze gebieden toe, waardoor de dichtheid verder toeneemt, zodat de op de omgeving uitgeoefende zwaartekracht nog groter wordt. Dit proces duurt voort totdat deze gebieden zoveel materie bevatten dat ze onder invloed van hun eigen zwaartekracht ineenstorten tot een gravitationeel gebonden object. Evenzo oefenen gebieden met minder materie dan hun omgeving een kleinere zwaartekracht uit op hun omgeving dan andersom het geval is. Het resultaat is dat er materie vanuit deze gebieden wegstromt naar hun omgeving, waardoor de dichtheid verder afneemt, zodat de op hun omgeving uitgeoefende zwaartekracht nog kleiner wordt. Het resultaat is dat deze gebieden geleidelijk steeds leger en leger worden om uiteindelijk de leegten te vormen die we in roodverschuivingskaarten aantreffen.

Gravitationele structuurvorming is een uitermate complex proces. De eerste fase verloopt nog relatief eenvoudig. De dichtheidsverstoringen zijn dan nog relatief klein en de resulterende materieverdeling is in essentie een versterkte versie van de oorspronkelijke verdeling. Dit *lineaire regime* kan daarom goed analytisch beschreven worden. Op het moment dat de dichtheidsverstoringen groter beginnen te worden, ontkoppelt de expansie van materieconcentraties van de globale kosmische uitdijing en beginnen deze samen te trekken. Het daaropvolgende proces van ineenstorting onder invloed van de eigen zwaartekracht is dermate complex dat er computersimulaties nodig zijn om het goed te kunnen bestuderen. Uit deze simulaties blijkt dat de ineenstorting verloopt volgens een reeks van karakteristieke *anisotropie* (afgeplatte) patronen. In het begin nemen de ineenstortende gebieden een vlakke, wandvormige



Figuur 3 — Computersimulatie van gravitationele structuurvorming. Afgebeeld is een gebied met een afmeting van ruim 300 bij 300 miljoen lichtjaar op vier verschillende tijdstippen. De tijd loopt van linksboven naar rechts onder.

structuur aan, gevolgd door een samentrekking tot een uitgerekte filamentaire vorm, voordat uiteindelijk de volledige ineenstorting tot een gravitationele gebonden object plaatsvindt. De vorming van anisotrope structuren is goed zichtbaar in Fig. 3.

De fysische eigenschappen van de dichtheidsfluctuaties in het jonge heelal zijn verantwoordelijk voor een tweede belangrijk aspect van gravitationele structuurvorming, namelijk de *hiërarchische* aard ervan. Omdat de fluctuaties op kleine schalen sterker zijn dan die op grote schalen, zullen de vroegst gevormde objecten klein zijn. Grottere objecten vormen doordat kleinere, eerder gevormde objecten, samenklonteren. Dit proces van *hiërarchische structuurvorming* lijkt inderdaad overeen te komen met hetgeen we waarnemen: sterrenstelsels zijn veel ouder dan de massievere en recentelijk ineengestortte clusters van sterrenstelsels. Op een nog grotere schaal hebben superclusters nog niet eens de ineenstortingsfase bereikt of zijn ze net begonnen met samentrekken. Doordat structuurvorming hiërarchisch verloopt, zijn grote structuren dan ook opgebouwd uit kleinere. Zo bestaan superclusters uit een reeks van kleinere groepen van sterrenstelsels, die elk een hogere concentratie van materie bevatten dan de supercluster als geheel. Als we inzoomen op kleinere schalen, dan zien we dat sterrenstel-

sels zelf vaak omringd worden door een aantal kleinere sattellietstelsels en dwergstelsels. Een voorbeeld hiervan is onze eigen Melkweg, die omringd wordt door twee relatief grote onregelmatige stelsels, de Magelhaense wolken, naast een groot aantal dwergstelsels. Ook clusters bevatten allerlei substructuur. Een voorbeeld is de Coma cluster, waarin verschillende dominante stelsels aanwezig zijn, die elk geïdentificeerd kunnen worden met een aparte kern van de cluster. Zelfs in de leegten kan substructuur worden waargenomen: kleinere leegten liggen vaak ingebet in uitgestrektere leegten.

Het resultaat van het gravitationele structuurvormingsproces is de al eerder beschreven ordening van clusters, superclusters en leegtes in het kosmische web. Uit computersimulaties blijkt dat deze cellulaire geometrie een universele uitkomst is van gravitationele structuurvorming.

Bestaande gereedschappen en hun tekortkomingen

De oerklanttheorie en gravitationele instabiliteitstheorie vormen het natuurkundige referentiekader waarbinnen we de vorming en ontwikkeling van structuren in ons heelal kunnen begrijpen. We hebben gezien dat een analytische beschrijving van de ontwikkeling van kosmische structuren uitsluitend mogelijk is in de lineaire fase[†]. Dit komt door de complexiteit van het kosmische structuurvormingsproces, dat gekarakteriseerd wordt door:

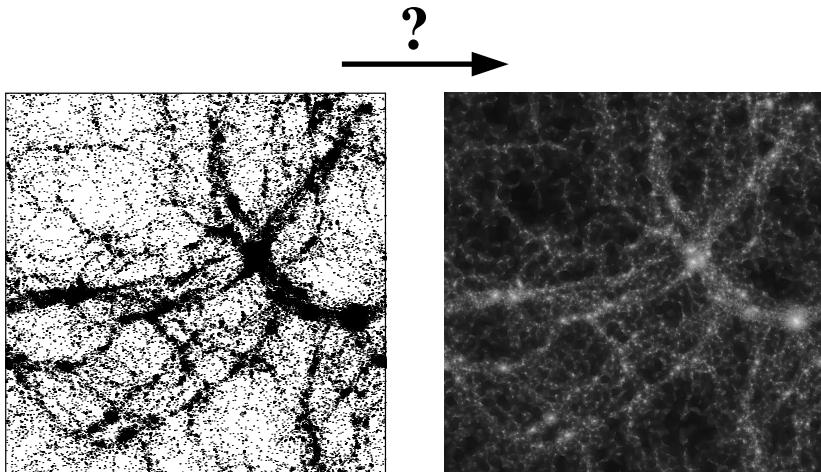
- hiërarchische structuurvorming;
- gravitationele ineinstorting volgens anisotrope patronen;
- een cellulaire geometrie met uitgestrekte lege gebieden.

Het zal duidelijk zijn dat zo'n hoge mate van complexiteit een kwantitatieve analyse enorm bemoeilijkt. Kosmische structuurvorming wordt daarom in de praktijk vaak bestudeerd door het vergelijken van experimentele waarnemingen, zoals roodverschuivingskaarten of computersimulaties, met voorspellingen van theoretische modellen.

Op dit punt lopen we tegen een fundamenteel probleem bij de analyse van roodverschuivingskaarten en computersimulaties aan: de transformatie van een discrete verzameling van posities van sterrenstelsels dan wel simulatiedeeltjes naar het corresponderende continue dichtheidsveld (zie Fig. 4). Het dichtheidsveld beschrijft op ieder punt van de ruimte de intensiteit van de materieverdeling. In dit proefschrift argumenteren we dat conventionele methoden om het dichtheidsveld te reconstrueren niet in staat zijn om de volledige complexiteit van de materieverdeling op grote schaal goed te beschrijven. De meeste bestaande technieken zijn speciaal ontworpen om één of hooguit enkele aspecten van de verdeling van sterrenstelsels te beschrijven, maar zijn tegelijkertijd volledig ongeschikt om andere aspecten te beschrijven.

Conventionele methoden om het continue dichtheidsveld te bepalen vanuit een discrete puntverdeling vallen uiteen in twee categorieën. De conceptueel meest eenvoudige zijn niet-adaptief en maken gebruik van een vast rooster of een vast gebied waarover de massa van een deeltje wordt uitgesmeerd. In Fig. 5 staat de werking van een op een rooster gebaseerde reconstructiemethode afgebeeld. In deze figuur zijn enkele van de nadelen van niet-adaptieve

[†]In de praktijk is het mogelijk om een analytische beschrijving uit te breiden naar de vroege semi-lineaire fasen en enkele bijzondere, relatief eenvoudige configuraties, maar dit levert dit geen goede basis voor een goed en volledig begrip van het kosmische structuurvormingsproces.

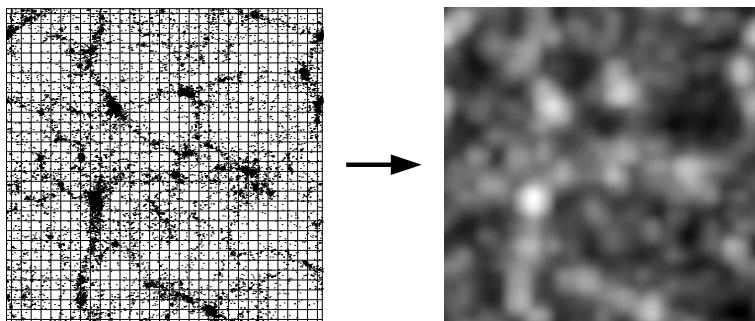


Figuur 4 — Het centrale probleem in dit proefschrift: de reconstructie van een continu dichtheidsveld (rechterpaneel) vanuit een discrete puntverdeling (linkerpaneel). Het voorbeeld dat in deze figuur staat afgebeeld betreft de uitkomst van een computersimulatie van kosmologische structuurvorming. Het figuur laat ook een ander probleem bij de analyse van sterrenkundige waarnemingen zien: de deeltjesverdeling wordt gekarakteriseerd door structuren die elk heel verschillende afmetingen, vormen en dichthesen hebben. Het dichtheidsveld dat in het rechterpaneel staat afgebeeld, is gereconstrueerd met de Delaunay Tessellation Field Estimator, de reconstructiemethode die in dit proefschrift beschreven wordt.

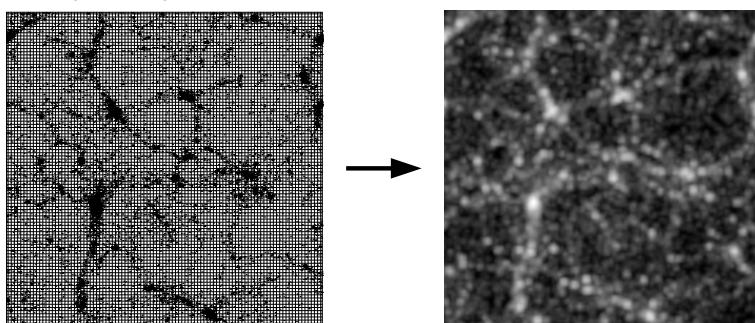
methoden duidelijk zichtbaar. De belangrijkste is dat de keuze van de resolutie (de grootte van de vakjes) in principe willekeurig is, terwijl het resulterende dichtheidsveld sterk afhankelijk is van de gekozen waarde. In de figuur is goed te zien dat bij lage resoluties (grote vakjes) gebieden met een hoge dichtheid slecht opgelost worden en dat afgeplatte structuren niet goed reconstrueerd worden. Gebieden met een lage dichtheid lijken wel goed beschreven te worden. Bij hoge resoluties (kleine vakjes) lijken gebieden met hoge dichtheid beter opgelost te worden, maar gebieden met lage dichtheid juist niet. In deze gebieden is er nu sprake van ruis. Vakjes waarin zich toevallig een deeltje bevindt, worden met een relatief hoge dichtheid gereconstrueerd, terwijl de dichtheid in de naastliggende vakjes gelijk is aan nul. Uitgebreide structuren worden niet meer als één geheel gereconstrueerd, maar opgebroken in substructuren. De vaste geometrie van het rooster zorgt ook voor andere problemen, zoals het feit dat de vorm van afgeplatte structuren in het gereconstrueerde dichtheidsveld sterk afhangt van de oriëntatie ten opzichte van het rooster.

Om aan deze bezwaren tegemoet te komen, zijn er verschillende adaptieve methoden voorgesteld. Deze maken bijvoorbeeld gebruik van adaptieve roosters waarvan de resolutie automatisch wordt aangepast aan de lokale dichtheid van de puntverdeling. Hoewel adaptieve technieken zeker een verbetering vormen ten opzichte van conventionele niet-adaptieve methoden, maken zij gebruik van allerlei subjectief in te stellen parameters, zoals de relatie tussen de lokale dichtheid en de grootte en vorm van het gebied waarover de massa van een deeltje wordt uitgesmeerd. Daarnaast is de vaste geometrie van het rooster meestal niet geschikt om sterk afgeplatte structuren te beschrijven.

32 bij 32 vakjes:



100 bij 100 vakjes:



Figuur 5 — Een conventionele niet-adaptieve op een rooster gebaseerde reconstructiemethode. Over de deeltjesverdeling wordt een vast rooster gelegd, waarbij de dichtheid in een vakje bepaald wordt door de hoeveelheid deeltjes in dit vakje. Duidelijk zichtbaar is hoe sterk het eindresultaat afhangt van de gekozen resolutie van het rooster.

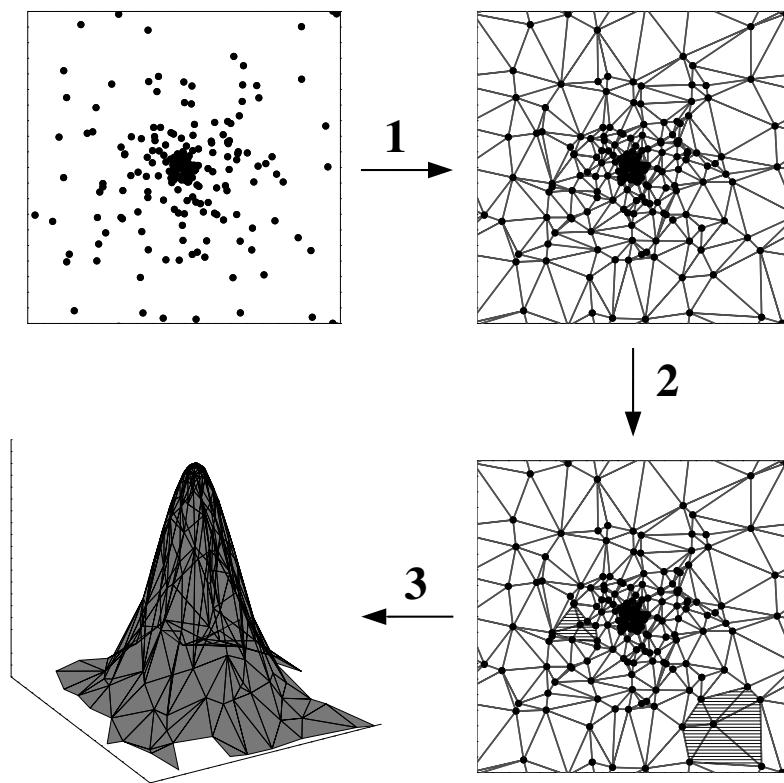
Dit proefschrift

In dit proefschrift beschrijven we een nieuw gereedschap om een dichtheidsveld te reconstrueren uit een discrete puntverdeling. Dit gereedschap, de *Delaunay Tessellation Field Estimator* (DTFE), is uitstekend geschikt voor de analyse van zeer complexe puntverdelingen zoals de verdeling van sterrenstelsels op grote schaal of computersimulaties van structuurvorming in het heelal.

The Delaunay Tessellation Field Estimator

In Hoofdstukken 2 en 3 hebben we de DTFE uitvoerig beschreven. In feite is het een wiskundig algoritme om een gegeven discrete puntverdeling om te zetten naar het corresponderende continue dichtheidsveld. Dit algoritme bestaat uit drie stappen, die afgebeeld staan in Fig. 6.

Het uitgangspunt van de DTFE is een gegeven discrete puntverdeling. In Fig. 6 is linksboven een puntverdeling afgebeeld waarin zich in het midden van de figuur een object bevindt, waarvan de dichtheid naar buiten toe afneemt. In de eerste stap van de DTFE wordt de Delaunay tessellatie van de puntverdeling geconstrueerd. Dit is een verdeling van de ruimte in



Figuur 6 — De Delaunay Tessellation Field Estimator (DTFE). Gegeven een puntverdeling (linksboven) wordt de Delaunay tessellatie geconstrueerd (rechtsboven). Vervolgens wordt op de plaats van de punten de dichtheid afgeschat door het omgekeerde te nemen van de oppervlakte van de omringende Delaunay driehoeken (rechtsonder). Tenslotte wordt de dichtheid op iedere willekeurige plaats gedefinieerd door aan te nemen dat de deze binnen een driehoekje gelijkmataig varieert (linksonder).

driehoeken (in drie dimensies zijn dit tetrahedra), waarbij de hoekpunten van de driehoeken gevormd worden door de oorspronkelijke puntverdeling (Fig. 6, rechtsboven).

De Delaunay tessellatie vormt het hart van de DTFE. In de figuur is duidelijk te zien dat deze zich automatisch aanpast aan de dichtheid en de geometrie van de puntverdeling: als de dichtheid hoog is en zich veel punten in een relatief klein gebied bevinden, dan zijn de driehoeken klein. Omgekeerd is het zo dat als de dichtheid laag is en zich weinig punten bevinden in een relatief groot gebied, dan zijn de driehoeken groot. De grootte van de driehoeken vormt dus een maat voor de lokale dichtheid van de puntverdeling.

Deze eigenschap van de Delaunay tessellatie wordt uitgebuit in stap 2, waarin de lokale dichtheid van de puntverdeling wordt afgeschat op de plaats van de punten. Hiertoe wordt voor elk punt de oppervlakte bepaald van de omringende driehoeken (Fig. 6, rechtsonder). Vervolgens wordt de dichtheid in dit punt gedefinieerd als het omgekeerde van deze oppervlakte (maal een normaliseringssconstante). Het resultaat van deze stap is een verzameling dichtheidswaarden op de plaats van de punten.

In stap 3 worden deze dichtheidsschattingen vervolgens geïnterpoleerd naar elke willekeurige plaats, door aan te nemen dat in elke driehoek de dichtheid gelijkmatig varieert. In Fig. 6 is rechtsonder het resulterende dichtheidsveld afgebeeld, waarbij de hoogte de gereconstrueerde dichtheid aangeeft. Inderdaad is het zo dat midden in de figuur de hoogste dichtheden gevonden worden, en dat deze naar buiten toe kleiner worden.

De belangrijkste eigenschap van de DTFE is dat deze zich automatisch aanpast aan zowel de lokale dichtheid als de lokale geometrie van de puntverdeling. Dit levert belangrijke voordelen op ten opzichte van bestaande reconstructiemethoden. In Hoofdstuk 3 hebben we aangetoond dat van alle bestudeerde reconstructiemethoden de DTFE de hoogste resolutie heeft. In Hoofdstuk 4 hebben we specifiek naar de eigenschappen van de DTFE met betrekking tot sterk hiërarchische en afgeplatte puntverdelingen gekeken. We hebben laten zien dat de DTFE in beide gevallen significant beter presteert dan bestaande reconstructiemethoden. In Hoofdstuk 5 hebben we de DTFE reconstructie van een computersimulatie van het kosmische web vergeleken met de reconstructie van een vaak gebruikte adaptieve methode. We hebben laten zien dat DTFE vooral beter presteert op de grensgebieden van de leegten en filament- en wandvormige structuren, waar bestaande methoden de dichtheid overschatten, alsook in de binnengebieden van filament- en wandvormige structuren, waar bestaande methoden de dichtheid juist onderschatten.

In Hoofdstuk 8 van dit proefschrift hebben we de statistische eigenschappen van de DTFE beschreven. We hebben laten zien hoe de significantie van gereconstrueerde dichtheidsvelden bepaald kan worden en wat de invloed van verschillende soorten fouten en onzekerheden is op het uiteindelijk verkregen dichtheidsveld.

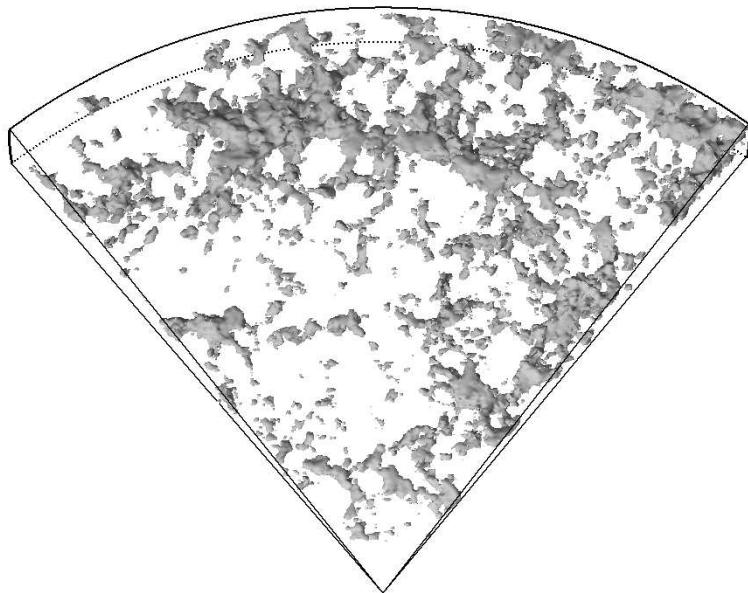
Een atlas van het nabije heelal

In dit proefschrift hebben we ook een aantal toepassingen beschreven waar de DTFE een bijdrage kan leveren. Een voor de hand liggende toepassing is de reconstructie van het kosmische dichtheidsveld in het nabije heelal. In Hoofdstuk 7 van dit proefschrift hebben we het dichtheidsveld gereconstrueerd dat correspondeert met de sterrenstelsels in de 2dF roodverschuivingskaart. De resulterende twee- en drie-dimensionale kaarten geven een indrukwekkend beeld van de kosmische structuren in het nabije heelal.

In Fig. 7 is één van deze kaarten afgebeeld. Net als in de CfA roodverschuivingskaart (zie Fig. 2) bevindt de Melkweg zich helemaal onderaan de figuur en correspondeert elke rechte lijn die door de Melkweg gaat met één bepaalde positie aan de hemel. Duidelijk zichtbaar is een gigantische supercluster die bovenaan van links naar rechts door de figuur loopt. Deze supercluster, de *Sloan Grote Muur*, genoemd naar de Sloan roodverschuivingskaart waarin deze supercluster voor het eerst is waargenomen, is de allergrootste en meest massieve supercluster die op dit moment bekend is.

Computersimulaties van structuurvorming in het heelal

Eerder noemden we al dat het vanwege de complexiteit van de relevante natuurkundige processen nodig is om computersimulaties te gebruiken bij de bestudering van structuurvorming in het heelal. In zulke simulatieprogramma's worden de posities en snelheden van miljoenen simulatiedeeltjes gevolgd. Gegeven deze posities en snelheden op een bepaald tijdstip berekent de computer voor ieder deeltje de zwaartekracht ten gevolge van alle andere deeltjes



Figuur 7 — Drie-dimensionale kaart van de verdeling van sterrenstelsels in een dunne strip aan de noordelijke hemel.

in de simulatie. Daarna berekent de computer de posities en snelheden die de deeltjes een klein tijdje later zullen aannemen. Hiermee wordt de situatie een tijdje na het begin verkregen. Vervolgens berekent de computer opnieuw voor ieder deeltje de zwaartekracht en het effect daarvan op zijn positie en snelheid. Zo wordt de situatie op weer een klein tijdje later verkregen. Door dit proces voortdurend te herhalen, kan de ontwikkeling van kosmische structuurvorming heel nauwkeurig gevolgd worden. Het hart van een simulatieprogramma wordt gevormd door het algoritme om de zwaartekracht uit te rekenen. In Hoofdstuk 5 hebben we beschreven hoe de DTFE in zo'n simulatieprogramma geïmplementeerd zou kunnen worden. We hebben aangegeven dat dit tot sterk verbeterde dichtheidsbepalingen zal leiden. Dit is van groot belang omdat veel natuurkundige processen, zoals de hoeveelheid sterren die gevormd worden, van de precieze waarde van de dichtheid afhangen. Onze resultaten tonen aan dat simulatieprogramma's sterk verbeterd kunnen worden door de DTFE te gebruiken.

Het kosmische snelheidsveld

In Hoofdstuk 6 hebben we beschreven dat de DTFE niet alleen gebruikt kan worden om de dichtheid van een puntverdeling te reconstrueren, maar ook om andere continue velden te reconstrueren waarvan de waarde op een verzameling van discrete punten bekend is. De belangrijkste toepassing die we hebben beschreven betreft het kosmische snelheidsveld, dat samen met het dichtheidsveld de dynamica van het kosmische web bepaalt.

Waargenomen en gesimuleerde snelheidsvelden kunnen alleen vergelijkt worden met theoretische modellen onder de aanname dat de waargenomen velden een goede beschrijving vormen van het onderliggende snelheidsveld. In werkelijkheid is dit vaak niet het geval.

Waargenomen snelheden zijn gewoonlijk alleen bekend op lokaties waar zich – toevallig – sterrenstelsels of simulatieelijtjes bevinden. Theoretische voorspellingen zijn juist gebaseerd op de veronderstelling dat het snelheidsveld overall goed gedefinieerd is. Het is daarom niet altijd goed mogelijk om waargenomen snelheidsvelden te vergelijken met theoretische voorspellingen. Daarnaast is het zo dat de bezwaren die we hebben beschreven voor bestaande methoden om het dichtheidsveld te reconstrueren ook gelden voor methoden om het snelheidsveld te reconstrueren.

In Hoofdstuk 6 hebben we aangetoond dat de DTFE zowel op grote als op kleine schaal de karakteristieke elementen van het snelheidsveld goed beschrijft. De resulterende velden hebben de vereiste eigenschappen, namelijk dat ze op iedere plek goed gedefinieerd zijn. Met onze implementatie kunnen bovendien gerelateerde grootheden, zoals de uitdijing en samentrekking van het snelheidsveld, eenvoudig berekend. Vergelijkingen van DTFE snelheidsvelden met theoretische voorspellingen komen zeer goed met elkaar overeen. In Fig. 8 is een voorbeeld te zien van de DTFE reconstructie van het dichtheids- en het snelheidsveld van een aantal karakteristieke elementen van het kosmische web. Onze resultaten laten zien dat deze techniek een belangrijke stap voorwaarts is bij de analyse van kosmische snelheidsvelden op zowel grote als kleine schaal.

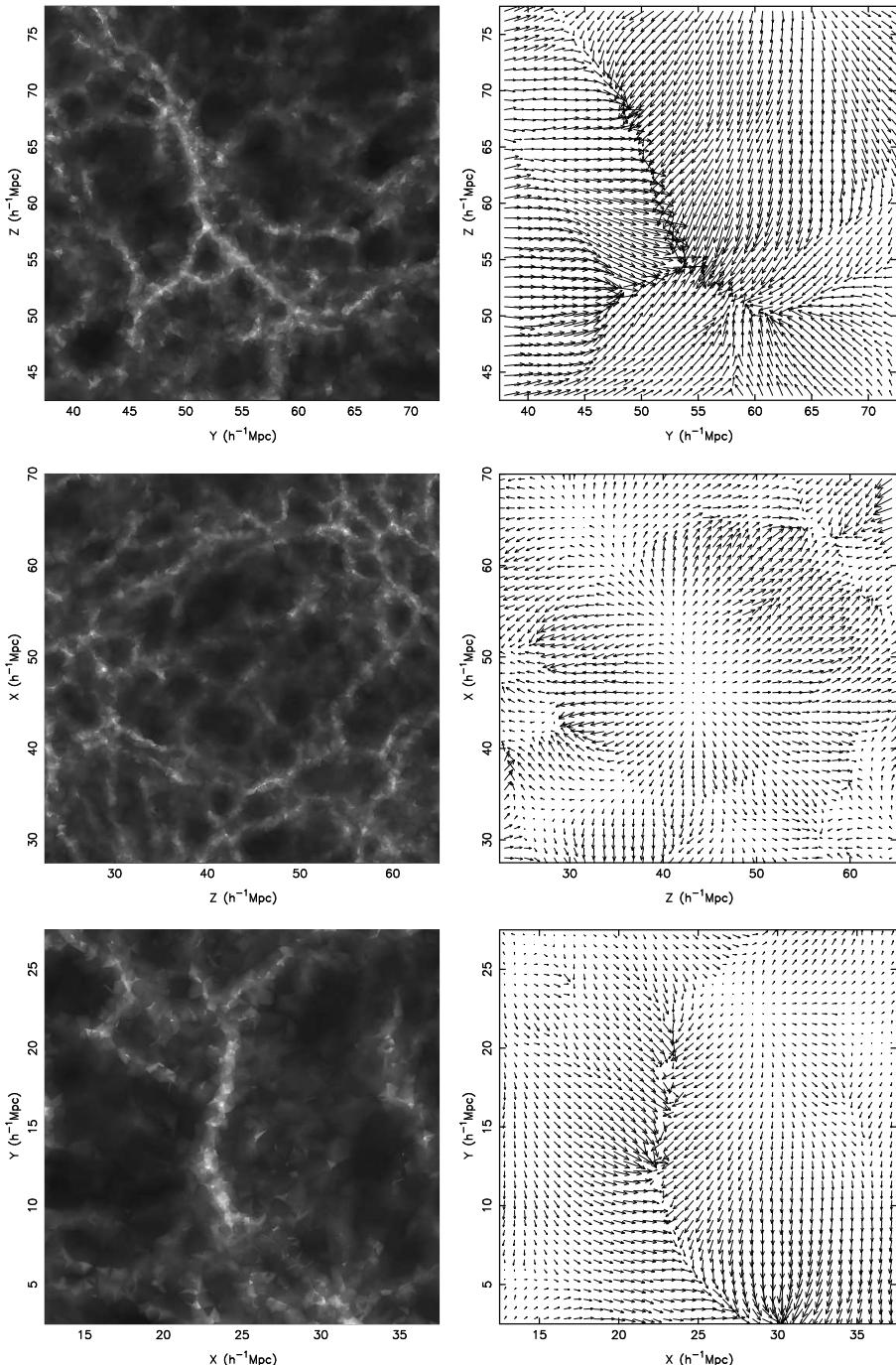
Evolutie en dynamica van het kosmische web

De DTFE is ontwikkeld om de complexe eigenschappen van het kosmische web goed te kunnen beschrijven. Dit betekent dat de DTFE een uitermate geschikt instrument is om de verdeling van sterrenstelsels op grote schaal te bestuderen. In Hoofdstuk 6 hebben we een aantal computersimulaties van kosmische structuurvorming geanalyseerd. Omdat met behulp van de DTFE zowel de snelheidsvelden als de dichtheid gereconstrueerd kunnen worden, hebben we in het bijzonder naar de dynamica van karakteristieke elementen van de verdeling van sterrenstelsels op grote schaal gekeken. We hebben een aantal analytische modellen voor leegtes beschreven en laten zien dat de DTFE reconstructie hier goed mee overeenkomt. In het middelste paneel van Fig. 8 is duidelijk te zien dat het snelheidsveld in de afgebeelde leegte naar buiten gericht is, hetgeen betekent dat deze leegte expandeert. Ook hebben we in Hoofdstuk 6 laten zien dat leegtes gebruikt kunnen worden om de waarde van de kosmologische constante te bepalen. Daarnaast hebben we gekeken naar het gedrag van de dichtheid en snelheid nabij superclusters en laten zien dat de DTFE het theoretisch verwachte invalpatroon goed beschrijft. Dit is goed te zien in het bovenste en het onderste paneel van Fig. 8.

Conclusies en vooruitblik

In dit proefschrift hebben we een nieuwe methode ontwikkeld om de verdeling van sterrenstelsels op grote schaal te bestuderen. We hebben aangetoond dat deze nieuwe methode, de Delaunay Tessellation Field Estimator (DTFE), duidelijk beter presteert dan bestaande technieken. De toepassingen die we hebben beschreven laten zien dat de DTFE over een breed gebied een bijdrage aan het moderne kosmologisch onderzoek kan leveren.

Ook andere onderzoekers hebben de DTFE immiddels gebruikt voor verschillende doelen. Zo hebben onderzoekers aan de universiteit van Bonn in Duitsland de DTFE gebruikt om de effecten van donkere materie op gravitatielenzen te meten. Onderzoekers aan de universiteit van Kansas in de Verenigde Staten en Pune in India hebben aangetoond dat de DTFE



Figuur 8 — Structuur en snelheidsveld van een aantal karakteristieke elementen van het kosmische web. Van boven naar beneden: een deel van het kosmische web met in het midden een cluster waar drie superclusters bijeenkommen, een lege en een supercluster.

uitermate geschikt is om de topologie van de verdeling van sterrenstelsels op grote schaal te bepalen. Onderzoekers van de universiteit in Colorado in de Verenigde Staten hebben de DTFE als basis gebruikt voor een methode om gravitationeel gebonden objecten in grote computersimulaties te detecteren. Tenslotte hebben onderzoekers aan de universiteit van Jezuzalem te Israël de DTFE gebruikt om de dynamica te onderzoeken van objecten die onder invloed van hun eigen zwaartekracht ineenstorten.

Ook de onderzoeksgrondslag kosmologie van de universiteit in Groningen heeft de DTFE gebruikt voor het bestuderen van verschillende aspecten van de verdeling van sterrenstelsels op grote schaal. Zo is met behulp van de DTFE een snelheidskaart van het nabije heelal gemaakt. Daarnaast is de DTFE gebruikt als basis voor enkele geavanceerde methoden om de verschillende elementen van het kosmische web in computersimulaties en roodverschuivingskaarten te detecteren. Eén van deze methoden, het *Multiscale Morphology Filter*, kan clusters, filamentaire en wandvormige structuren identificeren. Op basis hiervan kan een systematische studie naar de eigenschappen van deze structuren plaatsvinden. De andere methode, het *Cosmic Watershed Algorithm*, kan juist leegten detecteren.

Het is duidelijk dat de in dit proefschrift beschreven Delaunay Tessellation Field Estimator een belangrijke stap voorwaarts vormt bij de bestudering van de verdeling van sterrenstelsels op grote schaal. Het onderzoek naar grote schaal structuur boekt nog steeds snelle vooruitgang door het beschikbaar komen van nieuwe data en steeds grotere en geavanceerdere computersimulaties. Dit zal voorlopig zo blijven. Met de betere data en nieuwe theoretische en computergereedschappen zoals de DTFE hebben kosmologen het vooruitzicht om de komende jaren enkele van de fundamentele nog openstaande vragen over ons kosmologisch wereldbeeld te beantwoorden.

List of Publications

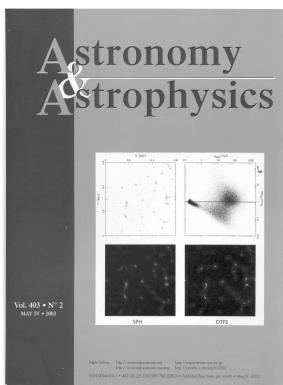
Publications in refereed journals

Density estimators in particle hydrodynamics: DTFE versus regular SPH

F.I. Pelupessy, W.E. Schaap and R. van de Weygaert

Astronomy & Astrophysics, 2003, 403, 389

This article featured on the cover of Astronomy & Astrophysics:



Deprojection of luminosity functions of galaxies in the Coma cluster

M. Beijersbergen, W.E. Schaap and J.M. van der Hulst

Astronomy & Astrophysics, 2002, 390, 817

Continuous fields and discrete samples: reconstruction through Delaunay tessellations

W.E. Schaap and R. van de Weygaert

Astronomy & Astrophysics, 2000, 363, L29

The vertical extent and kinematics of the HI in NGC 2403

W.E. Schaap, R. Sancisi and R.A. Swaters

Astronomy & Astrophysics, 2000, 256, L49

Conference proceedings

Delaunay recovery of cosmic density and velocity probes

W.E. Schaap and R. van de Weygaert

Contribution to ‘Statistical challenges in modern astronomy III’, 2003, State College, Penn. State, USA, eds. G.J. Babu and E.D. Feigelson

Deprojection of luminosity functions of galaxies in the Coma cluster

M. Beijersbergen, W.E. Schaap and J.M. van der Hulst

Contribution to ‘Tracing cosmic evolution with galaxy clusters’, 2001, Italy, eds. S. Borgani, M. Mezzetti and R. Valdarnini

Reconstructing cosmological fields using tessellation methods

W.E. Schaap and R. van de Weygaert

Contribution to ‘Where’s the matter? Tracing dark and bright matter with the new generation of large scale surveys’, 2001, Marseille, France, eds. Treyer and Tresser

Tessellation reconstruction techniques

R. van de Weygaert and W.E. Schaap

Contribution to ‘Mining the sky’, MPA/ESO/MPE joint astronomy conference, 2000, Garching, Germany, eds. A. Banday et al.

Popular articles

De grootschalige structuur van het heelal

W.E. Schaap

Zenit, juli/augustus 2001

This article featured on the cover of Zenit:



Dankwoord

Allereerst wil ik mijn begeleider en promotor Rien van de Weijgaert bedanken. Beste Rien, ik bewonder je enthousiasme voor de wetenschap in het algemeen en de sterrenkunde in het bijzonder. Ik dank je voor je steun en geduld gedurende de afgelopen jaren. Hoewel je oorspronkelijk een heel ander onderzoek in gedachten had, ben ik je dankbaar dat je me de vrijheid hebt gegeven om een eigen weg in te slaan. Het resultaat daarvan ligt nu voor je en ik hoop dat je er net zo trots op bent als ik.

Met Inti Pelupessy heb ik samengewerkt aan Hoofdstuk 5. Inti, bedankt voor de plezierige samenwerking! With Emilio Romano-Díaz I have worked on Chapter 6. My dear compadre, this is only a minor part of our collaboration and the friendship we have experienced. Mucho gracias! Joop Schaye ben ik dankbaar voor zijn vriendschap en gastvrijheid tijdens werkbezoeken aan Cambridge en Princeton.

Tjeerd van Albada wil ik bedanken voor zijn bereidheid om als promotor te fungeren. De leden van de beoordelingscommissie ben ik erkentelijk voor het lezen van dit proefschrift.

Alle bewoners van het Kapteyn Instituut wil ik danken voor de goede sfeer en de geboden hulp. Mijn kameraden, Richard, Emilio, Katarina en Chris, wil ik in het bijzonder bedanken. Veel dank ben ik ook verschuldigd aan de computergroep en het secretariaat.

Enkele van de in dit proefschrift gebruikte simulaties zijn uitgevoerd door het Virgo Supercomputing Consortium waarbij gebruik is gemaakt van computers van het Computing Centre van de Max-Planck Society in Garching en van het Edinburgh Parallel Computing Centre. Deze data zijn publiek toegankelijk op <http://www.mpa-garching.mpg.de>. Ik ben het 2dFGRS team erkentelijk voor het publiek toegankelijk maken van deze survey. Peder Norberg en Shaun Cole dank ik voor hun 2dFGRS mask software.

Ik wil NWO, het Kapteyn Instituut, het LKBF, de NAVO en het MPA danken voor hun financiële steun waardoor ik zomerscholen en conferenties heb kunnen bezoeken in Dwingeloo, Cambridge, Hengeloef, Victoria, Marseille, State College, Leiden en Lunteren, alsmede werkbezoeken heb kunnen afleggen aan het MPA te Garching en het IAS te Princeton. Tevens dank ik de Rijksuniversiteit Groningen, het Kapteyn Instituut en het LKBF voor de financiële ondersteuning bij het drukken van dit proefschrift.

Mijn ouders ben ik dankbaar voor de steun die ze me altijd hebben gegeven, en in het bijzonder voor de praktische ondersteuning door geregeld op Tobias te passen.

De meeste dank ben ik natuurlijk verschuldigd aan Paula. Lieve Paula, bedankt dat je er altijd voor mij was, ook als het proefschrift wel erg veel tijd en aandacht in beslag nam! Ruim twee jaar al is Tobias ons gegeven, een ontzettend lieve jongen om voor altijd van te houden en voor te zorgen. Dat we met zijn drieën een gelukkige tijd tegemoet mogen gaan!

*Willem Schaap
Groningen, november 2006*