Constraining Cosmology with Lyman-alpha Emitters a Study Using HETDEX Parameters

Ralf Koehler

München 2009

Constraining Cosmology with Lyman-alpha Emitters a Study Using HETDEX Parameters

Ralf Koehler

Dissertation
an der Fakultät für Physik
der Ludwig–Maximilians–Universität
München

vorgelegt von
Ralf Koehler
aus München

München, den 20.01.2009

Contents

List of Figures

List of Tables

Zusammenfassung

Diese Dissertation befasst sich mit den technischen Aspekten der Messung und Auswertung von Leistungspektren von weit entfernten Galaxien um mit deren Hilfe kosmologische Parameter einzugrenzen.

Wir untersuchen die Methode der Extraktion von baryonischen Akkustischen Oszillationen (BAOs) aus dem Leistungsspektrum von Lyman-alpha emittierenden Galaxien auf seine Stabilität und Anfälligkeit für Fehler. Dazu schlagen wir ein parametrisches Verfahren vor um BAOs ohne Wissen um den genauen Inhalt des Universums zu extrahieren. Wir benutzen die Hubble Volume Simulationen um zu zeigen, dass auf diese Weise extrahierte BAOs auch trotz vorhandener Rotverschiebungseffekte, nicht linearem Wachstum und ohne Kenntnis des Bias einen zuverlässigen Datensatz darstellen. Wir zeigen, dass die Oszillationslänge der BAOs verwendet werden kann um in einem geometrischen Test robuste Resultate über verschiedene kosmologische Modelle zu erhalten. Wir untersuchen den Einfluss von grobmaschigen Abtastmethoden auf die Fähigkeit die Amplitude und Form des Leistungsspektrums innerhalb der Parameter des HETDEX Versuchs zu extrahieren. Dazu simulieren 400 Kuben mit einer Seitenlänge von $500\,h^{-1}\mathrm{Mpc}$ und zeigen, dass selbst eine extrem inhomogene Fensterfunktion und der Einfluss von Kalibirierungsfehlern, den Informationsgehalt des Datensatzes nur wenig veränden. Obwohl das Leistungsspektrum durch beide Effekte schwer verzerrt wird, kann das wahre Leistungsspektrum daraus extrahiert werden. Korrelationen die durch die Fensterfunktion und Kallibrierungsfehler eingeführt werden verringern die Fhigkeit des resultierenden Leistungsspektrums kosmologische Parameter einzugrenzen nur wenig. Der Unterschied zwischen den Überdichten der Galaxienverteilung und der zu Grunde liegenden Dichteverteilung der dunklen Materie, das so genannte Biasing, ist fr die Vorhersage von kosmologischen Tests mit Hilfe des Leistungsspektrum extrem wichtig. Wir untersuchen die beste Methode um den Biasing Faktor aus Beobachtungsdaten mit extrem länglicher Geometrie und sehr geringer Punktdichte, wie er von VIRUS-P bereit gestellt werden wird, zu extrahieren. Wir zeigen, wie das Biasing mit Hilfe von Fast Fourier Transformationen mit quaderförmigen Dimensionen zuverlässig geschätzt werden kann, die mit den Korrelationsfunktionen desselben Datensatzes auch gut übereinstimmen. Zuletzt beschreiben wir die Algorithmen und die Genauigkeit des Reduktionsprogrammes CURE, welches in der Lage ist Daten, wie sie vom im HETDEX Experiment genutzten VIRUS Spektrographen bereitgestellt werden, in Echtzeit und vollständig automatisch zu reduzieren.

Abstract

The equation of state of dark energy is currently one of the most discussed topics in astrophysics with a large number of ongoing or planned surveys. This thesis focuses on the technical aspects of measuring and recovering the power spectra of distant galaxies to constrain cosmological parameters.

We investigate the robustness of the extraction of Baryonic Acoustic Oscillations (BAOs) from the power spectrum of Lyman-alpha emitters at high redshifts. We propose a parametric method to recover the BAOs without any specific knowledge about the actual composition of the universe. We use the Hubble Volume simulation to show that BAOs extracted in this way represent a robust data set, even in the presence of redshift-space effects, non-linear growth and without any knowledge of biasing. We show that BAOs are able to robustly constrain cosmological models using the BAO scale as a standard ruler for a geometric test. We explore the impact of proposed sparse sampling techniques on the ability to recover the amplitude and shape of the power spectrum using the parameters of the HETDEX survey. We simulate 400 with a length of $500\,h^{-1}\mathrm{Mpc}$ to show that the highly inhomogeneous window function and zero-point effects do not decrease the information content of the observed data set by a large factor. Although the power spectrum is heavily distorted by both effects, the true power spectrum can be recovered. Correlations introduced by the window function and zero-point errors are slightly decreasing the ability of the resulting power spectrum to constrain cosmological paremeters. The difference in overdensities between galaxies and the underlying dark matter field, the so-called biasing, is crucial in predicting accuracy of a cosmological test using galaxies as tracer particles for the underlying power spectrum. We investigate the best method to extract a bias estimate out of pencil-beam like surveys like the VIRUS-P survey currently under way. VIRUS-P will deliver a very low density data set with a highly elongated geometry. We show that the biasing can be estimated robustly using Fast Fourier Transforms with non-equal dimensions that are in good agreement with results obtained facilitating the correlation function. We finally describe the algorithms and accuracy of the CURE pipeline that is able to reduce the data provided by the VIRUS spectrograph used in HETDEX automatically and almost in real time.

Chapter 1

Introduction

1.1 Motivation

In the last ten years, cosmology has arrived in what a lot of scientists call "precision cosmology". The times when only orders of magnitude results could be given are over. At least since the COBE (COsmic Microwave Background Explorer) mission, we are able to measure cosmological parameters on the percent and sometimes sub-percent niveau.

Another exciting part of the universe was discovered only recently. Also about ten years ago Riess et al. (1998); Perlmutter et al. (1998) discovered, using light curves of supernovae as standard candles, the accelerated expansion of the universe. We have come a long way since that, and various models that explain the accelerated expansion have been proposed. Some invoke the infamous cosmological constant, others claim that gravity itself is not Einsteinian at large scales, some argue that the Einsteinian description is enough and the measured expansion is only an illusion created by local inhomogeneities. But by far the most popular explanation is dark energy (see Sect. 1.4).

Whatever the explanation is, it will have a strong impact on the future evolution at the universe. Using the dark energy description, about 75% of the energy content in the universe are comprised of dark energy (see Komatsu et al., 2008). In this scenario, the universe is no longer dominated by matter (be it dark or baryonic), but by something else which we do not understand yet.

White (2007) argues, that it is not the task of astronomers to do the particle physicists and theorists job and put a large amount of resources into experiments to measure the nature of dark energy. Other organizations, like the Dark Energy Task Force (see Albrecht et al., 2006), show that there is indeed strong interest in the astronomers community to do exactly that.

Whatever the personal opinion is, the topic of the accelerated expansion of the universe is at least controversial, and interesting. Cosmology and the evolution and fate of the universe have always been one of the most interesting scientific topics that are able to gain a wide audience even outside the community astronomers and even physicits. The work done in this thesis tries to be just another small brick in the theoretical and experimental

framework that will, hopefully, one day explain the nature of the cosmic expansion and growth history, which is, maybe, some kind of dark energy.

1.2 Power Spectrum Cosmology

Since the 1980s (see e.g. Frenk et al., 1983; Bardeen et al., 1987; Bond & Efstathiou, 1987) power spectra have been extensively used and refined to discriminate between various cosmological parameters. The approach is quite simple: The structure we observe in the universe today, was only able to form, because small initial quantum-fluctuations (Bardeen et al., 1983). These fluctuations were blown up by inflation (see Guth, 1981) to cosmological scales, and are the perturbations in the otherwise very homogeneous universe, that allowed galaxies and clusters of galaxies to form. These overdensities can be described statistically using a power spectrum. We know the shape of the power spectrum of the Cosmic Microwave Background (CMB) and the initial matter density distribution depending on various cosmological parameters (see Eisenstein & Hu, 1998), like the total matter content, Ω_{m}, the dark energy content, Ω_Λ, or the curvature of the universe, Ω_{K}. Both spectra can be calculated using simple plasma physics (see Sect. 2.2).

However, the shape and amplitude can be distorted. This is caused either by the space-time geometry, through which we observe the power spectrum, or by to structure growth, which, upon other effects, enhances the amplitude of the power spectrum. Both effects can be calculated to a degree which enables us to derive cosmological parameters to good accuracy by comparing the observed power spectra with theoretical predictions. The most prominent experiments today is the Wilkinson Microwave Anisotropy Probe (WMAP), that has brought us the most accurate estimates of cosmological parameters available (see Komatsu et al., 2008).

Much of the work in this thesis will revolve around power spectra in various coordinate systems. It is thus useful to give a short summary of the mathematical calculation and main attributes of power spectra.

1.2.1 Power Spectrum

Throughout this work we use a Fourier transformation convention such that the inverse transformation in real space r of a quantity in k space becomes:

$$\delta(\vec{r}) \;=\; \frac{1}{(2\pi)^3} \int d^3k \, \delta(\vec{k}) \, e^{-i\vec{k}\cdot\vec{r}} \;. \tag{1.1}$$

When talking about power spectra, this work will always refer to the power spectra of overdensities (in the case of the Cosmic Microwave Background temperature fluctuations), $\delta(\vec{r})$. They are readily calculated from the mean local density, $\rho(\vec{r})$:

$$\delta(\vec{r}) \;=\; \frac{\rho(\vec{r}) - \bar{\rho}}{\bar{\rho}} \;. \tag{1.2}$$

The mean local density is dependent on redshift, distance to the observer and other parameters, which are in turn dependent on the position $\vec{r}$. The correlation function $\xi(r)$ describes the correlation of the density field with itself (autocorrelation) at a given distance r:

$$\xi(r) = \langle \delta(\vec{r'} + \vec{r})\, \delta(\vec{r'}) \rangle . \tag{1.3}$$

The power spectrum, $P(k)$, is now the Fourier transform of the spatial two point correlation function $\xi(r)$, and vice versa,

$$P(k) = \int d^3 r\, \xi(\vec{r})\, e^{i\vec{k}\cdot\vec{r}} . \tag{1.4}$$

With the definition of the Dirac delta distribution,

$$\delta^{\mathrm{D}}(\vec{k}) = \frac{1}{(2\pi)^3} \int d^3 r\, e^{\pm i\vec{k}\cdot\vec{r}} , \tag{1.5}$$

and the Fourier transformed overdensities,

$$\delta(\vec{k}) = \int d^3 r\, \delta(\vec{r})\, e^{i\vec{k}\cdot\vec{r}} . \tag{1.6}$$

the relation between the power spectrum and the fluctuation can be expressed explicitly:

$$\langle \delta(\vec{k})\, \delta^*(\vec{k'}) \rangle = (2\pi)^3\, P(k)\, \delta^{\mathrm{D}}(\vec{k} - \vec{k'}) . \tag{1.7}$$

Note that an asterisk denotes the complex conjugate. The power spectrum depending on $\vec{k}$ is a direct measure of the power carried per fluctuation mode $\vec{k}$.

1.2.2 Geometry

A geometrical test involves the transformation of a specific scale length, in this case the scales provided by the theoretically predicted power spectrum, depending on space-time geometry. A power spectrum survey measures angular positions, ϕ, on the sky, and redshifts, which are transformed into the co-moving coordinates, x, using a certain reference cosmology. The co-moving coordinates are used to calculate a test power spectrum, which can then compared to a theoretically predicted power spectrum that acts as a standard ruler. The transformations for angular, $x_\parallel$, and line of sight, $x_\perp$, coordinates

$$x_\parallel = \int_{z_1}^{z_2} \frac{c}{H(z)}\, dz , \tag{1.8}$$

$$x_\perp = \phi \int_0^{z_2} \frac{c}{H(z)}\, dz , \tag{1.9}$$

depends mostly on the Hubble constant, $H(z)$. In a flat universe, the evolution of the constant can be expressed like:

$$H(z) = H_0 \sqrt{\Omega_m \left(1+z\right)^3 \, + \, \Omega_\Lambda \, e^{3 \int_0^z \, d\ln(1+z') \, [1+w(z')]}} \,, \tag{1.10}$$

with c being the speed of light and the Hubble constant today $H_0 \approx 72 \frac{km}{s \, \mathrm{Mpc}}$. The measured transformed distances can than be compared to the predictions. However, it is often easier to calculate the power spectrum only once and use a "stretching factor", s between two different cosmologies to stretch it. Because we have two angular dimensions and one line of sight dimension, the stretch factor can be calculated from the ratio of scale lengths between a reference cosmology, x and a test cosmology, $\hat{x}$, like:

$$s \left(\Omega_\Lambda, \Omega_{\mathrm{m}}, w\right) = \frac{1}{3} \frac{\hat{x}_\parallel}{x_\parallel} + \frac{2}{3} \frac{\hat{x}_\perp}{x_\perp} \,. \tag{1.11}$$

The one dimensionally average power spectrum scales with this stretching factor as follows,

$$P(k, s) = s^{-3} \, P\left(k \, s\right) \,. \tag{1.12}$$

The biggest impact, the boost or reduction of amplitude is explained by the volume independence of the power spectrum. If the same power is detected in a larger volume, the amplitude of the power spectrum goes down and vice versa.

1.2.3 Growth

Growth measures the speed of structure growth (or growth suppression when looking backward in time) due to gravitational forces. On first order dark energy is delaying structure growth, as it leads to an acceleration of the expansion of the universe. This results in structures being formed at a later time compared to a universe without dark energy.

Linear growth can be calculated by solving the growth equation:

$$\frac{\partial^2 \delta}{\partial t^2} + 2 \, \frac{\dot{a}}{a} \frac{\partial \delta}{\partial t} = 4\pi \, G \, \rho \, \delta \,, \tag{1.13}$$

with G being the the gravitational constant, a the expansion factor of the universe and the time t.

Assuming a flat universe the growth function of the growing mode can be approximated (see Peebles, 1993):

$$\delta(t) \sim \frac{\dot{a}}{a} \int_0^t \frac{da}{\dot{a}^3} = H(z) \frac{5 \, \Omega_m}{2} \int_z^\infty \frac{1+z'}{H^3(z')} \, dz' = D(z) \,, \tag{1.14}$$

with $\dot{} = d/dt$, and the scale factor $a \propto (1+z)^{-1}$. The power spectrum itself scales with the growth ratio D only in amplitude

$$P(\vec{k}, D) = D^2(z) \, P(\vec{k}) \,. \tag{1.15}$$

By measuring the amplitude of the power spectrum at different times, for example at the epoch of recombination and at a redshift of $z = 1$, we can pinpoint a reference cosmology, that can be compared to measurements at other redshifts to constrain cosmological parameters.

1.3   Lyman Alpha Emitters

Lyman-alpha emitters (LAEs) (see Charlot & Fall, 1993) are young star-forming galaxies. The mostly ultraviolet emission of young and hot stars is absorbed by neutral hydrogen and re-emitted as Lyman-alpha emission at a wavelength of $121.6\,nm$. This absorption and re-emission process can occur a number of times, till the radiation finally escapes the galaxy.

The main advantage of LAEs in surveys is the ability to detect the Lyman-alpha emission line of high-redshift galaxies in the visible spectrum. The emission line is much easier to detect than the galaxies of continuum emission and thus enables us to observe galaxies which are not accessible by any other means. After a long period of unsuccessful searches for LAEs, the first emitters were found in the mid 1990 (see e.g. Steidel et al., 1996; Hu et al., 1998) and have led to a run for finding the most distant emitters today (see e.g. Taniguchi et al., 2005; Ota et al., 2008).

Because of their properties, LAEs are ideal tracers for the dark matter distribution at high redshifts. However, some problems have to be solved to facilitate their full potential.

1.3.1   Biasing

The properties of Lyman-alpha emitters are still not very well understood. Although there have been efforts to model the population and evolution of LAEs (see Delliou et al., 2006; Orsi et al., 2008), statistical properties are still not well known. Their mostly irregular shape, the fact they are young and maybe still in formation further complicates the process of understanding LAEs.

Especially important for power spectrum cosmology is the so-called biasing. It describes the fact that galaxies tend to form only in the highest overdensities of the underlying dark matter distribution. The amplitude of the observed power spectrum, $P_{\mathrm{obs}}(k)$ is thus boosted by the scale dependent biasing $b(k)$, compared to the dark matter power spectrum, $P_{\mathrm{DM}}(k)$:

$$P_{\mathrm{obs}}(k) = b^2(k)\,P_{\mathrm{DM}}(k)\,. \tag{1.16}$$

Although biasing does not affect cosmological tests looking only at the scale lengths of the power spectrum, it seriously affects growth tests. The best biasing estimates for LAEs at the moment are summarized by Orsi et al. (2008). The error of these measurements is still in the range of 80-100%. Knowing the biasing of the observed galaxy population is also important for the prediction of the expected observation accuracies. Some statistical errors are not dependent on the amplitude of the power spectrum. If the amplitude of the power

spectrum increases and these errors stay constant, the overall accuracy of the measurement increases. As biasing enters the power spectrum as its square, a small difference in basing can change the result by a large amount.

1.3.2 Contamination

Another problem in using the population of Lyman-alpha emitters as tracers for the underlying dark matter density field is the possibility of contamination. LAEs are not the only line-emission galaxies in the targeted spectral range. OII emission lines ($372.7\,nm$) originating from much closer galaxies (at lower redshift) can be mistaken for highly redshifted Lyman-alpha emission, if the spectral range is not brad enough to find the related OIII line. Or an OIII line emission ($496.1/500.7\,nm$) could be mistaken for even higher redshift LAEs, if the spectral resolution is not good enough to resolve the OII lines. The observed power spectrum, $P_{\mathrm{obs}}(\vec{k})$, is then a sum of the true LAE power spectrum, $P_{\mathrm{LAE}}(\vec{k})$ and the OII power spectrum, $P_{\mathrm{OII}}(\vec{k}s)$:

$$P_{\mathrm{obs}}(\vec{k}) = P_{\mathrm{LAE}}(\vec{k}) + f^2 \, P_{\mathrm{OII}}(\vec{k}s). \tag{1.17}$$

The OII power spectrum is stretched by a cosmology dependent stretching factor s, as the redshift information is mistaken for LAE redshifts at the given wavelength. The sum is weighted by the fraction of contaminating OIIs, f, from the whole sample.

1.4 Dark Energy

Observations show that the acceleration of the universe seems not to decrease, like one would expect from a matter dominated universe, but to increase. These observations are made using and combining various techniques, like the measurement of the angular power spectrum of the cosmic microwave background (see Komatsu et al., 2008), the observation of light curves of supernovae (see e.g. Riess et al., 2007), the distribution of galaxies on the sky (see e.g. Cole et al., 2005; Percival et al., 2001b) or weak lensing signals of the dark matter distribution (see e.g. Hoekstra et al., 2006). The expansion of the universe is expressed by Einsteins field equation:

$$\frac{\ddot{a}}{a} = -\frac{4}{3}\pi G \left(\rho + 3p \right) , \tag{1.18}$$

$$w = \frac{p}{\rho} , \tag{1.19}$$

with the density, ρ, and pressure, p, of the dominant constituent of the universe. The parameter w is the equation-of-state parameter (see Turner & White, 1997). One can easily see, that a w value, smaller than -1/3 leads to a positive acceleration $\frac{\ddot{a}}{a}$.

Various forms of solutions have been proposed for the problem that known today as "dark energy" (see e.g. Peebles & Ratra, 2003).

The first solution proposed was the Cosmological Constant (see e.g. Copeland et al., 2006), introduced and later dismissed by Einstein himself. A Cosmological Constant has a w-value of exactly -1. The biggest problem of a cosmological Constant is that it can not be explained by any means in the modern theoretical framework (see Weinberg, 1989). The best explanation, the zero-point energy of vacuum , is off by 120 order of magnitude.

Another explanation are various forms of scalar or vector fields, called dark energy, quintessence, k-essence, or phantom energy (see e.g. Copeland et al., 2006). They vary, upon other things, in their w-value both in absolute value and in time. Depending on the model, they solve issues like the question why the dark energy density is comparable to the dark matter density today, which means that it has to be extremely fine tuned at earlier times in the universe. All scalar field solutions are similar to the scalar fields responsible for the primordial inflation of the universe.

A further theory claims that Einsteinian gravity itself does not hold at the very large scales of the universe. It is called Modified Gravity (see e.g. Nojiri & Odintsov, 2006) and expands Einsteinian gravity at large scales, while maintaining the limits we get from constraints in the solar system at small scales.

More exotic solutions can be found using string theory. The braneworld model (see e.g. Shani & Shtanov, 2003) describes our universe as a 3-dimensional brane that is embedded in a 4 or more dimensional bulk. All forces except gravity stay on our brane, while gravity is leaking into the bulk and can thus generate phenomena like the ones observed and termed "dark energy".

Finally, some models claim that one does not need any new scalar fields, dimensions or modified gravity it all. The so called backreaction models (see e.g. Kolb et al., 2006) claim, thatlarge scale inhomogeneities make is believe that the close universe (both in space and time) is expanding faster than the more distant (both in scape and time) universe.

At the moment only astronomical observations probing the geometry and growth rate of the structure of the universe seem to have the power to discriminate between the various proposed models.

1.5 HETDEX

The Hobby-Eberly-Telescope (see Fig. 1.1), located in West Texas, has one of the largest constructed mirrors with a dimension of 11.1 times 9.8 meters. However, only about $9.2\,m$ effective aperture is used at any given moment, as the telescope tracks elevation by moving the tracker instead of the mirror. It is thus ideally suited to do surveys of certain fractions of the sky.

HETDEX, the Hobby-Eberly-Telescope Dark Energy eXperiment, is designed to survey 0.8 million Lyman-alpaha emitters at a redshift range of $1.9 < z\,3.5$ to measure their power spectrum and constrain the expansion and growth history of the universe. It was originally designed to measure the constant and time dependent term of w, but has just started to look at other ways to constrain cosmological parameters using the expected data set.

Figure 1.1: The Hobby Eberly Telescope

1.5.1   VIRUS

To achieve this, HETDEX will build and facilitate the Visible Integral-Field Replicable Unit Spectrograph (VIRUS).

The VIRUS consists of about 92 wide-field integral-field units (IFUs) consisting of about 500 fibers each. The fibers are fed to 184 spectrographs with a resolution of $R \approx 800$ and a spectral range of 350 to 550 nm. Each fiber will have a diameter of $\sim$ 1.5 arsecs with a somewhat larger separation between each fiber, leading to a field of view of about 1 arcmin for each fiberhead. Using three dither positions of 300 seconds each per observation, to reach a fill-factor in each fiber head of 1, VIRUS will reach a flux limit of $3 \times 10^{-17}\,\mathrm{ergs\,cm^{-2}\,s^{-1}}$ at 5-σ.

The 90 fiber heads of the IFUs will be arranged in the focal plane of the telescope in a regular grid. The new wide-field corrector developed for the instrument has a field of view of 20 arcmin diameter, leading to a total fill-factor of $\sim 1/7$th. This will help to increase the volume of surveys taken with the VIRUS at the expense of sparser sampling (see Chapt. 3). The instrument itself is stationary, because of its total weight of 30,000 kg. It is connected to the fiber heads by fibers with a length of $20\,m$.

1.5.2   The Survey

HETDEX is going to survey an area of $420\,deg^2$ for three years to obtain an expected 0.8 million Lyman-alpha emitters. With the expected redshift coverage of $1.9 < z < 3.5$, the survey will cover a co-moving volume of over $3h^{-3}\,$Gpc. An expected 4000 fields will be observed using a total of 1400 hours.

The survey region is shown in Fig. 1.3. It is located right on top of the big dipper.
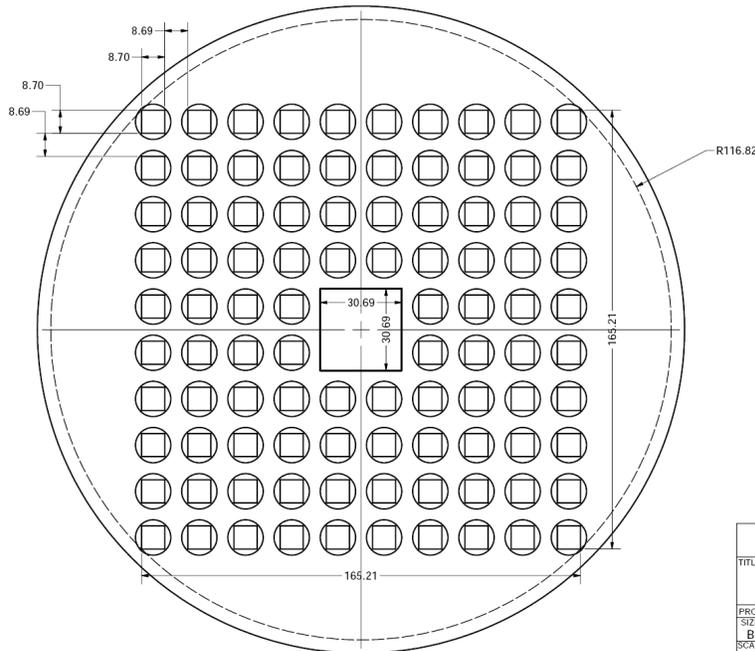
Figure 1.2: Proposed layout of the VIRUS fiberheads in one shot.

Besides the 0.8 million Lyman-alpha emitters, it is going to get spectroscopic data from about 1 million OII emitting galaxies, about 0.4 million other galaxies, 0.25 million stars and about 2000 Abell richness clusters.

HETDEX will mostlikely need a complementary imaging survey to separate detected OII emission lines from LAEs. This is possible, as OII lines have a much higher equivalent width, so that their continuum should be detectable in a modestly deep survey in the V ad B band to about 25.5 AB magnitudes.

1.5.3 Cosmology

HETDEX, being a spectroscopic survey, is going to measure the power spectrum in all three dimensions using spectroscopic information. This gives rise to various opportunities on constraining cosmology:

First and foremost, the oscillating features created by baryons (Baryonic Acoustic Oscillations, or BAOs) can be used to get a bias free and very robust measurement of the co-moving distance in angular direction and the co-moving distance in line of sight direction which is a direct measure for the expansion rate $H(z)$ (see 1.8). The robustness and performance of BAOs is investigated in Chapt. 2.

However, HETDEX will try and measure the whole shape of the power spectrum to fully exploit the available information. Detecting the effect of dark energy in a measure of the expansion rate $H(z)$ would be the first direct detection of dark energy. Using the whole information contained in the power spectrum, HETDEX would be able to detect

Reddening map with baseline survey limits
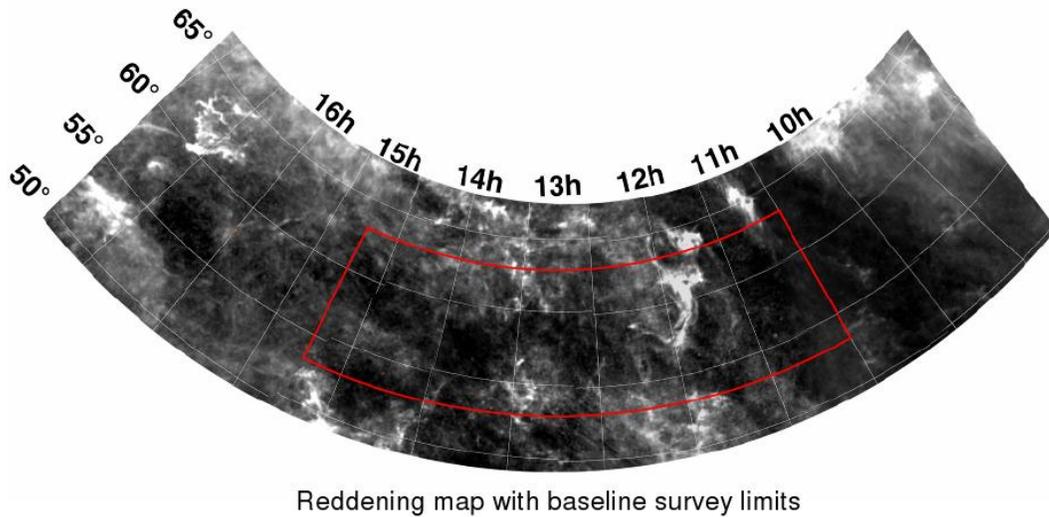
Figure 1.3: HETDEX survey layout with reddening map.

dark energy directly at a 3σ level. Figure 1.4 shows the predicted accuracies in $H(z)$ compared to other experiments.

The measurement of the co-moving distance in angular dimensions, and thus the angular diameter distance D_A, can be used to directly constrain the curvature of the universe at the observed redshift (see Knox, 2006). HETDEX would be able to reach an accuracy of about 0.2% in curvature, which is about a magnitude better than constraints available today. This will help low-redshift experiments, by providing good priors to their calculations.

However, the biggest advantage of HETDEX, compared to other ground based surveys, is the high redshift and high redshift-range. This does not only help with constraining the expansion and growth history of the universe over a larger z-space, but also allows to directly compare two different redshifts using the same data set. This is especially important, as most models of dark energy require an evolution of the w parameter to avoid the fine tuning problem mentioned above.

1.5.4   VIRUS-P

To investigate the properties of Lyman-alpha emitters, especially their number density, luminosity function, and biasing, the VIRUS Prototype (VIRUS-P) survey was started in 2007.

The VIRUS-P consists of a single VIRUS IFU with 246 fibers and a field of view of $\sim 4\,\mathrm{arcmin}^2$. Using 6 hour exposures on the $2.7m$ telescope, VIRUS-P has a 5-σ flux limit of $6 \times 10^{-17}\,\mathrm{ergs\,cm^{-2}\,s^{-1}}$ at 430 nm.

Figure 1.5 shows the VIRUS-P instrument mounted on the $2.7m$ telescope conducting the pilot survey. Up to now more than 20 fields have been observed with 6 hour exposures. About 100 galaxies have been detected, half of which are securely confirmed as OII emitters
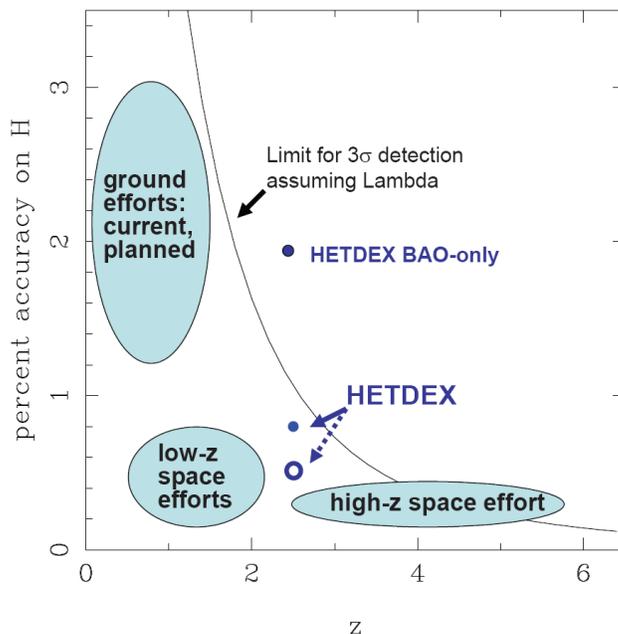
Figure 1.4: Predicted detection accuracy for HETDEX on H. Figure provided by Karl Gebhardt.

and one third as Lyman-alpha emitters.

It is planned to survey two fields with an area of $50 - 100\,\mathrm{arcmin}^2$ each until start of 2011. Until then an expected number of 300 Lyman-alpha emitters will be found, together with 500 OII emitters. The redshift range is the same as for the VIRUS, leading to a very elongated survey volume with a length-to-width ratio of about 60. Chapter 4 investigates how biasing can be estimated from pencil-beam shape surveys, like the VIRUS-P.

1.6   Overview

The first Chapter 2 will investigate the robustness of the Baryonic Acoustic Oscillations as a cosmological probe. It will suggest a method to extract BAOs from real life power spectra without any prior knowledge about the specific cosmological parameters. It will show that this method is robust in a non-linear, redshift-space environment with biased galaxies. Finally a cosmological test will be performed to show that the presented method is able to robustly recover the equation of state of dark energy w from a set of observed galaxies.

Chapter 4 will look at the ability to recover the three dimensional power spectrum using data from a survey applying sparse sampling strategies. Both the window function and zero-point effects have a strong impact on the shape, the errors and the correlations of the resulting data set. We show, that all effects can either be corrected for or have a small impact on the information content of the data set. We show this by performing a

Figure 1.5: The VIRUS Prototype mounted on Harlan J. Smith Telescope

cosmological test probing for both the expansion and growth history of the universe.

Chapter 3 will explore the various possibilities to recover biasing from a pencil-beam like survey. We show that standard power spectrum estimation carries the risk of being biased towards higher amplitudes at samples with very low point densities. We compare the ability of power spectra generated with two different algorithms and the correlation function to constrain galaxy biasing in a pencil beam like survey using a cosmological test.

Chapter 5 describes the CURE reduction pipline for VIRUS. It gives an overview of the basic reduction process and the modeling of observed data to achieve optimum sky subtraction. The detection algorithm of emission line objects is also described. Finally, the performance of CURE to detect emission line galaxies is tested using simulated emitters in real data.

Chapter 2

Probing Dark Energy with Baryonic Acoustic Oscillations at High Redshifts

Note: This chapter was published as Koehler et al. (2007). It expands and heavily uses results from Koehler (2005) (Master Thesis).

2.1 Introduction

Acoustic oscillations as observed in the temperature anisotropies of the cosmic microwave background radiation (CMB) are traditionally used to constrain the values of certain cosmological parameters. The intrinsic amplitudes and locations of the oscillations are determined by the densities and pressures of the various energy components in the very early and hot Universe. The oscillations are furthermore modified by their subsequent geometric projection onto the present hypersphere, where they are observed. It turns out that both intrinsic CMB oscillations and their geometric projection are mainly shaped by the physical properties at high redshifts. Excluding the very large scales where cosmic variance makes analysis quite difficult, the observed oscillations in the CMB are highly degenerate against certain changes of the energy density ρ_{DE} and the equation of state parameter, $w = p_{\mathrm{DE}}/\rho_{\mathrm{DE}}c^2$, of the dark energy, dominating at low redshifts (see Caldwell et al., 1998).

However, related so-called baryonic acoustic oscillations (BAOs) are expected to be observable at lower redshifts in the matter power spectrum. Therefore, the phases of the BAOs can be used as cosmic rulers in a similar manner as the CMB oscillations, but now at much smaller redshifts (Eisenstein et al., 1998). Here, the observed phases are affected by the late-time geometry and thus by the value of w, the prime cosmological parameter of the present investigation.

The BAOs are classical Doppler peaks in the density distribution of matter (e.g. Hu & Sugiyama, 1996). They are triggered in the hot Universe by oscillatory velocity patterns of the baryons. These velocity oscillations are generated, in the same environment as the

CMB density oscillations, by sound waves on scales where radiation pressure could stabilize the fluid against gravitational collapse. At later times, this oscillatory velocity pattern of the baryon field produces kinematically a new field of matter density fluctuations in form of quasi-regular matter oscillations. The amplitudes of these BAOs are determined by the ratio of baryonic matter to the overall matter density. The BAOs thus constitute a quasi-regular pattern of oscillating substructures superposed with small amplitudes onto a general density field which fluctuates irregularly with several 100 times higher amplitudes.

The imprints of the BAOs on galaxy and clusters distributions, gravitational shear maps etc. can provide at least in principle a clean cosmic ruler for precise tests of the w parameter. The characteristic scale, s, of these imprints is the comoving distance, sound waves can travel during the epoch where baryons and photons are strongly coupled through Compton and Thomson scattering (Compton drag). Cosmological tests based on the resulting sound horizon, s, as a metric ruler are expected to have only very small systematic errors because the phases of the BAOs located at scales small compared to s are solely determined by well-understood physical processes. A practical problem is, however, to separate the oscillatory BAO modes from the irregular fluctuation field with sufficient accuracy.

One might object that the small amplitudes of the BAOs can easily be washed out especially by structure growth which can mix perturbation modes with adjacent wavenumbers k. This is certainly true for oscillations in the non-linear regime (Meiksin, White & Peacock 1999). On larger scales, however, recent observations of 2dF galaxies at redshifts $z < 0.3$ clearly show several BAOs in the galaxy power spectrum (Cole et al., 2005). In addition, the sequence of BAOs in k-space is projected into a single wiggle in the space domain. The corresponding excess galaxy correlation was in fact observed in the two-point spatial correlation function of the luminous red SDSS galaxies at $z < 0.5$ (Eisenstein et al., 2005). In addition to these observational indications, results from recent numerical simulations suggest that within certain redshift and scale ranges, stable BAOs could exist as useful probes for cosmological investigations (e.g. Seo & Eisenstein, 2003; Springel et al., 2005; Jeong & Komatsu, 2006). However, several critical issues like the exact behavior of baryons during structure growth on BAO-scales can only be discussed with much larger simulations and a better understanding of galaxy formation.

Applications of BAOs for cosmological tests of w are confronted with the following situation. Present observational constraints of w using different combinations of CMB, galaxy clusters, galaxies, and gravitational lensing data are all found to be consistent with $w = -1$, i.e., the value for the cosmological constant. The 1σ error of the w values, derived from partially dependent observations, is around 10-20% (for a recent review see, e.g., Schuecker, 2005). To improve current estimates, we thus have to extract the BAOs and to measure their phases quite accurately in the presence of non-linear and scale-dependent effects. Therefore, even tiny systematic errors in the analysis on percent levels can have severe consequences for the observational accuracy of w.

The basic aim of the present paper is to describe a new, simple, and robust method to extract BAOs from (non)linear, scale and redshift-dependent biased, redshift-space galaxy power spectra under realistic lightcone survey conditions. For brevity, we call the method

'fit and extract' (FITEX). The method has the potential to reduce many of the above-mentioned sources of systematic errors. Our starting point is to use as few assumptions as possible for the extraction of the BAOs from a complex power spectrum. For reasons which will become clear later, we do the separation by fitting a flexible non-oscillating function to an observable which is directly related to the transfer function and not to the power spectrum itself. The crucial point is to show that BAOs extracted in this simple manner are still governed by simple physical processes, and that FITEX stays robust even under complex survey conditions. We believe that such model-independent approaches are quite important, especially in light of the fact that at least in the near future we cannot expect to model the abovementioned non-linear and scale-dependent effects with accuracies on the sub-percent level.

Our computations are in several cases optimized to galaxy surveys covering typically several 100 square degrees on the sky in the redshift range between $z = 2 - 4$. This choice is motivated by the Hobby Eberly Dark Energy Experiment (HETDEX), which is planned to measure the w parameter with several million Ly-α emitting galaxies at these redshifts (see Hill et al., 2008). We see the results of the present paper as a useful contribution to a realistic error forecast for this important project.

We organized the paper as follows. In Sect. 2.2 we motivate the basic idea of the FITEX algorithm. In the following sections we test its performance to separate BAOs from a complex power spectrum under linear conditions (Sect. 2.3), under quasi-non-linear conditions (Sect. 2.4), in redshift space (Sect. 2.5), and from biased samples (Sect. 2.6). Finally, we illustrate the performance of FITEX including all effects and use a cosmological test of w as a benchmark to investigate the quality of the BAO extraction together with the theoretical template (Sect. 2.7). Our results are mainly based on a deep data wedge extracted by the Virgo Consortium from the Hubble Volume Simulation. The paper thus presents for the first time constraints on the application of BAOs for cosmological tests under realistic lightcone survey conditions.

If not mentioned explicitly, we assume a spatially flat Friedmann-Lemaître Robertson-Walker world model with the Hubble constant in units of $h = H_0/(100 \, \mathrm{km \, s^{-1} \, Mpc^{-1}})$, the present values of the total matter density $\Omega_\mathrm{m} h^2 = 0.147$ and baryon density $\Omega_\mathrm{b} h^2 = 0.0196$, the density of relativistic matter (e.g. neutrinos) $\Omega_\nu = 0$, and the mean CMB temperature $T_\mathrm{CMB} = 2.728 \, \mathrm{K}$ (concordance cosmology).

2.2 Baryonic Acoustic Oscillations (BAOs)

In this section, we discuss the 'theoretical wiggle function', that is, a reference function (see Eq. 2.1 below) we use in our cosmological tests of w to match the BAOs extracted in a certain manner from the complex galaxy power spectrum. The power spectrum may be written in the form $P(k, z) = A(k, z, b) \, k^n \, T^2(k)$. The amplitude $A(k, z, b)$ includes the primordial amplitude, redshift and scale-dependent effects of redshift space distortions, linear and non-linear structure growth, and galaxy biasing which will be discussed in the course of the paper. The exponent n is the slope of the primordial power spectrum, and

$T(k)$ the transfer function. Our further treatment of the formation of BAOs is based on Hu & Sugiyama (1996) where a detailed description of the relevant physical processes can be found, and on the fitting equations derived by Hu et al. (1998).

Between the epoch of matter-radiation equality at redshift $z_{eq} \approx 3526$ (concordance cosmology) and the end of the Compton drag epoch at $z_d \approx 1026$, the baryons follow a regular velocity pattern which can generate new density fluctuations kinematically. On small scales and at z_d, this effect (velocity overshoot) overrides the intrinsic density fluctuations of the baryons. As adiabatic modes dominate the isocurvature modes, the baryon density oscillates as $\sin(ks)$. The phase of the BAOs is frozen out at z_d at the value ks, with $s \approx 152\,\mathrm{Mpc}$ being the comoving sound horizon at z_d and k the comoving wavenumber of the oscillation. The BAOs are thus $\pi/2$ out of phase with the corresponding CMB fluctuations. The sound horizon s at z_d is the standard ruler we are looking for. Its value calibrates the phases of the theoretical wiggle function and can be measured quite easily with CMB experiments on the sub-percent level. This includes the contribution of unknown relativistic energy components (see Eisenstein & White, 2004). For consistency with the Hubble Volume Simulations (see Sect. 2.4) we use $\Omega_\nu = 0$.

The sinusoidal fluctuations in the baryon density are dampened by expansion drag, gravitational forcing, and Silk damping. In addition, cosmic expansion forces the velocity contributions to fall off at large scales, and amplitudes decline when dark matter dominates the energy density. While these effects only reduce the amplitude of the fluctuations and do not affect the phase, the damping may be summarized by the pseudo transfer function $T_w(k) \sim j_0(k\,s)e^{-(k/k_{Silk})^{m_s}}/[1 + (\beta_b\,/\,k\,s)^3]$, with j_0 the spherical Bessel function of order zero, the exponent $m_s \approx 1.4$ which is basically independent from cosmology, $k_{Silk} = 1.6(\Omega_b h^2)^{0.52}(\Omega_m h^2)^{0.73}[1+(10.4\Omega_m h^2)^{-0.95}]$, and $\beta_b = 0.5 + f_b + (3-2f_b)\sqrt{(17.2\Omega_m h^2)^2 + 1}$ with $f_b = \Omega_b/\Omega_m$. However, velocity overshoot dominates only on scales small compared to s. On larger scales, the original sound horizon, s, appears to be reduced by $(\tilde{s}/s)^3 = 1 + (\beta/ks)^3$, with $\beta = 8.41(\Omega_m h^2)^{0.435}$. This correction of s is about 0.2% at $60\,h^{-1}\,\mathrm{Mpc}$ and 1.3% at $120\,h^{-1}\,\mathrm{Mpc}$. Though relatively small, the corrections are important as they change the length of the standard ruler s depending on k. Collecting all scale-dependent terms we get the oscillatory solution (theoretical wiggle function)

$$T_w(k) \sim \frac{e^{-(k/k_{Silk})^{m_s}}}{1 + (\beta_b\,/\,k\,s)^3}\,j_0(k\,\tilde{s})\,. \tag{2.1}$$

A way to extract BAOs from a complex power spectrum can be found when we specify the relation between the wiggle function, $T_w(k)$, and the total transfer function $T(k)$, which we discuss now. Each particle species, in the present case CDM and baryonic matter with the corresponding densities Ω_c and Ω_b, should have separate effective transfer functions, T_c and T_b. Though, after the drag epoch at z_d, baryons appear basically pressureless and will fall into the potential wells of CDM. This results in a transfer function valid for both species of matter,

$$T(k) = \frac{\Omega_c}{\Omega_m}T_c(k) + \frac{\Omega_b}{\Omega_m}\,T_b(k)\,, \tag{2.2}$$

with $T_c(k) = f\tilde{T}_0(k, 1, \beta_c) + (1-f)\tilde{T}_0(k, \alpha_c, \beta_c)$, written in terms of the generalized transfer function $\tilde{T}_0(k, \alpha_c, \beta_c)$ as defined in Eisenstein & Hu (1998). Here, $f = 1/[1 + (ks/5.4)^4]$ smooths the combination of the almost baryon-free and baryon-loaded solutions near s, and

$$T_b(k) = \tilde{T}_b(k) + T_w(k), \tag{2.3}$$

where $\tilde{T}_b(k) = \frac{j_0(k\tilde{s})\tilde{T}_0(k,1,1)}{1+(ks/5.2)^2}$. These approximations are better than 2% for $\Omega_b / \Omega_m < 0.5$. Note that $\Omega_b/\Omega_m \to 0$ corresponds to $\alpha_c, \beta_c \to 1$. If $\Omega_c >> \Omega_b$, the effects of dark matter dominate over velocity overshoot of the BAOs. Written in this way, we immediately see that the oscillatory part (2.1) is added on top of a non-oscillatory part and can be computed by subtracting off a smooth continuum from the complete transfer function. This process establishes the basic methodology of FITEX.

The different terms of the transfer function are shown in Fig. 2.1. The upper panel shows the components which determine the global shape of the transfer function. Note that an additional oscillating component, the first term in Eq. (2.3), is necessary to describe the transition from large to small scales with sufficient accuracy. This is a direct consequence of the fact that the main effect of the baryons on the transfer function is a damping of the overall growth of dark matter between z_{eq} and z_{d} which significantly reduces the fluctuation power at scales smaller than s. The resulting break is the imprint of the baryons on the power spectrum which is easiest to observe. The term $\tilde{T}_b$ thus contributes mainly to the shape as it is rapidly declining on scales smaller than s to model the damping effect of the baryons.

In the lower panel of Fig. 2.1, the two parts of the baryon transfer function are shown. One can clearly see the first to fifth oscillations. Note that $\tilde{T}_b$ declines very fast and is almost zero after the second oscillation. It is compared to the second term of the baryonic transfer function, $T_w(k)$, and a reference function, $T_{w,\mathrm{ref}}$, which is formally constructed to contain all oscillations. Thus, the reference function includes both the small-scale oscillations from $T_w(k)$ and the large-scale oscillations from $\tilde{T}_b(k)$. The latter is, however, also determined by the dark matter function $T_0(k, 1, 1)$ (see below Eq. 2.3). Therefore, excluding or including knowledge of the shape of the dark matter function for the extraction of BAOs corresponds to working either with $T_w(k)$ or with $T_{w,\mathrm{ref}}$. Both $T_w(k)$ and $T_{w,\mathrm{ref}}$ agree reasonably well beyond $k \geq 0.05 \, h \, \mathrm{Mpc}^{-1}$, especially when one keeps in mind that only the phases of the oscillations are relevant as the amplitudes are subject to distortions anyway (see below).

The good agreement between the phases of $T_w(k)$ and $T_{w,\mathrm{ref}}$ means, that quite simple physics as described by T_w (mainly velocity overshoot and Silk damping) dominate the oscillations expected to be seen in the overall power spectrum. However, on larger scales, the physics get more complex due to the growing influence of CDM and the backreaction of the baryons on structure growth. Thus T_w and $T_{w,\mathrm{ref}}$ differ accordingly. Nevertheless, on small scales, T_w is a good theoretical wiggle function for BAOs phases which can be used as a standard ruler for cosmological tests, with a minimum of theoretical assumptions.
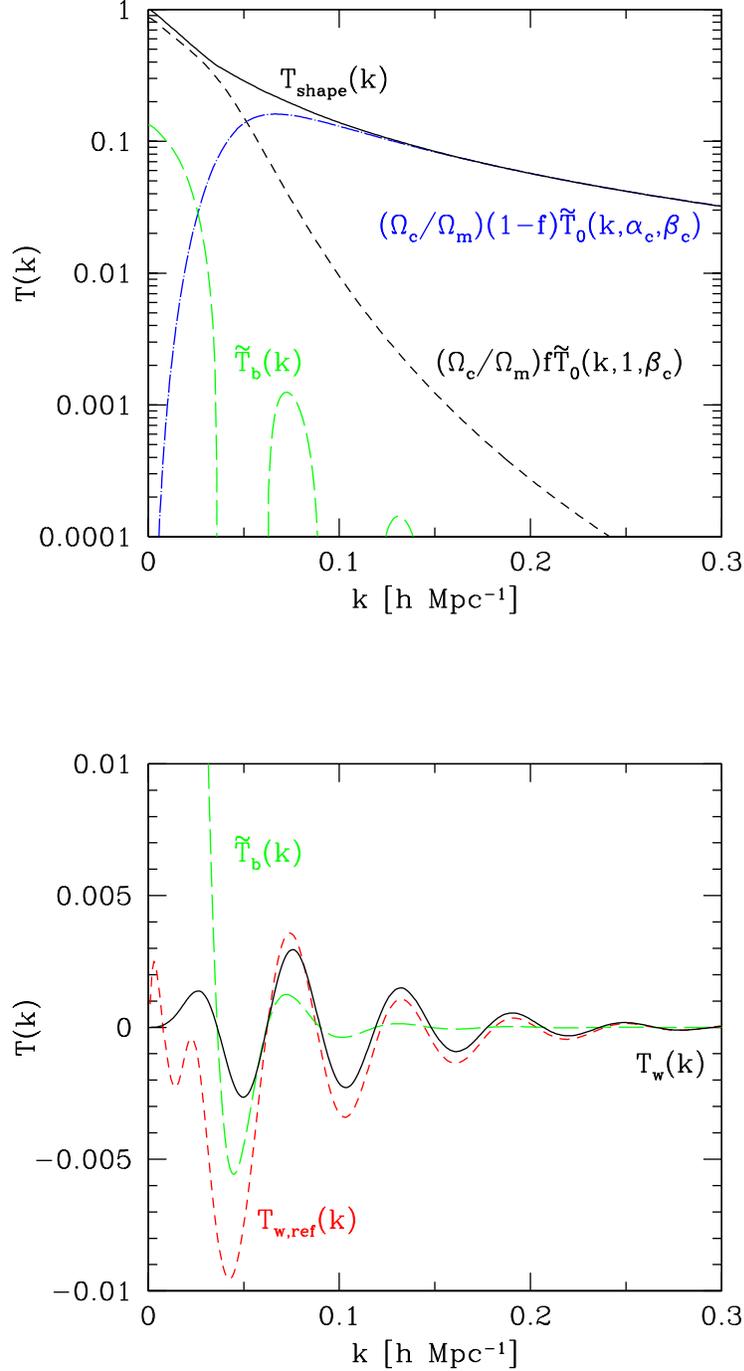
Figure 2.1: Comparison of different terms of the transfer function computed with $\Omega_{\rm b}\,h^2 = 0.0196$, $\Omega_{\rm m}\,h^2 = 0.147$ and $T_{\rm CMB} = 2.728\,{\rm K}$. **Upper panel:** Non-oscillatory parts. The function $T_{\rm shape}$ is the sum of the other three. **Lower panel:** Oscillating parts. The residuals of the subtraction between the transfer function and its corresponding non-oscillatory fit is shown as long-dashed line. Note the different scaling of the y-axes. See main text for more details.

2.3 Extracting BAOs under Linear Conditions

The main advantages of constraining cosmological parameters with BAOs are its small systematic errors, the ability to discriminate between geometrical effects (homogeneous universe) and structure growth effects (inhomogeneous universe) of cosmological parameters, and the potential to constrain dark energy without assuming a certain dark matter model. The observed phase of the BAOs is only affected by the geometry of the universe, while the observed amplitude of the oscillations is mainly affected by structure growth and other amplitude effects (see Sects. 2.4 to 2.6). It is thus very important to extract the BAOs from the power spectrum in a way that does not mix these two benchmarks. As suggested by the previous discussion, BAOs, described by simple physical processes, should be extracted by *subtracting* the shape of the non-oscillatory part of the *transfer* function. This method is able to disentangle phase effects from amplitude effects in an effective manner.

In contrast to this method, previous studies (Blake & Glazebrook, 2003; Angulo et al., 2005; Seo & Eisenstein, 2003; Hütsi, 2006; White, 2005) used the CDM-based theoretical non-oscillatory *power spectrum*, $P_{\mathrm{ref}}(k, z)$, given in Eisenstein & Hu (1998) to analyze the BAOs. This is done by *dividing* the measured power spectrum, $P_{\mathrm{obs}}(k, z)$, by this reference power spectrum. Assuming that the transfer function can be split into a shape part, T_{shape}, and an oscillatory part, T_{wiggle}, where $T_{\mathrm{shape}} > T_{\mathrm{wiggle}}$ (Sect. 2.2), the result of such a division is

$$\frac{P_{\mathrm{obs}}(k, z)}{P_{\mathrm{ref}}(k, z)} = \frac{A_{\mathrm{obs}}(k,z)\,[T_{\mathrm{shape}}(k) + T_{\mathrm{wiggle}}(k)]^2\,k^n}{A_{\mathrm{ref}}(k,z)\,T_{\mathrm{shape}}^2(k)\,k^n} \tag{2.4}$$

$$\approx 2\,\frac{A_{\mathrm{obs}}(k,z)\,T_{\mathrm{wiggle}}(k)}{A_{\mathrm{ref}}(k,z)\,T_{\mathrm{shape}}(k)} + \frac{A_{\mathrm{obs}}(k,z)}{A_{\mathrm{ref}}(k,z)}\,. \tag{2.5}$$

Eq.(2.4) gives a spectral ratio, which strongly deviates from the simple functional form of Eq. (2.1) expected from basic physical principles, and thus unnecessarily complicates any further analysis of the resulting "BAOs". A possible result of such a division is given in (Fig. 6 Springel et al., 2005). It can easily be seen that phase information and amplitude information is mixed.

In addition, the use of a *theoretical* non-oscillatoy reference power spectrum, $P_{\mathrm{ref}}(k, z)$, or the related *theoretical* "boosted" transfer function, $A_{\mathrm{ref}}(k, z)T_{\mathrm{ref}}(k, z) = \sqrt{P_{\mathrm{ref}}(k, z)/k^n}$, has a number of drawbacks:

(1) The shape of the transfer function and thus the power spectrum vary with the cosmological parameters. To compute a reference power spectrum one has to know the exact values of several cosmological parameters. Furthermore one has to assume a CDM model as a prior. (2) The reference power spectrum has to be flexible to compensate for various distortions: structure growth, redshift space distortions, biasing etc. distort the transfer function and thus its shape. Most of the effects vary with redshift as well as with k. (3) The reference power spectrum has to be exact on the sub-percent level. The BAOs make up only $\sim 2\%$ of the transfer function depending on the ratio $\Omega_{\mathrm{b}}/\Omega_{\mathrm{m}}$. An analytical function has to be more accurate to extract the oscillations from the transfer function. (4)

It is non-trivial to provide an assumption-free template for the resulting power ratio to perform a cosmological test. Instead, very accurate knowledge of the cosmology as well as amplitude effects is needed.

Most of these issues are well-known and useful corrections are already available from the simulations. However, these calibrations have certainly not reached the sub-percent accuracy level which is necessary for the cosmological tests, and one can doubt whether this is achievable at all. Furthermore, one of the main advantages of the BAO method, the ability not to assume a certain dark matter model, is negated.

The Eqs. (2.1-2.3) instead suggest that the computation of the difference of transfer functions to extract BAOs is the most direct approach. Therefore, FITEX does not *divide* the observed *power spectrum* by a *theoretical* non-oscillating reference power spectrum but *subtracts* a *phenomenological* non-oscillating fitting function from the observed boosted *transfer function*, to extract the BAOs. Formally FITDEX can be described as follows:

$$\sqrt{\frac{P(k, z, b,)}{k}} - F(k, b, z) = \sqrt{A(k, z, b)} \frac{\Omega_{\mathrm{b}}}{\Omega_{\mathrm{m}}} T_{\mathrm{w}}(k) \,. \tag{2.6}$$

We found that the formula

$$F(k) \; = \; \frac{A}{1 \, + \, B \, k^{\delta}} \, e^{(k \, / \, k_1)^{\alpha}} \tag{2.7}$$

is able to fit all non-oscillating distortions to a satisfactory degree and leaves enough free parameters to allow for a wide range of transfer functions. Eq. (2.7) has no oscillatory components and can thus only trace the shape of a transfer function. Most of the drawbacks of the reference power spectrum division can be avoided:

(1) No cosmology or dark matter model has to be assumed to extract the BAOs. (2) The formula is flexible enough to compensate for various non-oscillating distortions without having to care about the source or physics of the distortions (see Sect. 2.4 - 2.6). (3) The formula is able to fit power spectra reliably on the sub percent level (see Sect. 2.4). (4) It is possible to provide a theoretical template function that can be used without assuming knowledge about dark matter and various amplitude effects by leaving the amplitude as a free parameter, as phase information and amplitude information is disentangled. The amplitude could be used for a further cosmological test.

To test FITEX, i.e., the combination of both the fitting function $F(k)$ and the wiggle function $T_w(k)$ under linear conditions, we first computed normalized transfer functions with CMBfast with the parameter values $w = -1$ (the transfer functions are affected by w only on scales above several Gpc, Ma et al. (1999)), $\Omega_{\mathrm{m}} = 0.3$, $\Omega_{\mathrm{b}} = 0.04$, $h = 0.70$, the primordial slope $n = 1$, the average CMB temperature $T_{\mathrm{CMB}} = 2.728 \, \mathrm{K}$, the He mass fraction after primordial nucleosynthesis $Y_{\mathrm{He}} = 0.24$, and the number of neutrino families $N_{\nu} = 3.04$. In the second step, the phenomenological continuum function (Eq. 2.7) was fitted to the multi-component transfer functions in the range $0.01 < k < 0.3h \, \mathrm{Mpc}^{-1}$. The results were subtracted from the original transfer functions to extract the BAOs. One example with $\Omega_{\mathrm{m}} = 0.3$ is shown in Fig. 2.2. In the first panel, the multi-component transfer function is plotted (solid line) as well as the phenomenological fitting function

(dashed line). In the second and third panels, the differences between the transfer function and the continuum fit are plotted (solid lines), together with the theoretical wiggle functions (dashed lines) computed with Eq. (2.1).

With typical baryon fractions, $f_b \sim 0.15$, baryonic oscillations are hardly visible as they make up only $\sim 2\%$ of the multi-component transfer function. This illustrates the main observational challenge in using wiggles as a metric ruler for cosmological investigations. On small scales, $k > 0.15h\mathrm{Mpc}^{-1}$, Fig. 2.2 shows that the theoretical wiggle function describes the baryonic oscillations to a satisfactory degree. This is expected, as the transfer function is dominated by the effects of velocity overshoot and Silk damping on these scales and the wiggle function was designed to model these effects. On large scales, $k < 0.05\,h\mathrm{Mpc}^{-1}$, the BAOs have to cope with competing CDM effects that dominate in these k-ranges. As the influence of CDM increases, the transfer function can not be described by baryonic physics alone and further corrections and assumptions have to be applied. Two related effects are:

(1) Turnover-effect: The turnover in the matter power spectrum occurs on the scale k_{eq} of the particle horizon at z_{eq}, which coincides with the approximate location of the first wiggle. No observation has hitherto clearly revealed this turnover in spectral power. The main problem in identifying the turnover is mostly due to the fact, that it is located at large scales where precise measurements of spectral power are difficult: the surveys cover too small volumes with slice-like shapes leading to strong smoothing (up to factors of two for the 2dF survey, Percival et al. (2001a)) and significant leakage in the derived power spectra. Under such conditions it appears quite difficult if not impossible to discriminate the first wiggle from the turnover when even the position of the turnover could not yet been determined to some accuracy. To do this, an excellent signal to noise ratio would be needed as well as precise knowledge of the underlying linear theory matter power spectrum. This includes information of distortions of the power spectrum to the sub-percent level. As a consequence, a model-independent approach like separating the BAOs from the multi-component transfer function with a phenomenological fitting function will only be able to detect the first wiggle when the baryon fraction is about as high as 40 percent or more, as we found with CMBfast simulations. However, we will show that our cosmological test is most sensitive to changes of w on scales around $k = 0.1h\mathrm{Mpc}^{-1}$. Therefore, the inability of a model-independent approach to detect the first BAO does not matter much as this first wiggle barely contributes any information to the cosmological test, even more as the amplitude of the first wiggle is comparatively low as it is dampened by CDM effects (see Sect. 2.7).

(2) Phase-shift effects: As mentioned in Sect. 2.2, the increasing effect of dark matter on large scales introduces a k-dependent phase-shift of the BAOs. The theoretical wiggle function (Eq. 2.1) already includes a phenomenological correction for this effect (replacing s by $\tilde{s}$). A similar effect on the phases, but in the opposite direction, is introduced by the damping of baryons on structure growth. In fact, the main effect of baryons on the multi-component power spectrum is, assuming a standard cosmology, a sharp break in power starting on scales smaller than the turnover. Formally, this break is caused by the rapid decay of the $\tilde{T}_b$ component of the transfer function (see Fig. 2.1). As the break occurs right after the turnover, it is very hard to distinguish shape components from oscillations.
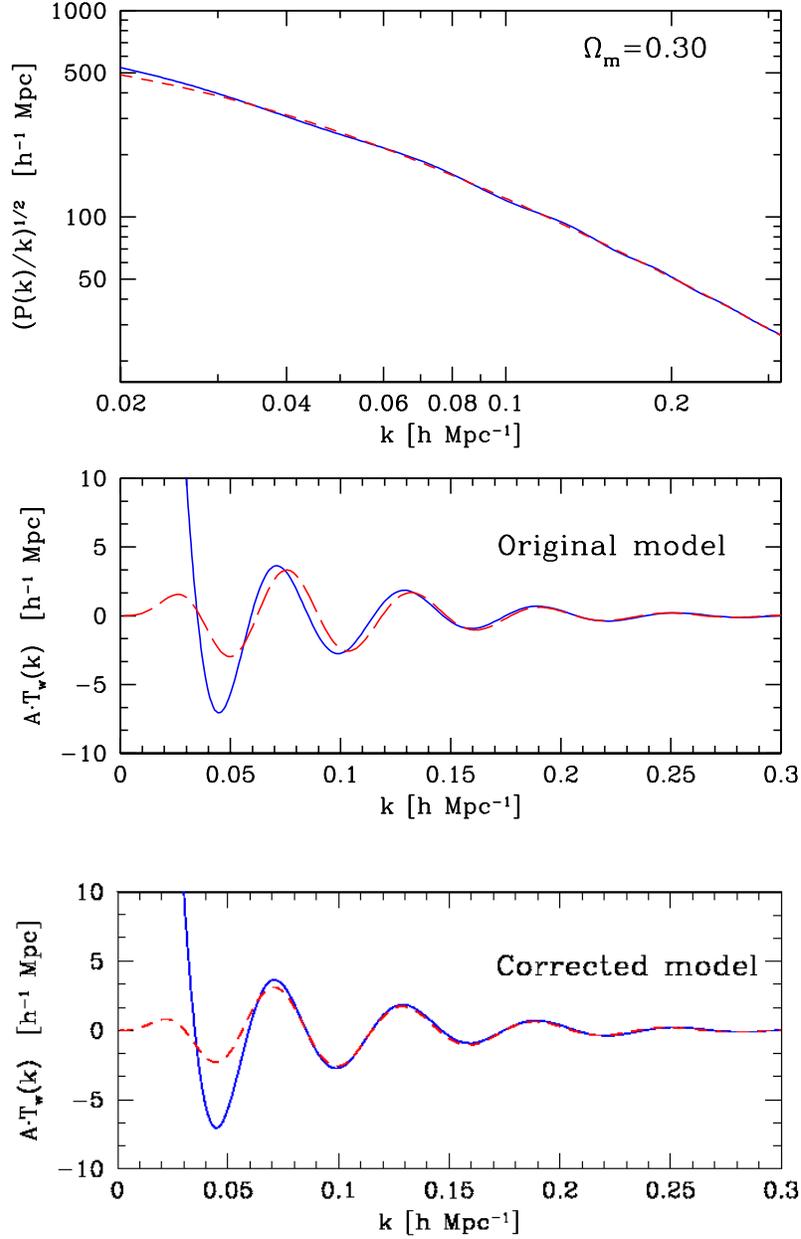
Figure 2.2: Performance of the phenomenological fitting function under linear conditions. The upper panel shows the CMB fast transfer function (solid line) and the best non-oscillating fit (dashed line). In the lower panels, the residuals of the subtraction are plotted (solid line) as well as the theoretical transfer function (dashed line). The original model gives the performance of the original theoretical function (Eq. 2.1) while the corrected model uses the modified function (Eq. 2.8)

For the extraction of BAOs in that scale range we have found that this spectral break must be modeled in detail (with an accuracy on the sub-percent level) or artificial phase-shifts with sizes of the order of interesting w effects are introduced when naive fitting functions ignoring this baryonic backreaction are subtracted from the multi-component power spectrum. Thus the model independent fitting function presented in this work (Eq. 2.7), but also traditional wiggle-free dark matter transfer functions with a slightly different functional approach like those given in Bardeen et al. (1986) or Efstathiou et al. (1992), introduce artificial phase-shifts of the generic form

$$\left(\frac{\hat{s}}{\tilde{s}}\right)^{3/2} \;=\; 0.98 \;+\; \left[\frac{5.1\,(\Omega_{\mathrm{m}}\,h^2)^{0.47}}{k\,s}\right]^{3/2} \;,\qquad (2.8)$$

which gives a useful summary of the phase-shifts (about 0.7% at $60h^{-1}$Mpc and 4% at $120h^{-1}$Mpc) even for very different baryon fractions. This equation was derived phenomenologically from fits to CMBfast simulations. The second panel in Fig. 2.2 shows an example for the concordance cosmology. It is seen that the phase-shifts are most prominent at intermediate scales, $k < 0.15\,h\,\mathrm{Mpc}^{-1}$, and counteracts the phase-shifts introduced by large-scale CDM potentials. The dashed line in the fourth panel shows the corrected theoretical wiggle function, replacing $\tilde{s}$ in Eq. (2.1) by $\hat{s}$ from Eq. (2.8). It is clearly seen that the extracted wiggle function follows the predicted theoretical, and phenomenologically corrected, wiggle function on the 0.5%-level in the range $0.05 < k < 0.3\,h\,\mathrm{Mpc}^{-1}$ which is relevant for a cosmological test of the parameter w. In the following, Eqs. (2.1) and (2.8) constitute our final theoretical wiggle function which we use in FITEX to match the BAOs extracted from the simulated data.

2.4 Extracting BAOs under Quasi-Nonlinear Conditions

The simulated data used to test FITEX for BAO extraction under more realistic conditions are provided by the ΛCDM version of the Hubble Volume Simulations conducted by the Virgo Consortium (Evrard et al., 2002). One billion dark matter particles were simulated with $h = 0.7$, $\Omega_{\mathrm{m}} = 0.3$, $\Omega_{\Lambda} = 0.7$, $\Omega_{\mathrm{b}} = 0.04$, $\Omega_{\nu} = 0$, and the normalization of the matter power spectrum, $\sigma_8 = 0.9$, in a cube with a comoving length of $L = 3000\,h^{-1}\,\mathrm{Mpc}$ and a mass of $2.25 \cdot 10^{12}\,h^{-1}\,\mathcal{M}_{\odot}$ per particle. The simulations were started with a 'glass-like' load (see Baugh et al., 1995). During the simulation itself, long range gravitational forces were computed on a 1024^3 grid yielding a Nyquist critical frequency of $k_c = 1.07\,h\,\mathrm{Mpc}^{-1}$. The short range gravitational forces were computed via direct summation and softened on a scale of $0.1\,h^{-1}\,\mathrm{Mpc}$. We used the $10 \times 10\,\mathrm{deg}^2$ fraction of the XW extended deep wedge which uses periodic boundary conditions from a redshift of $z = 4.4$ to provide a survey lightcone up to the redshift $z_{\mathrm{max}} = 6.8$. The lightcone output comprises of data that includes cluster evolution and thus mimics real life observations. Larger distance from the observer correspond to higher redshifts where structures are less pronounced due to linear and non-linear growth.
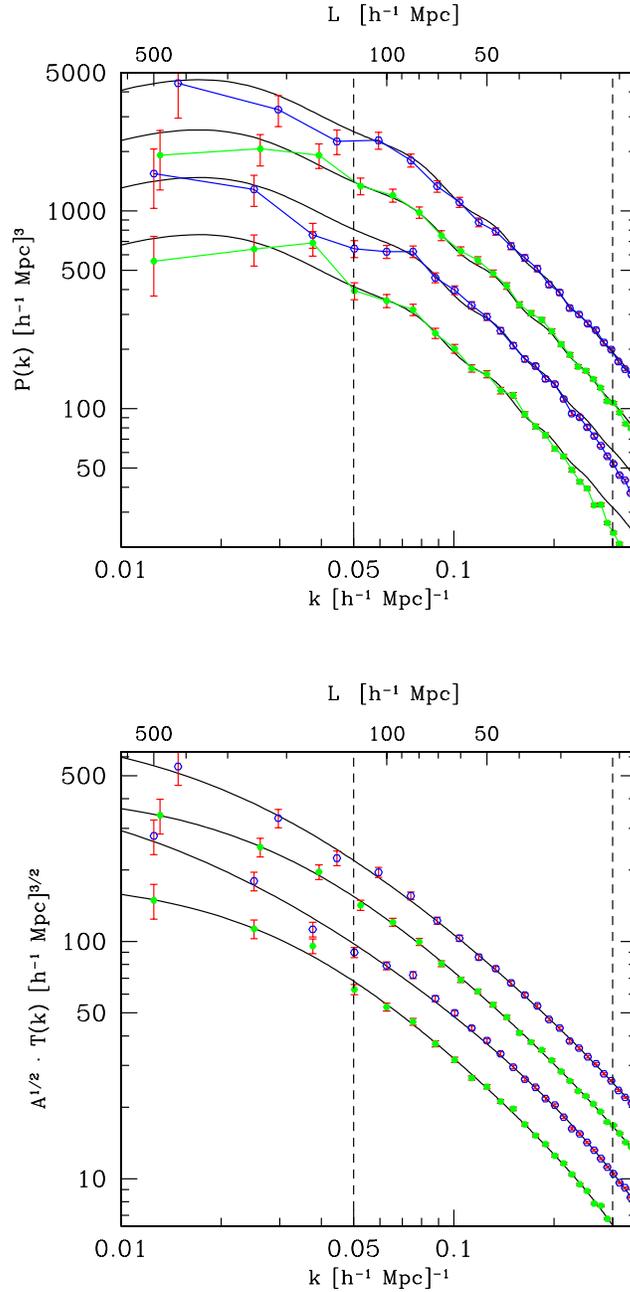
Figure 2.3: **Upper panel:** Power spectra of the matter distribution from the cubes 7 (upper spectrum) to 10 (lower spectrum) of the XW deep wedge N-body simulation with superposed linear theory predictions. The linear matter power spectrum was taken from Eisenstein & Hu (1998). Amplitude corrections for linear structure growth were implemented in the standard manner. For better illustration, the 2nd, 3rd and 4th spectrum is shifted downwards by the factors dex(0.1), dex(0.2), dex(0.3), respectively. The extraction of BAOs described in the main text is restricted to the k-range bordered by the two vertical dashed lines. **Lower panel:** Same as upper panel for the corresponding boosted transfer functions with superposed fits of the phenomenological fitting functions. The fit at $z = 2.3$ uses the following parameters: $A = 540.0$, $B = 35.45$, $k_1 = -0.32354$, $\alpha = 2.493$, $\delta = 1.420$.

Table 2.1: Parameters of the cubes for the power spectrum estimation. Col. 1: Cube number. Col. 2: Comoving length of the cube (concordance cosmology). Col. 3: Redshift range covered by the cube. Col. 4: Fundamental mode of the discrete Fourier transformation which corresponds to the sample bin size of the power spectrum in k space. Col. 5: Nyquist critical wavenumber of the Fast Fourier Transformation.

Sample	length $[h^{-1}\,\mathrm{Mpc}]$	z-range	Δk $[h\,\mathrm{Mpc}^{-1}]$	k_c $[h\,\mathrm{Mpc}^{-1}]$
cube1	200	$0.58 - 0.68$	0.031	4.0
cube2	226	$0.68 - 0.79$	0.028	3.6
cube3	256	$0.79 - 0.92$	0.025	3.1
cube4	291	$0.92 - 1.1$	0.022	2.8
cube5	329	$1.1 - 1.3$	0.019	2.4
cube6	373	$1.3 - 1.6$	0.017	2.2
cube7	423	$1.6 - 2.0$	0.015	1.9
cube8	479	$2.0 - 2.5$	0.013	1.7
cube9	500	$2.5 - 3.2$	0.013	1.6
cube10	500	$3.2 - 4.1$	0.013	1.6
cube11	500	$4.1 - 5.4$	0.013	1.6

Power spectra, $P_{\mathrm{obs}}(k, z, b)$, were estimated at different redshifts and for different galaxy biasing schemes b (see below) with the Fast Fourier Transform using the variance-optimized method of Feldman et al. (1994) for cubes along the line-of-sight (LOS) of the XW wedge. All the cubes were selected to fully fit into the lightcone in order to not introduce a non-trivial window function. The parameters of the cubes are given in Tab. 2.1. Note that the spectroscopic SDSS survey of normal galaxies covers an effective volume which is comparable to cubes with $L \approx 500\,h^{-1}\,\mathrm{Mpc}$. The errors of the power spectral densities at wavenumber k were estimated with standard mode counting arguments by $\sigma_P/P(k) = \sqrt{2\pi}\{1 + 1/[P(k)\bar{n}]\}/kL$ to provide a first estimate. Here, $\bar{n}$ is the mean comoving particle number density in the box, and L the length of the box. We found that the initial glass load reduces the shotnoise level significantly below the standard Poisson case at redshifts $z > 2$ (see Smith et al., 2003). Instead of correcting our spectral estimator for this artifact (see, e.g. Smith et al., 2003), we could in the end neglect its impact on the cosmological test, because our phenomenological fitting function was flexible enough to compensate for the small residuals of this artifact seen in our standard scale range $k < 0.3\,h\,\mathrm{Mpc}^{-1}$.

In order to extract the BAOs from the power spectra $P_{\mathrm{obs}}(k, z, b)$, we computed the corresponding boosted transfer functions first, $\sqrt{A(k, z, b)}\,T(k) = \sqrt{P_{\mathrm{obs}}(k, z, b)/k}$, assuming a primordial spectrum with $n = 1$ (see below). Fig. 2.3 shows the estimated power spectra and the corresponding boosted transfer functions for an unbiased dark matter particle

distribution in configuration space. Error bars include shot noise and sample variance. In most spectra the 2nd and 3rd BAOs are clearly visible. For $k \geq 0.05\,h\,\mathrm{Mpc}^{-1}$, the phenomenological fitting function (Eq. 2.7) can be computed very accurately as it is constrained by a large number of spectral data with relatively small errors, that can be traced reasonably well by a power law. Sample variance is negligible, as in this k-range, each spectral point represents the average over a large number of modes. For $k < 0.05\,h\,\mathrm{Mpc}^{-1}$, only a very small number of spectral densities, each estimated with a small number of independent modes, exists and the phenomenological function has to fit a wiggle that coincides with the turnover. Therefore, the results in the range $0 < k < 0.05\,h\,\mathrm{Mpc}^{-1}$ vary quite a bit. However, we already excluded this range from the cosmological test (see vertical dashed lines in Fig. 2.3 and the discussion at the end of Sect. 2.3).

The extracted BAOs for the four cubes 7-10 of the XW deep wedge simulation corresponding to the redshifts $z = 1.8\text{-}3.7$ are shown in Fig. 2.4. Spectral densities in the range $0.05 < k < 0.3\,h\,\mathrm{Mpc}^{-1}$ which is relevant for unbiased cosmological tests are plotted (dots) together with the theoretical wiggle function (Eqs. 2.1, 2.8) multiplied by an appropriate amplitude $\sqrt{A}$ (solid line). We see that even for very large galaxy samples filling a box-like survey volume with a length of $L = 500\,h^{-1}\,\mathrm{Mpc}$ error bars are mainly determined by sample variance and are of the same order as the amplitudes of the BAOs. Therefore, only surveys with significantly larger survey volumes can show the BAOs with higher significance.

We further see that the amplitudes of the wiggle functions in Fig. 2.4 decrease with increasing redshift, due to growth suppression (we verified the amplitude change with z according to standard linear structure growth). This is a direct consequence of the fact, that distortions of the power spectrum are directly translated into the boosted transfer function and thus into distortions of the amplitude of the BAOs. To see how this works and how deviations from the assumed primordial spectrum with $n = 1$ affect the wiggle function, we rewrite the primordial spectrum in the form $k^{1+2\epsilon}$ with $|\epsilon|$ small compared to unity. For the assumed power spectrum, $P(k, z, b) = A(k, z, b)k^n T^2(k)$, this choice gives

$$\sqrt{\frac{P(k, z, b)}{k}} - F(k, b, z) = \sqrt{A(k, z, b)}k^\epsilon \frac{\Omega_{\mathrm{b}}}{\Omega_{\mathrm{m}}} T_{\mathrm{w}}(k)\,, \tag{2.9}$$

where the phenomenological function (Eq. 2.7) effectively fits the amplitude and shape-dependencies (see Eqs. 2.2, 2.3)

$$F(k, z, b) = \sqrt{A(k, b, z)}k^\epsilon \left[\frac{\Omega_{\mathrm{c}}}{\Omega_{\mathrm{m}}} T_{\mathrm{c}}(k) + \frac{\Omega_{\mathrm{b}}}{\Omega_{\mathrm{m}}} \tilde{T}_{\mathrm{b}}(k)\right]\,. \tag{2.10}$$

It is seen that the phases of the wiggle functions remain unchanged by the FITEX procedure. This result is shown in Fig. 2.5 where the extracted oscillations from power spectra with different spectral index models in the primordial power spectra are plotted. The oscillations were retrieved by calculating a power spectrum with a spectral index $n = 1$ (solid line), an index of $n = 0.9$ (long dashed line) and a simple running spectral index model $k^{1-0.1 \ln k}$ (short dashed line) and applying the FITDEX method. The parameters
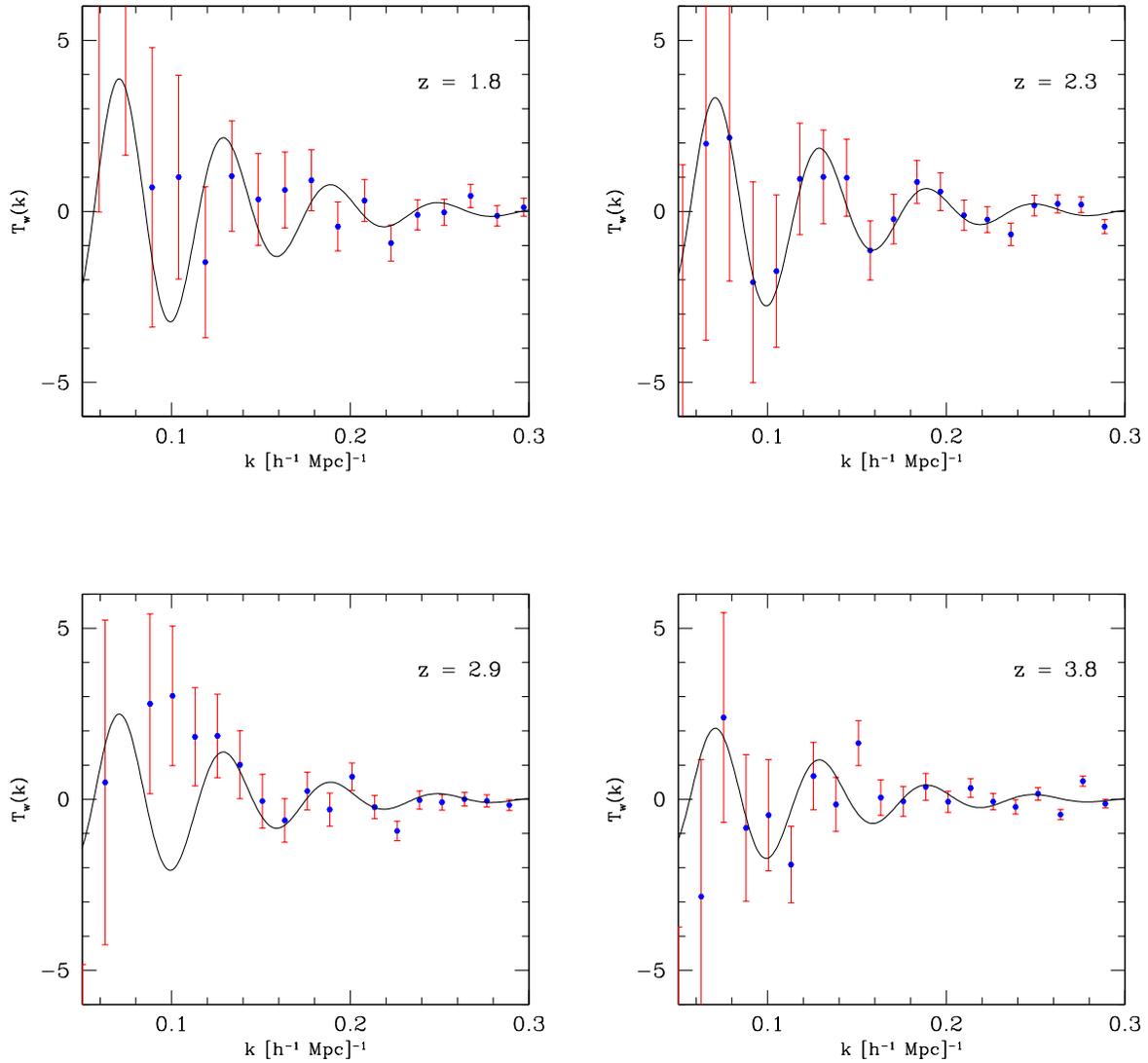
Figure 2.4: Extracted baryonic oscillations of dark matter particles (dots) with error bars including shot noise and sample variance at different redshifts ranging from $z = 1.8$ (upper left panel) to $z = 3.8$ (lower right panel). Superimposed is the theoretical wiggle function (Eq. 2.1) with an appropriate amplitude.

are chosen to conservatively cover the range of primordial power spectra not excluded by WMAP 3rd year data (see Bridges et al., 2006).
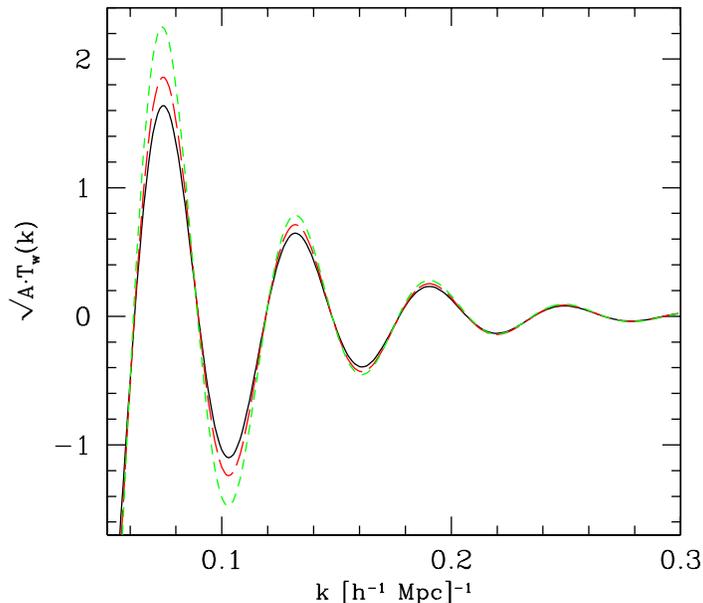
Figure 2.5: Wiggle extraction from power spectra with different primordial k dependencies. The resulting oscillations using FITDEX on primordial models with a spectral index $n = 1$ (solid line) is compared to a model with different spectral index $n = 0.9$ (long dashed line) and a running spectral index with $\alpha = 0.1$ (short dashed line).

Only the amplitudes of the wiggle functions are modulated by possibly scale-dependent factors. Although a lot of interesting information about galaxy formation and cosmology is contained in the amplitude factors of the wiggle function, we skip this detail here and perform an amplitude rescaling, $\sqrt{A(k,z,b)}k^\epsilon \Omega_{\rm b}/\Omega_{\rm m} \to \sqrt{A(k,b,z)}$, with $A(k,z,b)$ being a basically arbitrary function after rescaling, summarizing distortions like structure growth, galaxy biasing, redshift space effects, deviations from the Harrison-Zel'dovich case etc. Thus, with FITEX we are not deriving a pure undistorted wiggle function, but a spectrum which is subject to all the distortions that work on the power spectrum.

While the oscillations extracted from the Hubble Volume Simulations are consistent with the theoretical template, the error bars prevent a precise evaluation of the accuracy of FITEX (all 4 cubes will be combined later, see Fig. 2.15). To get a better estimate of non-linearities that might wash out wiggles at higher k values, we use a non-linear analytic form of the power spectrum. This was computed by Jeong & Komatsu (2006) using 3rd-order perturbation theory, which is able to model non-linear N-body matter power spectra to better than 1% at $z > 2$ and $k < 0.3\,h\,{\rm Mpc}^{-1}$.
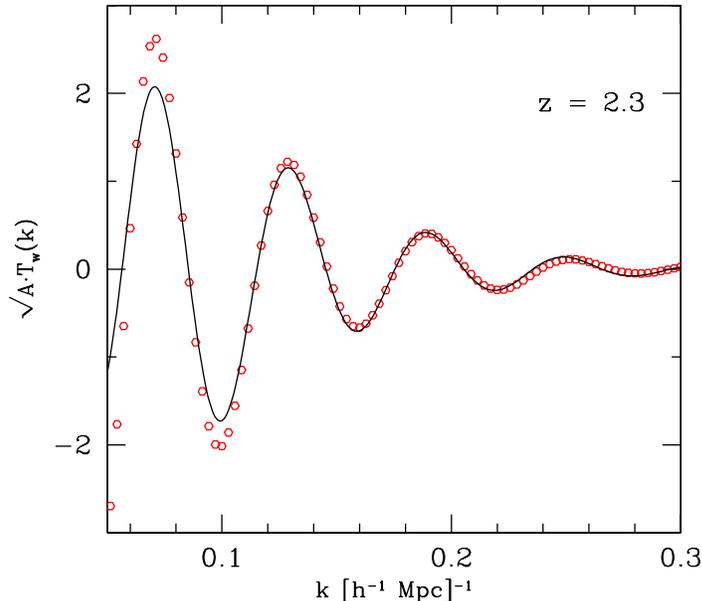
Figure 2.6: BAO extracted from an analytic power spectrum of Jeong & Komatsu (2006) that includes non-linear growth at $z = 2.3$ (dots). The continuous line is the theoretical wiggle function derived in Sect. 2.2, multiplied by an appropriate amplitude. See the main text for a more detailed discussion of the treatment of the amplitudes of the wiggle function and the analytic power spectrum.

Figure 2.6 shows the results of the FITEX extraction applied to the analytic non-linear power spectrum using Hubble Volume parameters. As expected, the amplitude deviates from the theoretical prediction at small k-values, where the shape contribution of the baryons dominates (see. Fig. 2.1), but the phase of the oscillations in the analytical non-linear function is recovered to better than 1% accuracy. This clearly shows, that, at least in theory, FITEX is able to model non-linear effects and recover the oscillations with high accuracy.

As an additional test, the power spectrum of the Hubble Volume Simulation (OCTANT data), a wedge covering on eighth of the sky from $z = 0 - 1.3$ including redshift-space effects, was computed with about 60 million particles in a fast Fourier box with a legnth of $1600 \, h^{-1} \, \mathrm{Mpc}$ and the FITDEX algorithm used to extract the oscillations (Fig. 2.7). The oscillations can be seen, although non-linear growth is heavily washing out the wiggles at $k > 0.22 \, h \, \mathrm{Mpc}^{-1}$.

Furthermore, we see a very good correspondence between a FITEX-extracted wiggle function obtained from N-body simulations as well as the analytic non-linear power spectrum and the theoretical wiggle function of FITEX (Eq. 2.1) in the HETDEX redshift
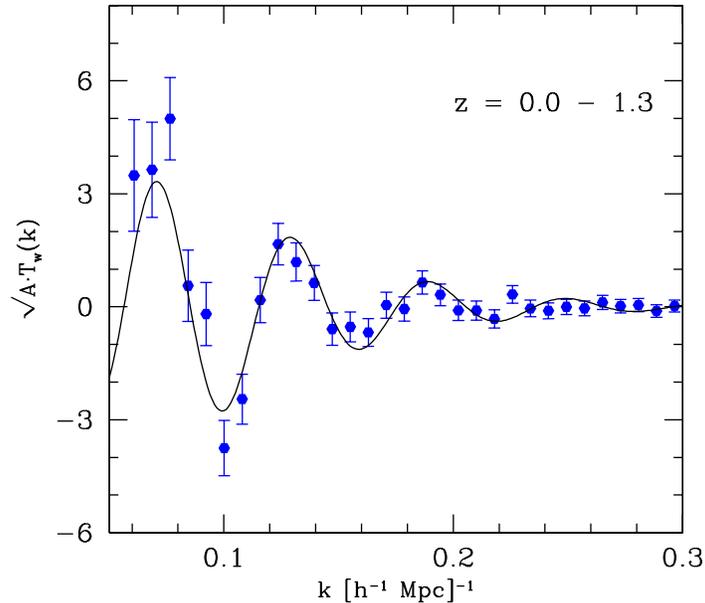
Figure 2.7: Extracted BAOs using the Hubble Volume Simulation covering one eighth of the sky from redshift 0 to 1.3. Error bars include sample variance and shotnoise but do not cover washing out effects introduced by the correlation of modes through non-linear growth and distant observer approximation.

range. Both linear and non-linear structure growth change the power spectrum and thus the BAOs, but FITEX still gives quite stable and easy to analyze results for the 2nd to 5th BAO. The presence of the BAOs is consistent with the results of recent simulations of e.g. Springel et al. (2005) and (Seo & Eisenstein, 2005). However, no specific method to extract the BAOs was applied and their simulations did not include lightcone effects. In the following we concentrate on the effects of redshift space and scale-dependent galaxy biasing on the results obtained with FITEX.

2.5   Extracting BAOs in Redshift Space

The main problem in relating redshift space coordinates with comoving coordinates is the differentiation between red-shifting due to the expansion of the Universe and red-shifting due to peculiar velocities of the measured particles. Two effects might be important.

(1) In the linear regime and in combination with the distant observer approximation, peculiar velocities introduce a boosting factor β along the LOS. Only the k-modes parallel to the LOS, $k_{\parallel}$, are boosted by the square of the factor $1 + \beta(z,b)k_{\parallel}^2/(k_{\parallel}^2 + k_{\perp}^2)$ with

$\beta(z, b) = -d \ln D(z)/[d \ln(1 + z)b(k, z)]$, $D(z)$ the linear structure growth, and $b(k, z)$ the biasing parameter (Kaiser 1987).

(2) In the non-linear regime, velocity dispersion of galaxies in (partially) virialized structures cause the well-known 'fingers of God' effect in redshift space. We assume that the effect is mainly restricted to small scales and is describable by a simple exponential damping of the power spectrum by the factor $\exp(-\alpha k_\parallel^2)$. Here, the redshift-dependent parameter α gives a measure of the average pairwise velocity dispersion in the given redshift shell.
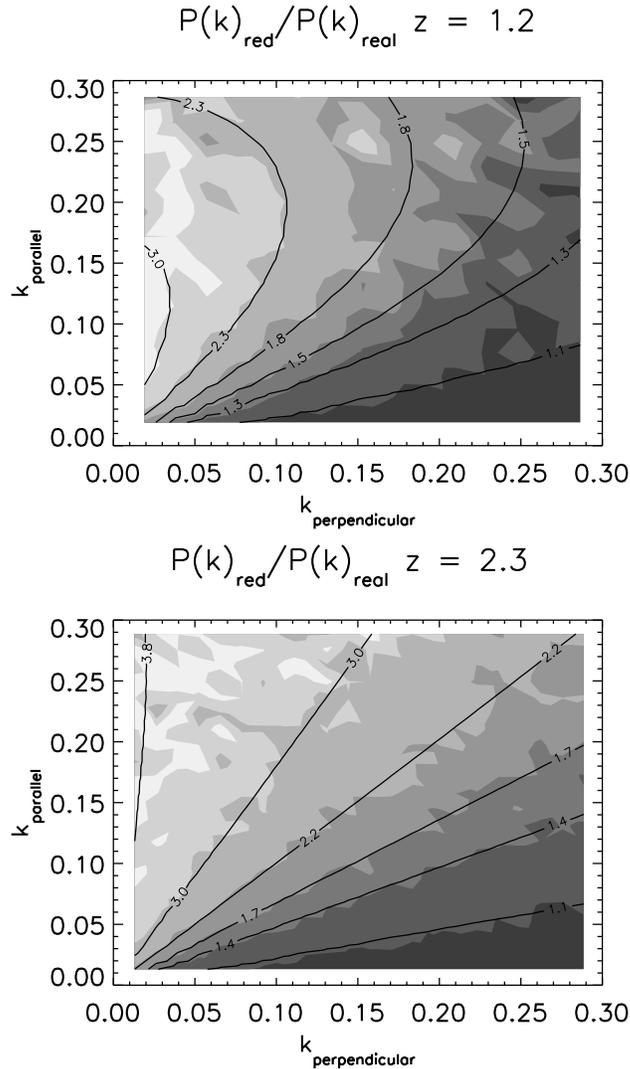
Figure 2.8: Redshift space spectra divided by their corresponding real space spectra at different redshifts (filled areas). Overplotted lines represent the best fit of the kernel of Eq. (2.11).

The combination of (1) and (2) constitute one version of the standard dispersion model of peculiar velocities which relates the one-dimensional redshift and real space power spec-

tra by

$$P_s(k, z, b) = P(k, z, b) \int_0^1 d\mu \left[1 + \beta(z, b)\mu^2 \right]^2 e^{-\alpha(k\mu)^2} , \qquad (2.11)$$

where μ is the cosine between the direction of the actual wavevector and the LOS, i.e., $\mu k = k_\parallel$, and $k^2 = k_\parallel^2 + k_\perp^2$ (for more detailed models see Scoccimarro 2004). On large scales, the real space power and thus the wiggle function is boosted in a k-independent manner (Kaiser factor). On small scales, the power spectrum is dampened in a k-dependent manner by the factor $0.5\sqrt{\pi} \, \mathrm{erf}(\alpha k)/\alpha k$, with erf the standard error function of statistics. In Schuecker et al. (1996) it is shown that Gaussian redshift errors can be described with the same model (for $\beta = 0$). Eq. (2.11) suggests that only the amplitudes but not the phases of the BAOs can be changed by this type of redshift space distortion.
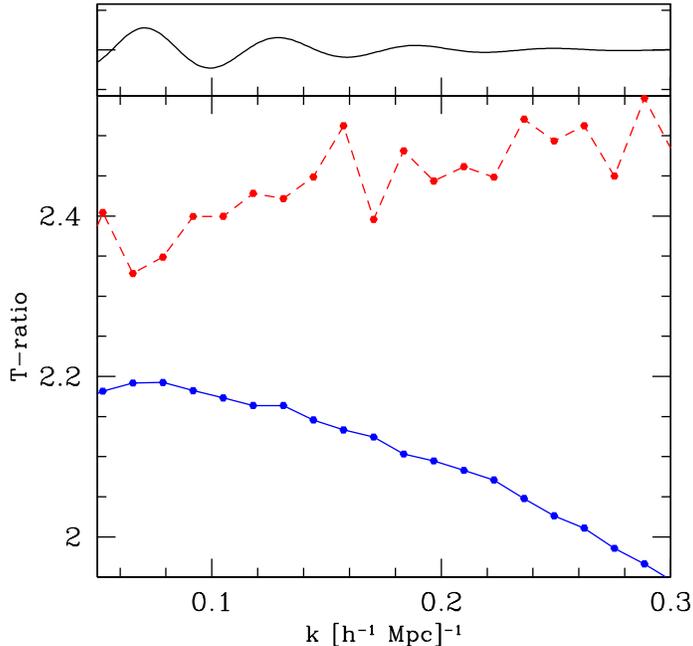
Figure 2.9: Ratios of boosted redshift and real-space transfer functions for different redshifts. As expected, the redshift space effects increase with decreasing redshift in a smooth manner. Increasing ratios at high redshifts are an artifact caused by the glass-like initial load of the N-body simulations. To illustrate the relation between redshift space effects and amplitudes of BAOs, the ratio between BAOs and the transfer function is plotted in the upper panel.

Figure 2.8 illustrates the combined effect in more detail. Two redshift space power spectra, as obtained from the Hubble Volume Simulation, were divided by their corresponding real space spectra at different redshifts and plotted as filled areas. The kernel of Eq. (2.11) was fitted to this ratio and the best fit results superposed (black solid lines). We found that the model is accurate on the 5%-level over the complete redshift range. For $z > 2$,
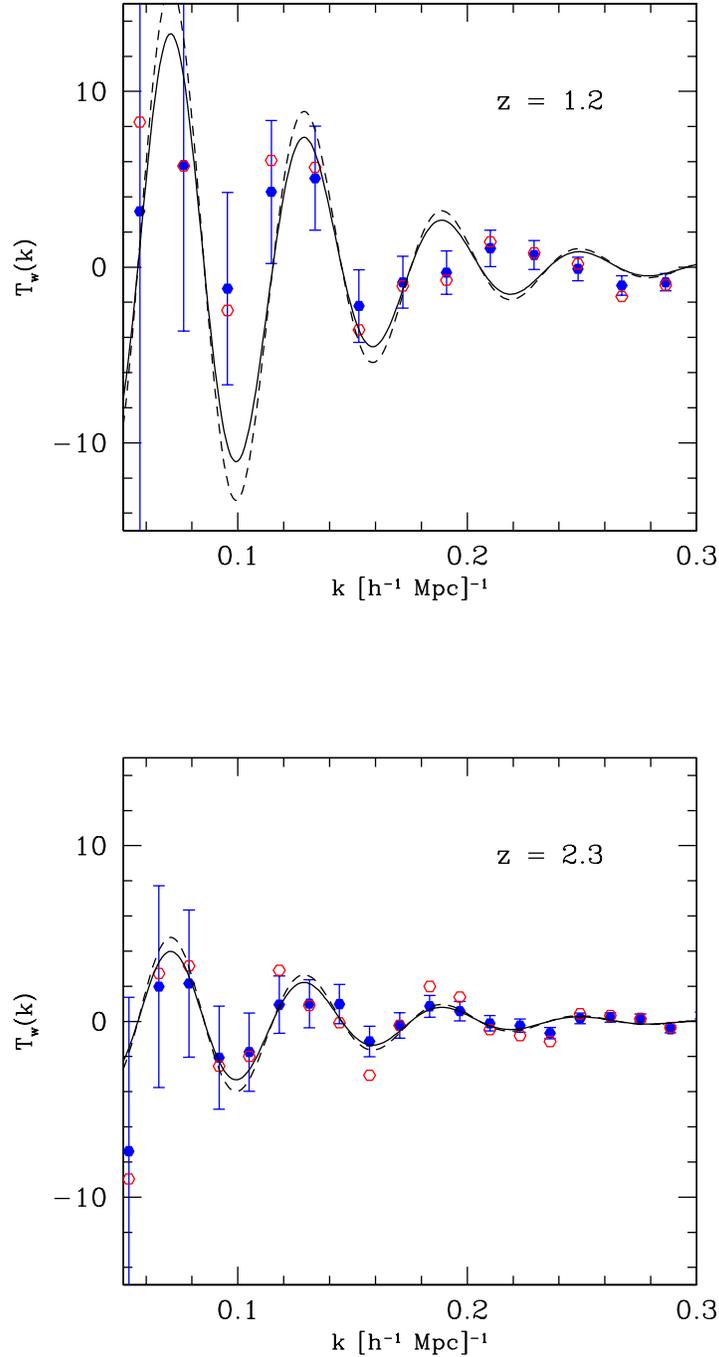
Figure 2.10: Comparison of wiggle functions at $z = 1.2$ (upper panel) and $z = 2.3$ (lower panel) obtained from real space (filled circles and continuous lines) and redshift space (open circles and dashed lines). Note the overall boost of the wiggle functions by redshift space effects which, however, leaves the phases of the BAOs invariant. The non-linear growth effects, start to wash out the wiggles at $k > 0.2 \, h \, \mathrm{Mpc}^{-1}$ in the upper panel.

the parameter α is virtually zero because no virialization had time to take place on the scales shown. As redshift decreases, galaxy clusters start to form and virialize and the α parameter starts to distort not the amplitude in the two dimensional power spectrum depending on the angular component, like β does, but changes the amplitude depending on the modulus of the k vector. This means that the shape of the wiggles could be influenced by this effect as the amplitude of the power spectrum and thus the amplitude of the wiggles is dampened at high k-values.

Fortunately, this anisotropy only gradually changes the amplitude of each consecutive mode over which is averaged in a k-shell from perpendicular to parallel modes. This is illustrated in Fig. 2.9 where the ratios of boosted transfer functions are plotted. A rough idea about the relation between redshift space effects and the amplitudes of BAOs is provided by the ratio between BAO and the transfer function plotted in the upper part of Fig. 2.9.

Two examples of wiggle functions boosted by redshift-space effects and extracted with FITEX are plotted in Fig. 2.10. The result of the anisotropic redshift space is that modes with a more parallel component get more weight than those with a more perpendicular component, which are not boosted. Whereas the overall shape of the wiggle function appears to be slightly changed, the phases of the wiggles are not distorted at all. Instead, the whole wiggle function receives a boost of the monopole component of the Kaiser factor.

The fitting function is obviously able to model the redshift space distortions on the sub-percent level and oscillations can be derived. If the wiggles were not extracted with FITEX-like methods but by dividing the observed power spectrum by a standard non-oscillatory linear theory power spectrum, as was done by in Springel et al. (2005) and Cole et al. (2005), the baseline of the oscillations would start to depart from zero at smaller scales. Such an oscillation spectrum could not easily be used in a cosmological test as it fit only nontrivial theoretical model spectra.

2.6   Extracting BAOs from Biased Samples

In the Hubble Volume, CDM particles are simulated. What is actually observed are galaxies made out of dark matter and baryonic matter, especially in the case of the proposed HETDEX survey, where Lyman-α emitting galaxies are used as tracers of the underlying matter distribution. Because of the different properties of dark matter and baryonic matter (Benson et al., 2000; Delliou et al., 2006), both need not necessarily follow exactly the same density distribution.

To test the stability of FITEX under such conditions, we introduce some sort of biasing into the N-body simulations. Formally, biasing is described by the mean biasing function (Dekel & Lahav, 1999),

$$b(\delta_d)\,\delta_d \;=\; \langle \delta_g \,|\, \delta_d \rangle \;=\; \int d\delta_g \, P(\delta_g \,|\, \delta_d)\, \delta_g \;, \tag{2.12}$$

that relates the dark matter distribution δ_d to the galaxy distribution δ_g (Somerville et al., 2001). Here, $P(\delta_g | \delta_d)$ is the local conditional biasing distribution, i.e. the probability that

a certain matter density contrast δ_d with a variance σ_d^2 is converted into a galaxy contrast δ_g with a variance σ_g^2. The biasing function can also be expanded into the moments, $\hat{b} = \langle b(\delta_d)\,\delta_d^2 \rangle/\sigma_d^2$, and $\tilde{b}^2 = \langle b^2(\delta_d)\delta_d^2 \rangle/\sigma_d^2$, with the biasing scatter $\sigma_b^2 = \langle (\delta_d - \langle \delta_g|\delta_d \rangle)^2 \rangle/\sigma_d^2$. Any local, non-linear and stochastic biasing relation can be described by these moments to second order. While $\tilde{b}/\hat{b}$ is a measure for the non-linearity and $\sigma_b/\hat{b}$ a measure of the scatter, the parameter $\hat{b}$ characterizes the linear biasing such, that if $\hat{b} = \tilde{b}$, then $\delta_g(\vec{x}) = \hat{b}\delta_d(\vec{x})$ (Wild et al. 2005).
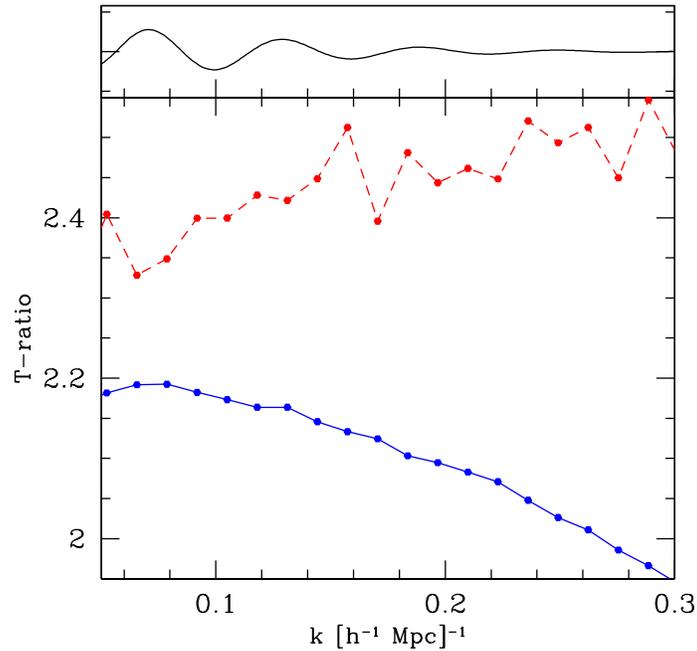
Figure 2.11: Ratios of biased and unbiased boosted transfer functions for two samples with different scale-dependent biasing at $z = 2.3$. A factor of 2 was subtracted from the strong bias (dashed line) for enhanced visibility, while the weak biasing (solid line) remains unchanged. The theoretical wiggle function is plotted on top of the figure to illustrate the scale of the biasing effects.

We use the simple stochastic and slightly non-linear Eulerian biasing scheme (e.g. Cole et al., 1998; Yoshida et al., 2001). First, local overdensities for each dark matter particle are computed by smoothing the distribution with a Gaussian, $\exp(-r^2/2r_0^2)$ and $r_0 = 3\,h^{-1}\,\mathrm{Mpc}$. The overdensities δ_d are then transformed into normalized overdensities, $\nu = \delta_d/\sigma_d$, with σ_d being the root mean square of the δ_d-values. Finally, a function is introduced that describes the probability of a dark matter particle with overdensity δ_d to be tagged as a 'galaxy', $P(\nu) \sim (\nu - \nu_c)^\alpha$, with ν_c being a threshold at which the probability is set to zero. This simulates the lower probability of galaxies forming in low density regions like voids. The parameter α controls the stochastic spread of the biasing function as well
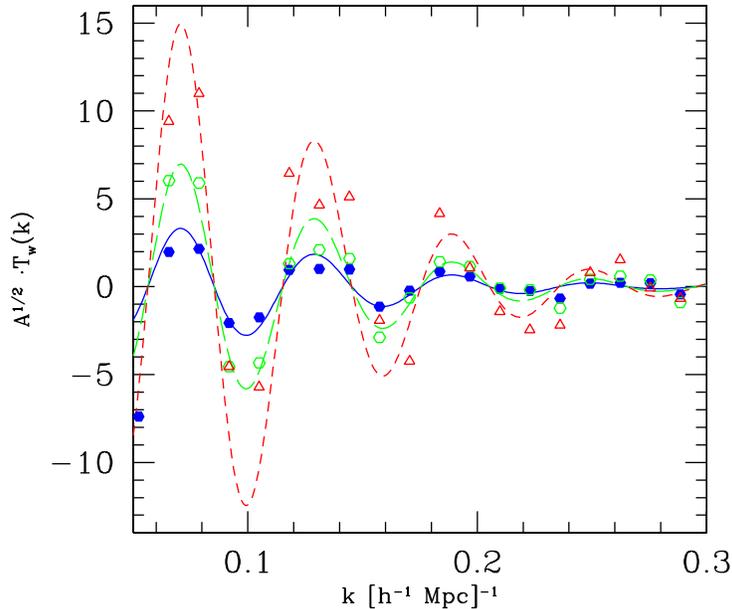
Figure 2.12: Boosted wiggle functions extracted with FITEX and theoretical wiggle functions for the unbiased sample (filled circles and continuous line), for the strongly biased sample (triangles and shot-dashed line) and for the less biased sample (open circles and long-dashed line). The theoretical wiggle functions are multiplied by the $\hat{b}$-moment of the biasing function.

as its non-linearity. A value of $\alpha = 0.2$ is a good compromise between reasonable non-linearity and a comparably tight spread. Two samples with $\nu_c = -0.2$ and 0.6 were generated. Both parameters, α and ν_c, were adjusted to model the biasing parameters and number densities found by observation. SUBARU data (Hamana et al., 2004) suggests a comoving LAE density of $\bar{n} = 0.00301 \pm 0.00194\,h^3\,\mathrm{Mpc}^{-3}$ and a biasing parameter of $\hat{b} = 4\text{-}5$ at a redshift $z = 5$, which should decrease slightly with redshift. As discussed in Sect. 2.4 the biasing function $b(\delta_d)$ should translate into a boost of the power spectrum $b^2(k, z)$. Because the average slope of the biasing function is $\hat{b}$, to first order we expect $P_{\mathrm{bias}}(k) = P(k)b^2(k, z) \sim P(k)\hat{b}^2(z)$.

The two prescriptions lead to a biasing that either decreases or increases with k (see Fig. 2.11). Both values are consistent with the calculated $\hat{b}$ parameter. The comparatively strong scale dependence of the biasing becomes evident when the ratios of the boosted transfer functions are compared with the ratio of the wiggle amplitude to the transfer function plotted on top of Fig. 2.11.

Figure 2.12 illustrates, that FITEX extracts BAOs even from biased galaxy distributions. Plotted are the wiggle functions of the less strongly biased sample (open circles), the more strongly biased sample (open triangles) and the unbiased sample (filled circles)

at $z = 2.3$. The unbiased and biased oscillations vary because of shot noise, as the biased samples have a 2 and 10 times lower point density, respectively. Superimposed are the theoretical wiggle functions (Eq. 2.1), normalized to the unbiased wiggle function (continuous line) and multiplied by the first moment of the biasing, $\hat{b}$ (dashed lines). The boost of the power spectrum obviously translates into an amplitude boost of the BAOs. As in redshift space, only the phenomenological fitting function as used in FITEX extracts the BAOs with a zero baseline, unlike when using a linear theory power spectrum.

However, not only the overall boost of the power spectrum, but also the k-dependency of the biasing function should directly affect the amplitude of the BAOs. We tried to verify this by fitting the $T_w(k)$ function supplemented by a 'linear biasing model' to describe the extracted wiggle function $T_{\text{ex}}(k) = (a_0 + a_1\,k)T_w(k)$. We notice that both parameters a_0 and a_1 are highly degenerate, mainly due to the fact that only data points at the extrema of the wiggle function significantly contribute to the fit. Data points at or near the nodes of the wiggle function are very insensitive to changes of the amplitude. Due to the degeneracy, any amplitude a_0 can easily be compensated by an appropriate value of the parameter a_1. Though the full fits had the expected slopes, the effects were very small and of no significance for the cosmological test. We may thus conclude that in a cosmological test, which is mostly about detecting phase-shifts, it seems to be enough to fit the theoretical wiggle function with only an amplitude parameter, when the k-dependent biasing is as weak as simulated here. However, more realistic simulations are needed to analyze higher order effects.

2.7 Cosmological Tests with BAOs

The cosmological test compares the phases of the extracted BAOs with the theoretical wiggle function, projected onto the hypersphere of the observer. As the theoretical template does not model the amplitude of the oscillations with high accuracy, we are not able to include amplitude effects in the cosmological test and marginalize over the wiggle amplitude. Future tests should include this parameter, which is sensitive to w by e.g. structure growth.

A general feature of the test is that one first has to assume a certain reference cosmology to get the metric scale of the wiggle function from the data, and to perform in a second step the comparison of the metric scale of the theoretical wiggle function relative to the chosen reference cosmology (Glazebrook & Blake, 2005). In the present paper, we are interested in the performance of the BAOs extracted with FITEX in constraining the redshift-independent part of the w parameter of the dark energy. The amplitude of the theoretical function is also fitted but later marginalized. The remaining cosmological parameters are fixed to their values given at the end of Sect. 2.1.

As metric scales, comoving distances of the BAOs parallel and perpendicular to the LOS are used. Recall that in the plane-parallel or distant observer approximation, comoving distances between two points with the same angular positions but different redshifts are $x_\parallel = \int_{z_1}^{z_2} \frac{dx}{dz} dz = \frac{c}{H_0} \int_{z_1}^{z_2} \frac{dz}{E(z)}$, where $H(z) = H_0 E(z)$ is the Hubble parameter at redshift z

and $E(z)$ the w-dependent transformation of the Hubble constant, H_0, from redshift zero to z (e.g. Peebles, 1993). To compute not distances between two points in space with different redshifts, but parallel scaling factors between the two sets of cosmologies denoted by primed and unprimed symbols, one has to take the limit $\Delta z = z_2 - z_1 \to 0$, which yields

$$ x'_\parallel = x_\parallel \frac{dx'}{dx} . \tag{2.13} $$

Comoving perpendicular separations represent distances two points with the same redshift but different angular positions are separated, $x_\perp = \theta \int_0^{z_2} \frac{dx}{dz} dz$, where θ is the angle between the two points. The scaling relation for the perpendicular component between two cosmologies thus becomes

$$ x'_\perp = x_\perp \frac{x'}{x} . \tag{2.14} $$

In Fig. 2.13 the transformation ratio of the wiggle function is plotted at different redshifts for two cosmologies that differ from the reference model only in w_0 (the redshift-independent part of w). The reference model is the concordance cosmology. Fig. 2.13 shows that the largest deviations from the reference model appear at redshifts in the range $0.5 < z < 1.5$. Differences in parallel and perpendicular scaling introduce anisotropies that could be measured, too (see below).
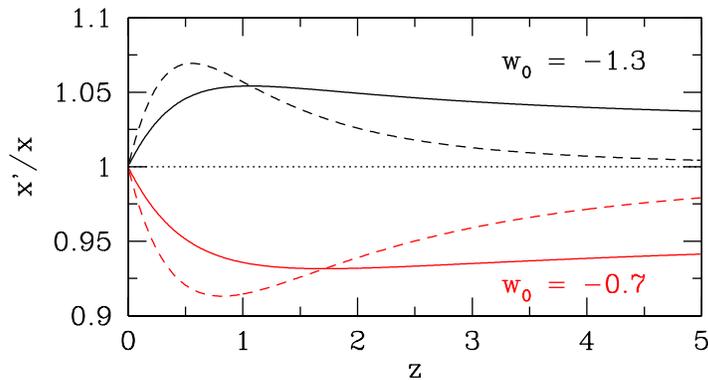
Figure 2.13: Transformation of perpendicular (solid) and parallel (dashed) distances in comparison to the concordance model with $w_0 = -0.7$ (upper part) and $w_0 = -1.3$ (lower part).

Figure 2.14 shows the scaled wiggle function at redshift $z = 2.5$ for $w_0 = -0.7$. The reference and the distorted wiggle function are plotted. Note that an elongation of the coordinates in real space leads to a compression of the coordinates in k-space and vice versa. In the upper panel, it can be seen that the wiggle function gets more and more out of phase as k increases. Thus the sensitivity of a cosmological test should increase at high values of k. However, the amplitudes of the oscillations decrease exponentially with k (e.g., Silk damping). This can be seen in the lower panel of Fig. 2.14 where the scaled wiggle function is subtracted from the reference ($w_0 = -1$) wiggle function: Though the
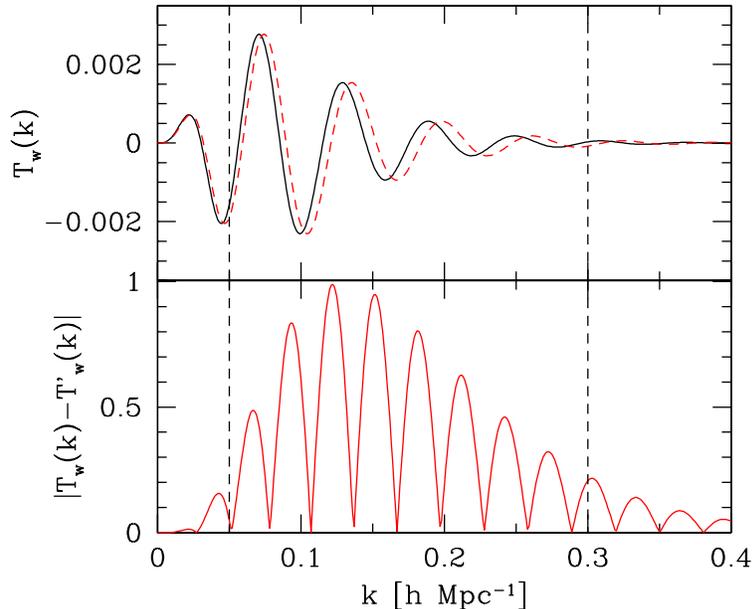
Figure 2.14: Scaled wiggle functions at $z = 2.5$ for $w_0 = -0.7$ (dashed line) compared to the $w_0 = -1$ cosmology (solid line, top panel). The normalized modulus of the difference between the reference cosmology and the scaled cosmology is shown in the bottom panel. The dashed vertical lines mark the region relevant for the cosmological test.

difference in phase is increasing, the amplitude decreases so fast that the difference between the projected oscillations increases at small values of k and decreases at high values of k.

The present version of the cosmological test comprises the following steps: (1) Angular and redshift coordinates of the 'observed' (simulated) data are transformed into comoving coordinates using a concordance model as the reference cosmology. (2) The comoving data sample is Fourier-transformed to estimate the boosted transfer functions. (3) BAOs are extracted with FITEX. (4) The theoretical wiggle function is computed for the corresponding cosmological model. (5) The parallel and perpendicular scaling factors between the test cosmology and the reference cosmology are computed at the redshift of the data sample. (6) The theoretical wiggle function is scaled by constructing a three-dimensional wiggle function, scaling each mode and collapsing it to one dimension. (7) The observed wiggle function is compared to the re-scaled theoretical wiggle function assuming Gaussian statistical errors (χ^2 test). Only the steps (4-7) have to be repeated to test a different cosmology, allowing a fast test of various parameters.

Note that the three-dimensional wiggle function is only constructed to simulate the effects of different parallel and perpendicular scaling factors. It does not address the anisotropy introduced by redshift space effects, because we found that this only boosts

the various modes but not change the phases of the theoretical wiggle function (see Sect 2.5). The boost effect of the redshift space distortions, along with other boosting effects like biasing and growth suppression, will force us to add an additional parameter to the cosmological test.

Figure 2.15 shows in the upper left panel the combined wiggle function of the four XW wedge cubes of the Hubble Volume Simulation. The combination is performed by averaging over k-bins of discrete size. The errors represent the standard deviation of all wiggle points in the corresponding k-bins. After averaging, the 2nd to 5th BAOs are now clearly visible. We thus conclude that a volume at least 4 times larger than SDSS has to be sampled in order to see the BAOs with FITEX with high significance. One has to keep in mind, though, that our cubes are located at much higher redshifts, where the smoothing effect of non-linear structure growth is obviously much smaller. We also neglected the fact, that larger sky coverages reduce the sizes of the fundamental modes, leading to higher (better) sampling rates of BAOs.

With the 100 square degree sample, the possibility of data points appearing at positions which severely inhibit the accuracy of the cosmological test is relatively high, due to the high sample variance. Especially as only a few data points, which are far away from positions where theoretical wiggle functions of different cosmological models intersect, have great statistical weight in the cosmological test. This fact is shown in Fig. 2.14, where the modulus of the difference between the theoretical transfer functions for two models depending on position in k-space is plotted.

Figure 2.16 shows the results of the cosmological test obtained with BAOs extracted with FITEX under different survey conditions. For surveys covering 100 square degrees we would expect errors of $\sigma_{w_0} = 0.25$. The asymmetry of the error distributions directly reflects the increased sensitivity of metric scales on w for $w_0 > -1$ compared to $w_0 < -1$. At these w_0 values, the theoretical wiggle function gets quite compressed, so that at large k values the amplitude decreases rapidly due to Silk damping. When $w_0 \rightarrow -\infty$, the theoretical wiggle function formally mimics the case of zero amplitude, so that low values of w_0 are excluded with about the same significance as the "no-wiggle" case of zero amplitude. The minima of all error distributions plotted in Fig. 2.16 are slightly off-set from the input $w_0 = -1$ value, but well within all formal 1σ ranges.

Note that the constraints given above are based solely on the phases of the oscillations in the one-dimensional power spectrum, assuming perfect knowledge of the sound horizon at drag epoch. The w-constraints are expected to further improve when the full two dimensional redshift space effects as well as the amplitudes of the BAOs are used.

Perhaps the most important result of the present investigation is the comparatively small decrease of only 8% of σ_{w_0} when comparing the w test with unbiased matter distributions in configuration space, (dashed line in Fig. 2.16) i.e., the simplest survey condition, with the test that includes all aforementioned effects (solid line). The robustness of the FITEX method to extract BAOs from boosted transfer functions with phenomenological fitting functions is mainly related to the fact that only the amplitudes of the wiggle functions are boosted by redshift-space and biasing effects, but the phases are not affected.
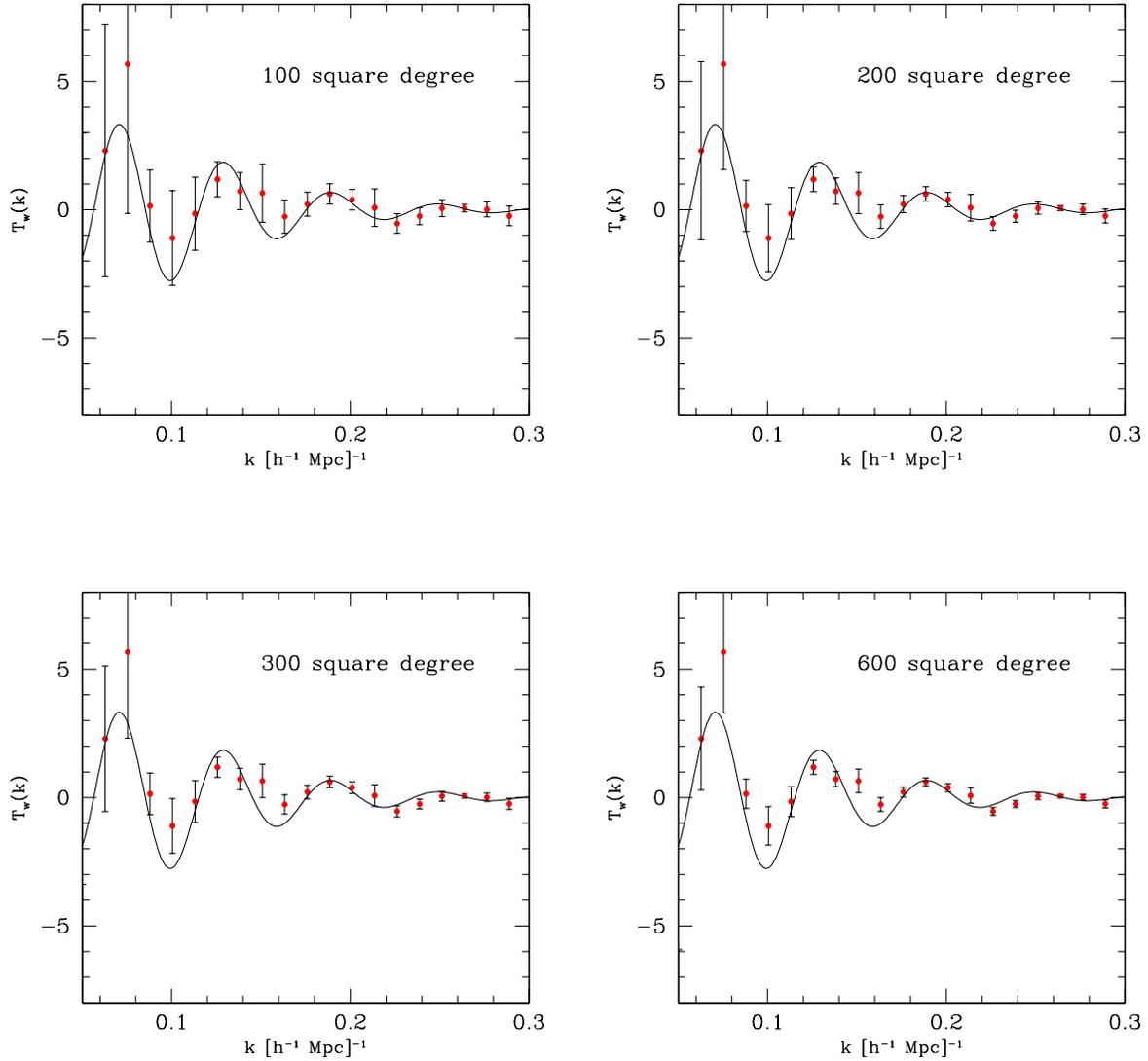
Figure 2.15: Combined wiggle functions from the four Hubble Volume cubes. The upper left panel shows the combined data point (dots) with self-consistent error bars for the 100 square degree field. The upper right panel shows expected error bars for 200 square degrees and the lower row with error bars for 300 and 600 square degrees, respectively. The continuous lines are the theoretical wiggle functions for the reference (concordance) cosmology.
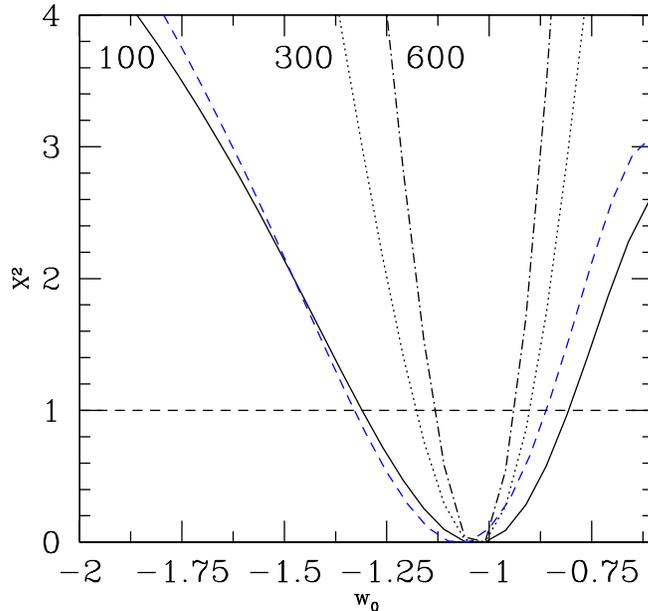
Figure 2.16: χ^2 distributions for the w_0 parameter obtained after marginalization over the amplitude parameter and for different conditions of BAO extraction. Short-dashed line: dark matter distribution in configuration space over 100 sq. deg obtained from the simulation of dark matter particles in configuration space. Continuous line: biased particle distribution in redshift space over 100 sq. deg. Dotted line: biased particle distribution in redshift space over 300 sq. deg (extrapolation). Dashed-dotted line: biased particle distribution in redshift space over 600 sq. deg (extrapolation). The horizontal dashed line marks the 1σ error level.

To give an idea of the performance of FITEX based on surveys of larger size, we extrapolated the results from the 100 square degree data. The upper right panel of Fig. 2.15 shows wiggle functions extracted from the 100 square degree data but with error bars expected for 200 - 600 square degrees. The error bars were extrapolated by scaling them according to the expected number of additional k-modes. The different survey fields are thus assumed to be not adjacent.

Figure 2.15 shows that increasing the survey volume to 200 square degrees does not reduce the error bars significantly, still allowing for crucial data points at peaks of the oscillating wiggle function to be reduced to zero amplitude. Error bars are still of the same order of magnitude as the oscillations. Only at 300 square degrees (lower left panel), the probability for smoothing out the oscillations due to statistical errors gets considerably low. Following our extrapolations, for a survey of 600 square degrees, the errors are expected to get much smaller than the amplitude of the oscillations and the chance of statistical deviations from the predicted function is of an order of magnitude that allows for very accurate cosmological probing.

The accuracies of the w_0 parameter (Fig.2.16) range from $\sigma_{w_0} = 0.25$ for 100 sq. deg which corresponds to 3 SDSS volumes, to $\sigma_{w_0} = 0.09$ for 600 sq. deg, corresponding to 21

SDSS volumes. These results are obtained for biased samples in redshift space, i.e., the most complex survey condition. The improvement is directly related to the larger number of independent k-modes in larger survey volumes and is consistent with results obtained with Monte-Carlo simulations (Blake & Glazebrook, 2003).

Future studies should also investigate the various parameters of the cosmological test to get constraints under less strong priors.

2.8   Discussion and Conclusions

We have seen that it is critical for the extraction of BAOs to either fit a phenomenological function to the continuum of the boosted transfer function (FITEX), or to calculate the shape of the boosted transfer function, including all distortions down to the sub-percent level. The latter approach has been carried out recently. Glazebrook & Blake (2005), Hütsi (2006), and Springel et al. (2005) used a linear theory power spectrum and divided the 'observed' data by this reference. Almost the same approach, but with the addition of linearizing the observational data before the division, to correct for non-linear structure growth was used in Angulo et al. (2005). There are two drawbacks of this method:

First, the ratio of a complex power spectrum and a reference power spectrum does not resemble a wiggle function as defined in the text, but something more complicated (see Eq. 2.4). The simple form of the theoretical wiggle function (Eq. 2.1) is strongly deformed by both the chosen transfer function of the reference spectrum and the distortions introduced by redshift space and galaxy biasing. These distortions make the comparison with theoretically expected model spectra unnecessarily complicated. FITEX transforms the observed power spectrum into a boosted transfer function where the subtraction of a non-oscillating phenomenological continuum function is directly related to $T_w(k)$, multiplied by an amplitude factor.

The second drawback is the linear theory power spectrum itself. By computing a correct shape of the power spectrum, one has to rely strongly on assumptions about the nature of the dark matter and various other physical processes. The method is no more the 'assumption free' approach that was intended. Furthermore, the calculation of the amplitude factor that includes all the distortions ranging from linear structure growth, non-linear structure growth to redshift space distortions and biasing is a problem. This calculation has to be accurate to the sub-percent level, as the oscillations themselves make up only $\sim 2\%$ of the transfer function.

A more phenomenological approach was conducted by White (2005). He assumed that redshift space distortions as well as linear and non-linear structure growth could be calculated with reasonable accuracy. Leaving only the ratio between dark matter power spectra and 'galaxy' power spectra, he found that one can fit this ratio with about 2% accuracy using a cubic polynomial. This adds further evidence that a phenomenological approach can model power spectrum distortions without assuming much about the physics of the distortions themselves.

With the FITEX method, we have shown that one is able to extract the BAOs without

any strong priors on the nature of dark matter or the shape of the power spectrum from the complex multi-component transfer function. All the aforementioned effects are modeled by the continuum function and need not be known very accurately, as long as the phases of the oscillations are not distorted or washed out. This extraction is accurate on the sub-percent level within k-ranges relevant for the cosmological test. Furthermore, FITDEX is able to disentangle phase information from amplitude information which could in the future be used to test the geometry of the Universe on the one hand and the growth history on the other hand. This type of cosmological test could be used to discriminate Dark Energy from Modified Gravity theories.

Chapter 3

The Impact of Sparse Sampling on Power Spectrum Cosmology

Note: this chapter will be submitted as Koehler et al. 2009. The work was performed under the supervision of Niv Drory and with comments and suggestions by Eiichiro Komatsu.

3.1 Introduction

In the past, Blake et al. (2006) investigated the effect of sparse sampling on the ability to recover Baryonic Acoustic Oscillations (BAOs) in the galaxy power spectrum. They also simulated its impact on a cosmological test to constrain cosmological parameters like the equation of state of dark energy w. Sparse sampling introduces correlations between the different Fourier modes (see Blake et al., 2006) and washes out the oscillations reducing their amplitude and thus their detectability. By using the full shape of the of the power spectrum, one is less prone to effects of correlations, but has to model the spectrum very accurately including non-linear growth and scale dependent biasing to the percent level. Recently, Jeong & Komatsu (2006) and Jeong & Komatsu (2008) have shown that it is possible to calculate these effects accurately at higher redshifts, using third order perturbation theory and exploiting information from the bi-spectrum.

Cosmic variance describes the error of a sub sample representing the statistical properties of the related super sample. The bigger the volume of a survey, the bigger the fraction of the universe covered and the smaller the cosmic variance. Cosmic variance affects mostly smaller k modes, as increasingly more k modes of larger value fit into a given volume. Shot noise describes the error of a sample with a finite point density representing the properties of an underlying continuous density field. The higher the point density the smaller the fluctuations in the density can be that can be accurately modeled. Shot noise affects mostly larger k modes, as a higher point density is needed to model smaller scales correctly.

To increase survey volume and thus minimize error from cosmic variance, sparse sampling surveys do not sample the sky continuously, but leave large gaps between measurements, to trade volume for point density. The related window functions heavily distort the

resulting power spectrum and introduce correlations both in the power itself and its error.

The upcoming Hobby-Eberly-Telescope Dark Energy eXperiment (HETDEX) (Hill et al., 2008) survey is utilizing Integral Field Spectroscopy Units (IFUs) and applies sparse sampling techniques (see Kaiser, 1986; Colombi et al., 1998) to search for Lyman Alpha Emitters (LAEs) as tracers for the underlying matter density field of the universe. The power spectra derived from this matter field can be used to test cosmological models, exploiting the impact of cosmological parameters like the equation of state of dark energy w on the shape of the power spectrum. The ability to constrain cosmological parameters depends on the relative accuracy of the measurement $\sigma(k)$, which is dominated by the two sources of errors: Cosmic variance and shot noise. The shot noise depends on the power in the sample $P(k)$ and the average point density $\bar{n}$, while cosmic variance depends on the size of the survey L:

$$\sigma(k) = \frac{\sqrt{2\,\pi}}{k\,L}\left[1 + \frac{1}{\bar{n}\,P(k)}\right]. \tag{3.1}$$

In the present work, we will investigate the impact of such a window function on the results of a cosmological test utilizing the full shape of the power spectrum. We will use a set of simulations described in Sect. 3.2 to directly test the impact of the proposed HETDEX IFU selection functions as well as additional zero-point effects (Sect. 3.3) on the power spectrum. Section 3.4 will investigate the effect on the errors of the power spectrum, the error correlation matrix, and show how and to what extent one is able to correct for the distortions. Finally we will show how the introduced additional errors and correlations are affecting the accuracy of a cosmological test exploiting both geometrical and growth effects of cosmological parameters in Sect. 3.5.

3.2   The Simulations

To test the various effects of different selection functions on the power spectrum and its errors we simulate 400 cubes with a baseline of $500\,h^{-1}\,\mathrm{Mpc}$ and a total volume of $50\,h^{-1}\,\mathrm{Gpc}$ using the Pinocchio (PINpointing Orbit-Crossing Collapsed Hierarchical Objects) semi analytical method (see Monaco et al., 2002). The Pinocchio algorithm uses a numerical representation of a linear density field to emulate non-linear structure growth by grouping particles into halos, simulating the effects of accretion and mergers during the procedure. It calculates positions, velocities, merger history, and angular momentum of dark matter halos, where positions are accurate to 2 Mpc and velocities to about $100\,\mathrm{km\,s^{-1}}$. We use the power spectrum generated with the fitting function of Eisenstein & Hu (1998) as initial condition.

Table 3.1 gives additional parameters of the simulations. The maximum and minimum k-values accessible to the simulation are given by the scale length of $500\,h^{-1}\,\mathrm{Mpc}$, and the Nyquist frequency $k = 1.61\,h\,\mathrm{Mpc}^{-1}$ of the simulations which equals $3.9\,h^{-1}\,\mathrm{Mpc}$, respectively. However, surveys with a baseline of smaller than $2\,h^{-1}\,\mathrm{Gpc}$ or which are

Table 3.1: Pinocchio Parameters

Par	Value	Description
L_{grid}	256	grid size in 1 dimension
n	1.3e6	initial particles
Ω_Λ	0.75	dark energy density
Ω_m	0.25	matter density
Ω_b	0.045	baryon density
H100	0.72	normalized Hubble constant
L	500	cube length in $h^{-1}\,\mathrm{Mpc}$
z	2.00	final redshift
σ_8	0.8	power normalization

highly elongated do not gain a relevant amount of information from scales larger than $500\,h^{-1}\,\mathrm{Mpc}$, and are thus also covered by the scale of the simulations used in this work.

Further testing suggest that the simulations should only be trusted to about $0.5\,h\,\mathrm{Mpc}^{-1}$ which equals a scale length of $12.6\,h^{-1}\,\mathrm{Mpc}$. This is mostly because of rapid power loss at larger k values due to the initial conditions of the simulations, where the particles are arranged on a regular grid. Similar effects can be found in other numerical simulations like the Hubble Volume Simulation (see Koehler et al., 2007), where an initial glass distribution distorts low k modes at high redshift.

The final redshift-space power spectrum averaged over all the 400 cubes is illustrated in Fig. 3.1. One can clearly see the power decay at higher values of k, where it deviates from the linear prediction. The given error bars are calculated from the variance of the full set of simulations and represent the errors of only one cube and agree very well with predictions (see Subsect. 3.4.2). The errors of the mean power spectrum itself are smaller by a factor of $\sqrt{400}$.

3.3 Selection Functions

Various selection functions work on the power spectrum to produce the final power observed.

The proposed HETDEX survey will utilize 96 IFUs, arrayed on a regular grid. The IFUs themselves have a fill factor of only 1/3rd, but a dithering pattern will ensure a complete sampling on the area of one IFU. This setup introduces, along with the radial selection function of a luminosity limited survey, a unique angular selection function.

We start looking at the radial selection function, which is dominated by the luminosity function of the LAEs and the thoughput of the telescope and instrument. Then, we will investigate the angular window function, dominated by the field of view of the IFUs and their fill factor of about 1/6th. And finally we will look at, zero-point offsets in the luminosity calibration that introduce correlations between nights, weeks and months, by systematically over or under estimating the true matter density in one night, week, or
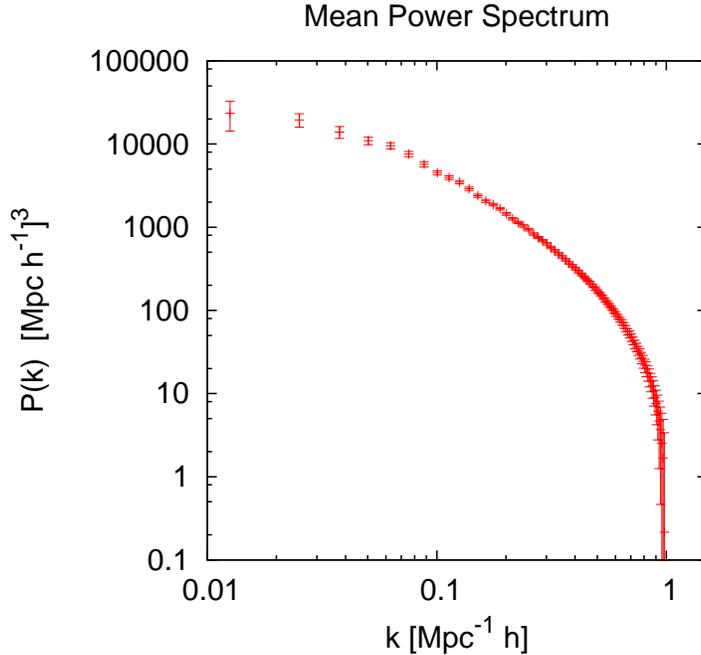
Figure 3.1: Mean power spectrum of 400 cubes with error bars of one cube calculated from the variance within the whole sample.

month.

3.3.1   Radial Window Function

As HETDEX is a flux limited survey, the luminosity function and the distance to the object dominate the radial selection function. However, because HETDEX utilizes IFUs to take spectra of every position on the sampled sky, the instruments throughput depending on wavelength and the sky spectrum also introduce selection effects in the radial direction. While the luminosity distance and instrument throughput are smooth functions of the distance to, and thus wavelength of, the LAE, the sky spectrum is not smooth at all and could potentially introduce strong correlations in the resulting power spectrum.

Figure 3.2 shows the radial selection function. It is calculated by multiplying the throughput with the object density depending on distance and dividing by the sky spectrum, as sky lines decrease the probability of finding a LAE at the related wavelength.

As the Pinocchio simulations only model dark matter halos and not galaxies, we simply apply the radial selection function by selecting each halo according to a probability given by the relative throughput of the selection function. This procedure neglects the fact that LAEs with higher luminosities might or might not have a higher biasing than LAEs with lower luminosities. However, this topic is not yet well investigated, and the effect is mostly a change in slope and amplitude of the biasing. For a cosmological test the biasing has to
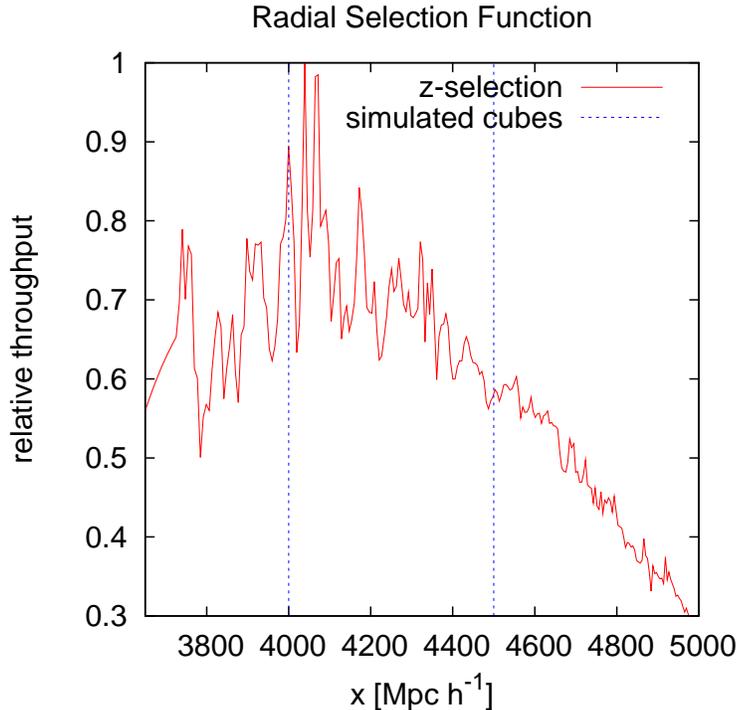
Figure 3.2: Predicted fraction of LAEs detected at a certain distance normalized to the maximum density. Vertical lines contain the region used for the simulations.

be calculated using the higher order statistics of the bi-spectrum as suggested in Jeong & Komatsu (2008). In this work, we assume that biasing was determined using the suggested methods to a maximum k value of $0.3 \, h \, \mathrm{Mpc}^{-1}$.

3.3.2 Angular Window Function

The proposed HETDEX selection function in angular dimension is heavily dominated by the setup of fiberheads in the Visible Integral-field Replicable Unit Spectrograph (VIRUS). VIRUS covers only a small fraction of the telescopes field of view. The ratio of covered area, and thus volume, to uncovered area and volume, the fill factor, is only about 1/6th. This allows HETDEX to find an optimal balance between the cosmic variance error that is dominated by the covered volume, and the shot noise error, which is dominated by the number of observed LAEs (see Eqn. 3.1).

As the IFU mask completely masks out a majority of the sky in angular direction, the resulting window function is conceptually different from angular selection functions of other large surveys where power spectra are estimated like the SDSS (Tegmark et al., 2006) or 2dF (Cole et al., 2005). In these surveys areas can be weighted according to their particle density to compensate for selection effects like extinction or luminosity limitations. In the case of HETDEX, the appropriate weight for a region with a points density of 0 would be infinity and can not be corrected for or weighted. The information in the masked out area

Angular Selection Function
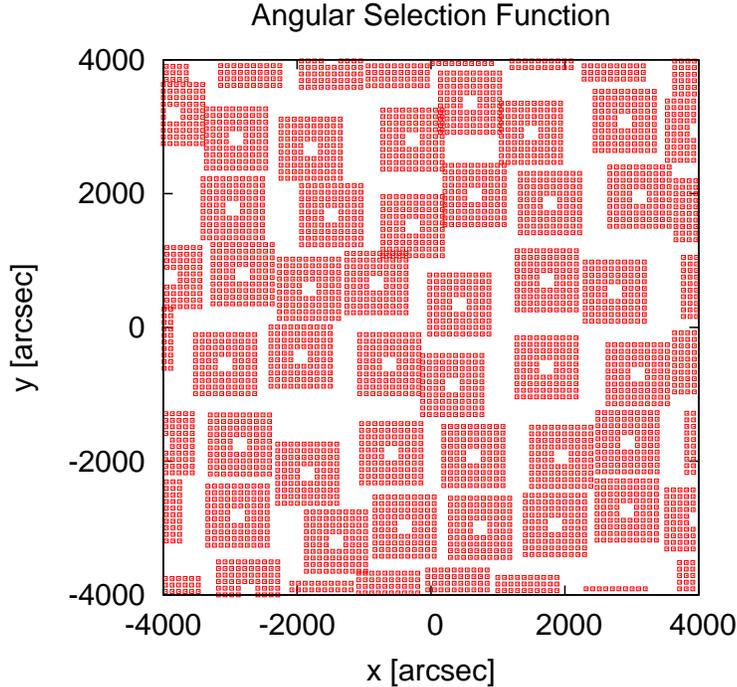
Figure 3.3: Angular selection functions displaying one shot of VIRUS (big squares) and the 96 fiber-heads in each shot (small squares). The figure shows only a small part of the whole survey.

is lost.

Figure 3.3 illustrates the IFU selection function on they sky in angular coordinates. It shows only a small part of the 400 degree survey area. A number of shots containing the 96 fiber-heads are arranged in a static and very regular pattern in one shot. Our simulations show, that it is not necessary to arrange the IFUs in a shot randomly or even randomize the IFU setting from shot to shot (see below), as the length scale of one IFU on the sky is only about $0.5\,h^{-1}\,\mathrm{Mpc}$, while the scales HETDEX is working on are about $50\,h^{-1}\,\mathrm{Mpc}$. The static setup of the IFUs in the shots is broken by a large gap in the middle, with a size of two times two fiberheads. This gap is used for other instruments.

The shots themselves are arranged in a semi random way using Monte Carlo methods. Firstly the shots are distributed on a regular grid, with some gaps between the shots. Then one shot is randomly selected and moved by a random small distance in a random direction. If the shot overlaps another shot, the new position is canceled, if there is no overlap (within a small allowance) the new position is accepted. This procedure is repeated until the peak in the resulting power spectrum on the scale of the shot-separation is washed out as much as possible.

We use this technique to simulate the semi random shot distribution resulting from the distribution of stars on the sky. HETDEX will want to cover as many foreground stars as possible for calibration. To achieve this, the shots will be nudged just a little from their original position on a regular grid to cover an optimal number of stars within their 1/6th

fill factor.
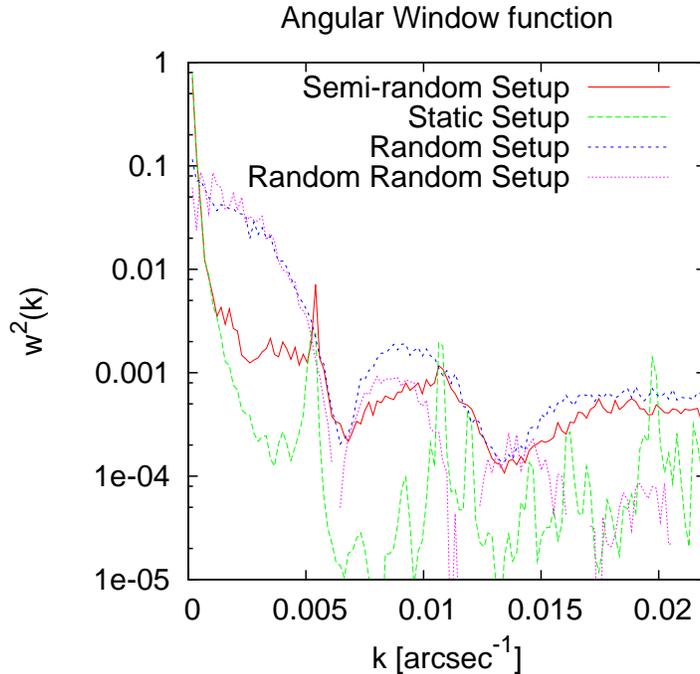
Figure 3.4: The angular window function of a selection based on shots on a regular grid (long dashed line), a semi-random shot distribution (solid line), a random distribution (short dashed line), and a random shot, random IFU distribution (dotted line).

The difference between this semi-random shot setup and a complete regular grid is, however, not that big. Figure 3.4 shows the averaged 1-dimensional angular window functions of both setups. In the semi-random setup, the two dominant peaks from the grid size itself at $k = 5 \times 10^{-3}$arcsec^{-1} and the harmonies at two and four times this value are washed out in comparison to the regular grid. However, the semi-random window function has more power itself, than the static function, which will lead to a stronger distortion of the power spectrum in the end.

A completely random placement of the shots with an average distance D was also considered. This has the advantage of having a simple window function with a $1/D$ profile, where all other features are washed out and smoothed as much as possible (see Kaiser, 1998). However, as D is comparable to scales of interest, the window function is very broad and has a lot of power itself. Furthermore, the completely random placing leads to heavy overlap, effectively decreasing the sampled volume by a considerable amount. Because of this, the completely random shot setup was dismissed.

Finally one can compare the completely random shot setup with IFUs placed on a static grid within a shot, and a setup where not only the shots are randomized, but also the IFUs setup within every shot. Figure 3.4 shows a difference on the scale of $k = 0.01$arcsec and higher where the correlations are dampened compared to the static IFU case. The

difference is, however, not big enough to compensate for the effort to mechanically reshuffle the setup of the IFUs for every shot.

3.3.3 Luminosity Calibration

Not only the pattern of shots itself, but also the clustering of shots within one night or even week or month might introduce correlations into the data. The shots of one HETDEX reference design night are clustered into two squares approximately 10 degrees apart. At a redshift of two, this equals more than the maximum scale length $500\,h^{-1}\,\mathrm{Mpc}$ of our simulations. Thus, we only capture the effect introduced by the correlations of one cluster.

   To simulate this effect, we randomly calculate the extinction in this night, based on real data. We assume that luminosity zero-point is calculated correctly with a lognormal error distribution

$$\delta_{\mathrm{zp}}(x) = \frac{1}{\sigma\sqrt{2\,\pi}} e^{-\frac{(\ln(x)-\mu)^2}{2\,\sigma^2}} \; , \tag{3.2}$$

   normalized to a standard deviation of of 5% and 10% and a mode of 1. The lognormal distribution is motivated by the experience from the VIRUS-P prototype survey (see Hill et al., 2008).

   Each shot in a night shares the same zero-point error, which means that all the shots in one night are 100% correlated. The error in number counts is then computed by calculating the difference in number density for the real extinction, and the errornous extinction depending on distance, and applying this ratio as a weight to all the detected LAEs in one night.

   We assume a three by three shot pattern for one of the two clusters a night, only one of which will fall into our simulated volume. Correlations between weeks and months are not considered in this work, as the nightly correlations will be much stronger and can thus, especially with a maximum correlation of 1, be used as an ideal test case to judge the impact of correlations of other time scales and correlation strengths.

3.3.4 Cartesian Window Function

In spherical coordinates, both the angular as well as the radial correlation function are independent of each other. However, a three dimensional power spectrum can not easily be calculated using spherical coordinates, as the resulting power spectrum scales change from radial shell to radial shell and can not easily be added up to a cumulative or mean power spectrum of the whole volume without washing out major features. Angular power spectra in redshift slices are used to achieve this (see Blake et al., 2007), but throw away information in form of the correlations between the slices. It was furthermore shown that lightcone effects and the breaking of the distant observer relation in rectangular volumes of about $10\,h^{-1}\,\mathrm{Gpc}$ and a redshift of 2 and more do not affect the calculated power spectrum much (see Wagner et al., 2008).
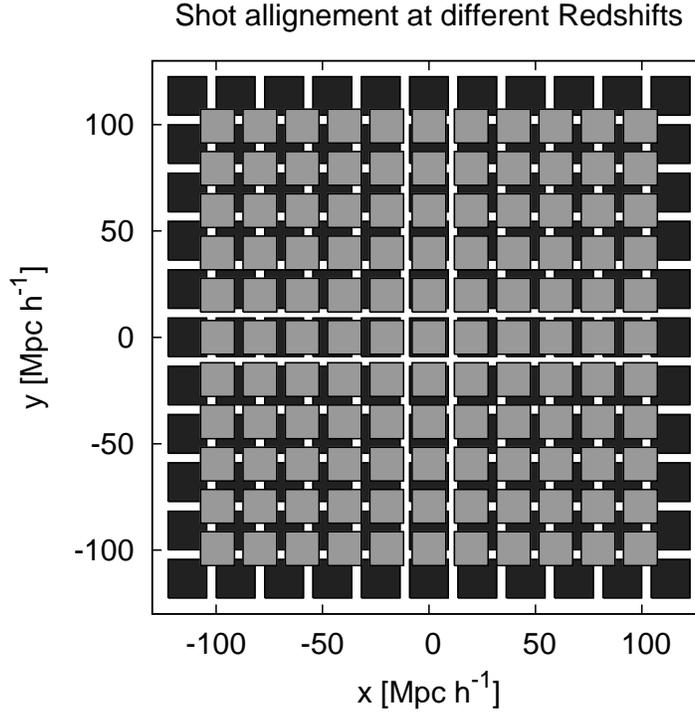
Figure 3.5: Spatial distribution of shots on a regular grid at a distance of $3500\,h^{-1}\,\mathrm{Mpc}$ (gray squares) and $4000\,h^{-1}\,\mathrm{Mpc}$ (black squares).

We will thus investigate the impact of both the angular and radial selection functions using a Cartesian Fourier transform to calculate the power spectrum. The breaking of the distant observer relation leads to a mixing of the formerly independent selection functions and heavily distorts the angular pattern depending on the distance to the observer.

Figure 3.5 shows how an angular selection function with equally spaced shots in angular coordinates would work on a sample of simulated LAEs in a distance of 3500 (blue) and $4000\,h^{-1}\,\mathrm{Mpc}$ (red). One can clearly see that any distinct scales, like the spacing of the various shots from each other, or the size of the shots themselves, are distorted with the distance from the center of the cube, and the distance to the observer. Because of this, the power on any angular scale is washed out and distributed within the whole cube in the Cartesian Fourier transform.

As a result, only the modes purely parallel to the line of sight carry the signal of the washed out radial selection function and the modes purely transversal to the line of sight carry the signal of the washed out angular selection function. Figure 3.6 shows a logarithmic representation of the two dimensional window function, where the two perpendicular dimensions were integrated to one. Modes that comprise of both parallel and perpendicular parts have practically no power and resulting correlations are kept to a minimum.
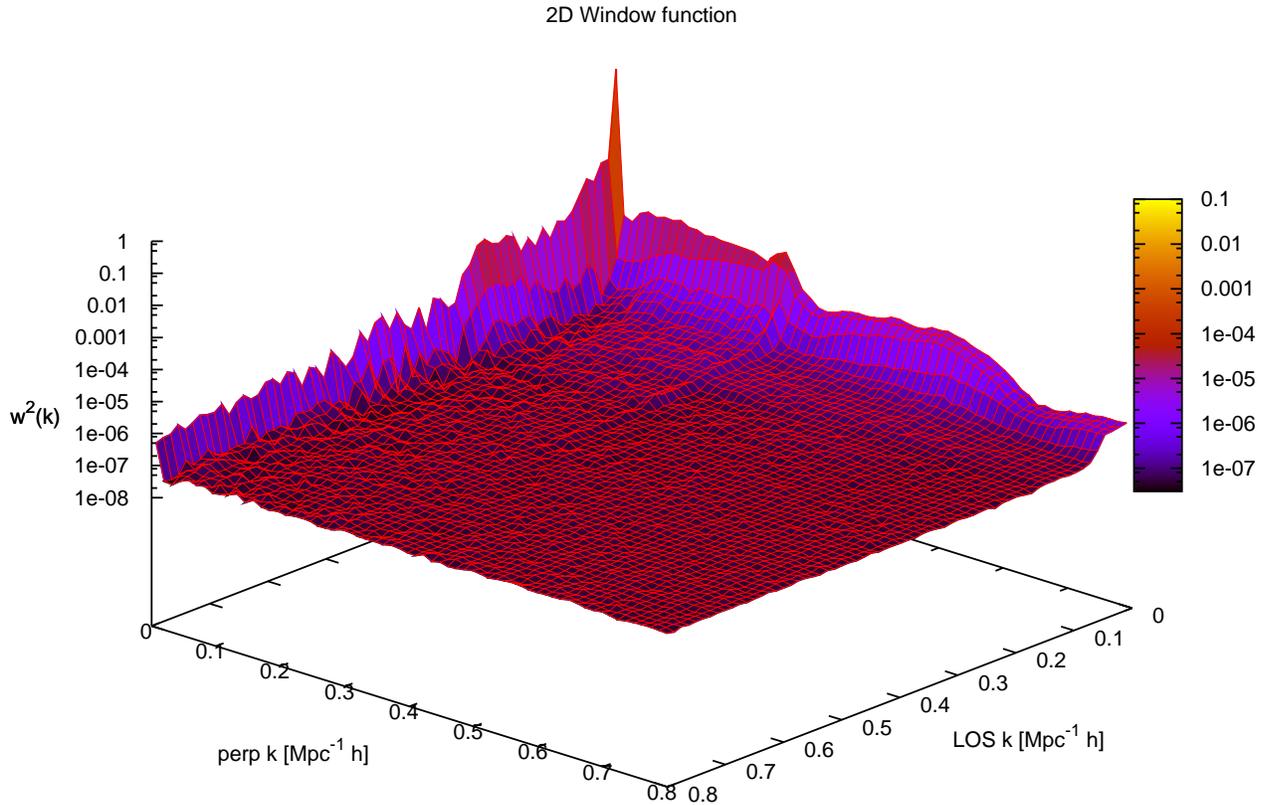
Figure 3.6: Window function in both the parallel and perpendicular direction. Note the logarithmic scale in z.

3.3.5   Samples

By applying all the selection functions to the 400 simulation cubes, we end up with five different samples:

- an undistorted sample, including redshift-space effects and (semi-analytic) non-linear growth

- a sample including the above plus radial selection effects

- a sample including the above plus angular selection effects

- a sample including the above plus 5% zero-point errors

- a sample like the above but with 10% zero-point errors

All the samples were generated by projecting the angular and radial selection functions to a Cartesian coordinate system, placing the cubes at a distance of 4000 to 4500 h^{-1} Mpc from the observer at an average redshift of 2.5.

A number of effects are not covered. They include, but are not limited to:

- lightcone effects (see Wagner et al., 2008)

- scale dependent biasing (see Jeong & Komatsu, 2008)

- luminosity dependent biasing (see Orsi et al., 2008)

- velocity dispersion effects of emitting medium (see Tapken et al., 2007)

All of them should, however, not affect the comparison of power spectra with different window functions.

3.4   Error Analysis

This section will show the effects of the various selection functions on the power spectrum itself, the errors of the power spectrum, and the correlations of the errors.

3.4.1   Systematic Errors

All three window functions introduce distortions to the power spectrum. An initial overdensity distribution $\delta(\vec{x})$ is transformed by an arbitrary selection function $n(\vec{x})$ like,

$$\tilde{\delta}(\vec{x}) = n(\vec{x})\delta(\vec{x}), \tag{3.3}$$

with $0 \leq \vec{k} \leq 1$. The FKP algorithm (see Feldman et al., 1994) calculates overdensities correctly by subtracting a homogeneous sample multiplied by the selection function from the inhomogeneous sample. The resulting estimated power spectrum after subtraction of shot noise is then a convolution of the true power spectrum with the window function:

$$\tilde{P}(\vec{k}) = P(\vec{k}) \otimes |n(\vec{k})|^2 \,. \tag{3.4}$$

The window function is given by

$$n(\vec{k}) = \int d^3x \, n(\vec{x}) \exp(i\vec{k} \cdot \vec{x}), \tag{3.5}$$

the square of the Fourier transformed selection function. The power spectrum itself is calculated using the FKP algorithm (see Feldman et al., 1994) using a grid with 256 grid points and a next grid point allocation algorithm (NGP). We do not correct for aliasing effects, but it was shown by Jing (2005) that the lower k-modes are not affected by aliasing.

The ratios between the undistorted power spectrum and the z-selected and IFU-selected power spectra of the 400 cubes are plotted in the upper panel of Fig. 3.7. The lower panel shows the ratio between the IFU-selected power spectra and the 5% and 10% zero-point errors power spectra, respectively.
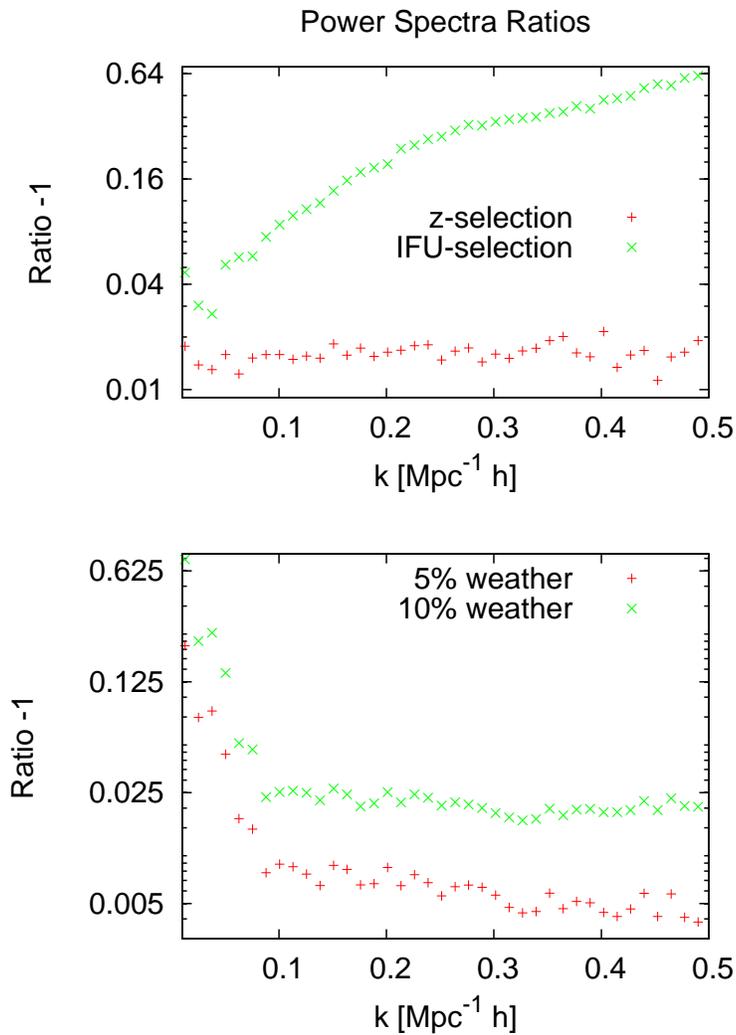
Figure 3.7: **Upper panel:** z-selected (+) and IFU-selected (x) power spectra divided by the undistorted redshift-space power spectra.
Lower panel: Power spectra from data with 5% zero-point errors (+) and 10% zero-point errors (x) divided by the IFU-selected power spectrum without zero-point effects.

While the absolute change in amplitude due to the radial window function and the zero-point errors are relatively small, the angular window function heavily distorts the observed power spectrum.

Another difference is the scale most affected by the distortions. Angular effects increase at smaller scales, while zero-point effects decrease at smaller scales. This is expected, as the angular window function is dominated by the relatively small scale of the shots, but much more by the scale of the IFUs fiberheads themselves which are, however, too small to be resolved by the power spectra of our simulations. Zero-point effects on the other hand are dominated by the much larger cluster scale of correlated shots in one night.

Correcting for Selection

To measure the impact of the distortions due to the window functions on a cosmological test we have to calculate how accurately the initial power spectrum can be recovered from the observations, or how well a distorted template can be calculated.

As a convolution is much easier to calculate than a deconvolution we try to model the distorted power spectrum by convolving a theoretical model power spectrum by the window function. It is not necessary to do this convolution for all possible scale factors in the following geometric test, as we are only probing for small variations from the reference cosmology. In this case, the linear expansion, and thus a simple re-scaling of the convoluted power spectrum to the tested scale factor (see Sect. 3.5) does suffice.

We compute the three dimensional window function numerically by applying the IFU- and z-selection functions on a high density random sample and compute the Fourier transform with fast Fourier transform algorithm (FFT).

Figure 3.8 shows the ratio between the convolutions with various degrees of resolutions and the power spectrum including the radial and angular selection effects. We start with 256 grid points per $500\,h^{-1}\,\mathrm{Mpc}$ cube and increase to 512 and 1024 grid points. However we achieve this not by increasing the actual number of grid points but use the same number of grid points on the full cube, a smaller sub-cube with half the length, and an again smaller sub-cube with one fourth the base length. This distorts the lower k-scales of the higher resolution spectra and only the high k-scales should be accurate. Furthermore, the different resolutions vary, because the smaller samples of the higher resolutions only capture a fraction of the overall, inhomogeneous, window function.

The convolution with the lowest resolution agrees well at low k-scales, which was expected. The highest resolution agrees well with the power spectrum including the selection effects at higher k-scales, while it is loosing power on large scales. However, the error bars due to different shot noise and sample variance are still of an order of 1% per data point, which is approximately the accuracy of a survey with $8\,h^{-3}\,\mathrm{Gpc}^3$ volume and 10 million tracer particles with a biasing of 2 at $k = 0.2\,h\,\mathrm{Mpc}^{-1}$ and a redshift of 2. HETDEX is reaching an accuracy of about 2.5% at this k-value.

Thus, we can conclude that the issue of correcting for the window function has to be examined further to exclude any systematic effects at an accuracy comparable to the surveys statistical error. The problem is not to calculate the window function to an arbitrary
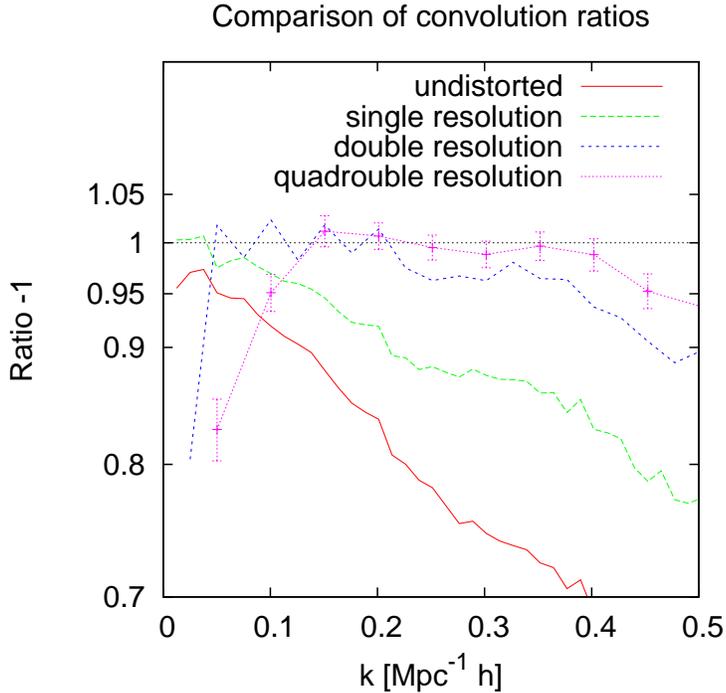
Figure 3.8: Ratio between the undistorted power spectrum (solid line), the single resolution convolution (long dashed line), double resolution convolution (short dashed line), quadruple resolution convolution (dashed dotted line) and the IFU-selected power spectrum.

accuracy, but confirming the accuracy by carrying out the convolution.

Correcting for Zero-Point Uncertainties

Assuming a Gaussian zero-point error distribution with a variance of σ_{zp}^2 and a window function $w(x)$ the overdensities $\hat{\delta}(\vec{x})$ with zero-point errors can be calculated like:

$$\hat{\delta}(\vec{x}) = n(\vec{x}) \left[\delta(\vec{x}) + \sigma_{\text{W}} w(\vec{x}) \delta(\vec{x}) \right] . \tag{3.6}$$

Because of the distributivity of convolution, we can express the power spectrum with zero-point effects and selection effects, $\hat{P}(k)$, by a sum of only the spectrum with selection effects, $\tilde{P}(k)$, and an additional power term:

$$\hat{P}(\vec{k}) = \tilde{P}(\vec{k}) + \sigma_{\text{W}}^2 \tilde{P}(\vec{k}) \otimes \left| w(\vec{k}) \right|^2 , \tag{3.7}$$

with $w(k)$ being the Fourier transformed window function.

The additional power term is proportional to the square of the error in the luminosity calibration σ_{zp}.

To test how well the relation holds in a simulated environment, we calculate the additional power by subtracting $\tilde{P}(k)$ from $\hat{P}(k)$ with 5% zero-point errors. As the sigma-ratio between the 10% and 5% error case is 2, we multiply the additional term by 4 and add it
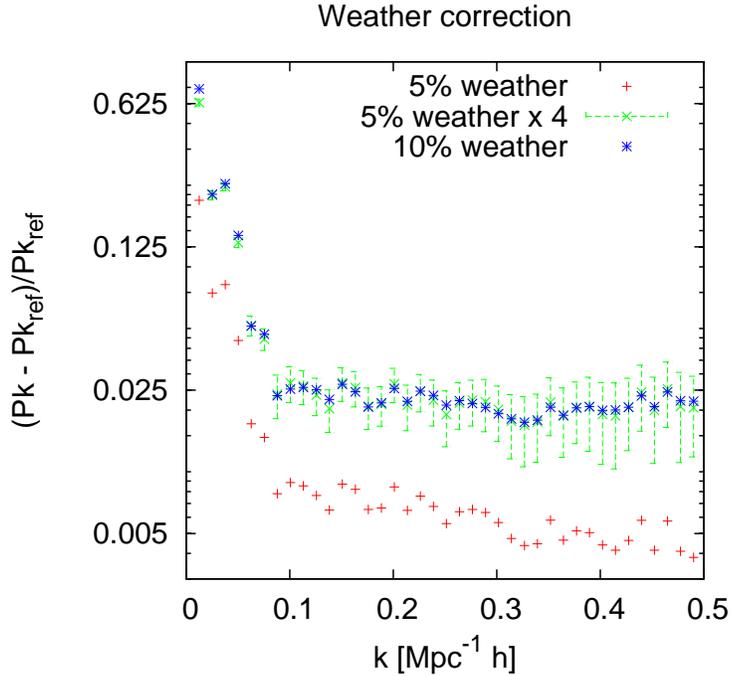
Figure 3.9: Ratio between the power spectrum with 5% zero-point errors (+), the 10% zero-point errors (*) and the IFU-selected power spectrum. The crosses with errorbars show the 5% data corrected in amplitude to match the 10% data.

to $\tilde{P}(k)$. The result should match the 10% error case. Figure 3.9 shows the excellent agreement of the original 10% zero-point error power spectrum and the model. The agreement is better than the random errors of the 400 cubes. These are not affecting this comparison, as both the 5% and 10% case share the same error in the initial power spectrum, due to cosmic variance and shot noise, but not in the statistically independent zero-point errors.

As the correction is better than the statistical error of the initial power spectrum, the impact of the zero-point effects can be corrected for and thus neglected in the cosmological test as long as the shape of the error distribution is well known. While a different shape is changing the form of the additional power term, the integral, measured to first order by σ_w, is proportional to the amplitude of the additional term. In the case of HETDEX the error of the luminosity calibration should be known to about 1%.

3.4.2 Statistical Errors

Using simple mode counting arguments, the relative error of a power spectrum $\sigma_P(k)$ is given by Eqn. 3.1.

Additional errors are introduced by mode coupling and mixing due to redshift and non-linear growth effects. Both of them are, to some extent, included in the semi-numerical simulation. Note that, because of the high redshift range of HETDEX, these errors are very small. More errors are introduced via correlations with the window functions and

zero-point effects. In this section, we will investigate the impact of the various selection effects on the absolute error.

The error of the error can be calculated as,

$$\text{Stddev}(\sigma) = \sigma \frac{1}{\sqrt{2N}} \, . \tag{3.8}$$

With $N = 400$ cubes, this results in a relative error accuracy of 3.5% per data point, or a cumulative 0.4% for all 40 data points of one power spectrum to a maximum k value of $0.3 \, h \, \text{Mpc}^{-1}$
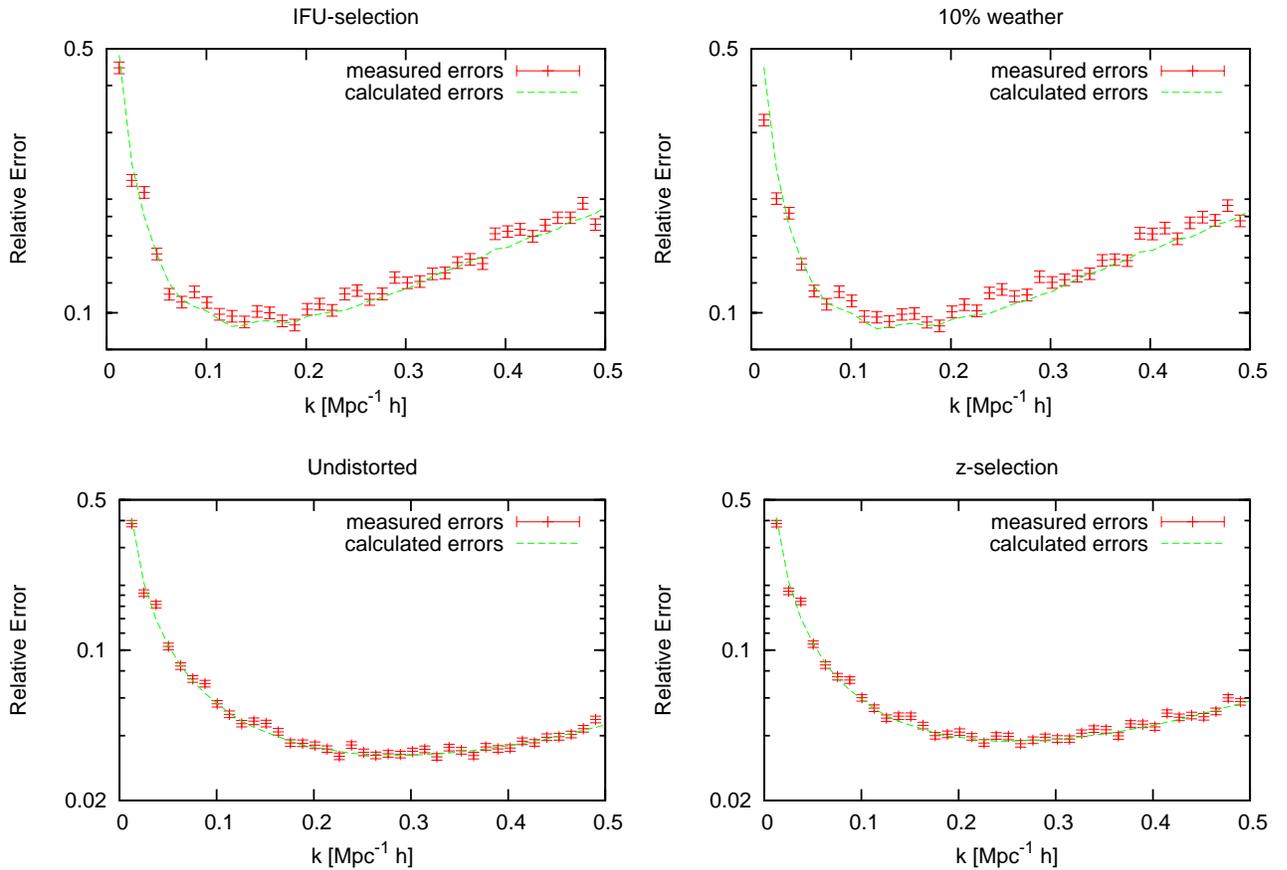
Figure 3.10: Relative errors calculated from Eqn.3.1 (solid line) and estimated from the numerical sample (data points). The error bars are given by Eqn. 3.8. From left to right and bottom to top: Undistorted power spectra, z-selected power spectra, IFU-selected power spectra, power spectra with 10% zero-point errors.

Figure 3.10 shows the errors computed using the statistical variance of the simulations themselves and the errors predicted by Eqn. 3.1. We plot the errors of the undistorted spectra in the lower right panel, and errors of a power spectrum with radial selection effects in the lower right panel. The upper left and right show spectra with additional

angular selection effects and 10% zero-point effects, respectively. The absolute error of the simulations is increasing from left to right and bottom to top, because more and more particles are excluded due to selection and thus increase shotnoise errors.

Because of the high redshift of $z = 2$ of the simulations, redshift-space effects and non-linear growth do not affect the error budget of the power spectrum much, and the undistorted redshift-space power spectrum matches the predictions perfectly, at least at scales relevant for HETDEX up to $k = 0.3\,h\,\mathrm{Mpc}^{-1}$. The radial selection function also does not introduce considerable errors, while the angular selection function increases the errors by a relative mean of 5%. The zero-point effects do not further alter the error budget considerably.

For a cosmological test this means that a relative increase of the error by 5% translates directly into a decrease of accuracy by a relative 5%.

3.4.3   Correlation Matrix

The correlation matrix gives the correlation of a mode k with another mode k', with the case of total correlation normalized to 1 and total anticorrelation normalized to -1. We calculate the mean correlation matrix $\rho(k, k')$ by calculating the mean power $\langle P(k) \rangle$ and the error $\sigma(k)$ from the $N = 400$ cubes, compute the correlation matrix for one cube, and average over all the cubes i using

$$\rho(k, k') = \frac{1}{N} \sum_{i=1}^{N} \frac{(\langle P(k) \rangle - P_i(k))\,(\langle P_i(k') \rangle - P(k'))}{\sigma(k)\,\sigma(k')}\ . \tag{3.9}$$

Figure 3.11 shows the correlation matrices of the various samples. Unfortunately the accuracy of each element in the correlation matrix is, given Eqn. 3.8 and Gaussian error propagation, only 5%. Figure 3.11 clearly shows fluctuations on the order of 5%.

Within the error bars, the z-selection does not have a considerable effect on correlations. This was expected, as the selection in line of sight direction only affects one third of the modes, in contrast to two thirds for the angular selection function. The function itself is quite smooth with uncorrelated modulations due to the spectrum of the night sky which modulates the selection in z-direction. Further resolution effects of the instrument can be neglected as z-resolution scale is far smaller than any scale of interest. The angular window function introduces noticeable correlations along all length scales. This agrees well with the assumption that the strong correlations form the angular window function are washed out in Cartesian space and distributed over all length scales. The correlations due to zero-point effects seem to increase rapidly with higher values of k. The main correlation length of the volume covered in one night is $60\,h^{-1}\,\mathrm{Mpc}$ which corresponds to a k value of $0.1\,h\,\mathrm{Mpc}^{-1}$. All scales smaller than this length should have the correlations from the angular window enhanced, as the zero-point distortions are working on exactly that pattern.

The lower panel of Fig. 3.11 is zoomed in on the k values relevant for HETDEX. Here, the correlations are dominated by the IFU selection function, while the effect of both z-selection and zero-point effects is small. The increase in correlation due to IFU-selection
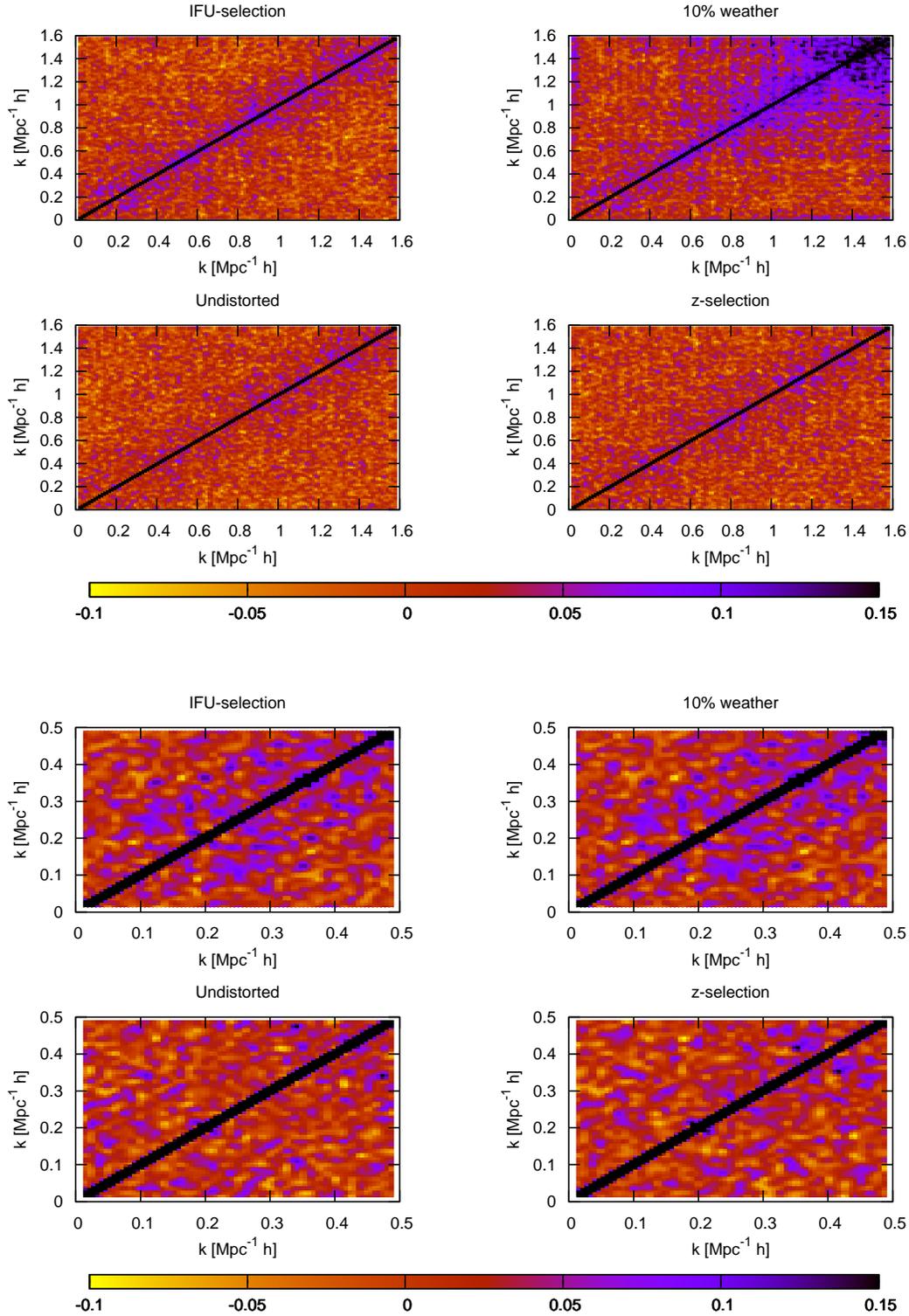
Figure 3.11: **Upper Panel:** Correlation matrices of, from left to right and bottom to top, undistorted power spectra, z-selected power spectra, IFU-selected power spectra, and power spectra with 10% zero-point errors.
Lower Panel: Same correlation matrices, but zoomed in on the region relevant for HETDEX.

at these scale seems to, at least with the resolution provided by the 400 cubes, affect all scales uniformly with a mean increase in correlation to about 0.04.

3.5 Cosmological Test

A wide variety of dark energy models are currently proposed (see Copeland et al., 2006) with various parametrization of the equation of state of dark energy, w, and its behavior in time. Models include, but are not limited to, standard dark energy (Carroll et al., 2003), brane world models (Sahni & Shtanov, 2003), or quintessence (Caldwell & Linder, 2005). Because the time dependent as well as the time independent characteristics of w are affecting the power spectrum as a function of redshift, we do not test for w directly, but only calculate the impact on geometry and growth. Given a certain dark energy model one can then convert these results to accuracies on w and its various parametrizations.

The effect of w on geometry is measured by calculating the goodness of fit from the data and a reference cosmology scaled by a dimensional scale factor, s. The effect on growth is calculated by computing the goodness of fit between the data and a reference cosmology with an amplitude scaled by a growth ratio d.

3.5.1 Geometric Test

The geometric cosmological test is based on the fact that the distance of an object on the sky both in radial and in angular direction is a function of cosmological parameters, like the dark energy density Ω_Λ, the total matter density, Ω_{m}, and the equation of state of dark energy, w. Surveys only measure the redshift, z, and the angular separation, ϕ, of objects. The co-moving line of sight distance, $x_{\parallel}$, and the co-moving angular distance, $x_{\perp}$, are calculated from redshift and angular separation:

$$x_{\parallel} = \int_{z_1}^{z_2} \frac{dx}{dz} \, dz \, , \tag{3.10}$$

$$x_{\perp} = \phi \int_0^{z_2} \frac{dx}{dz} \, dz \, , \tag{3.11}$$

using the relation,

$$\frac{dx}{dz} = \frac{c}{H(z, \Omega_\Lambda, \Omega_{\mathrm{m}}, w)} \, . \tag{3.12}$$

with $H(z)$ being the redshift dependent Hubble constant.

In the case of a flat universe, the Hubble constant can be approximated as,

$$H(z) = H_0 \sqrt{\Omega_m \, (1 + z)^3 \, + \, \Omega_\Lambda \, e^{3 \int_0^z \, d\ln(1+z') \, [1+w(z')]}} \, , \tag{3.13}$$

with H_0 being the Hubble constant measured today.

To perform a geometric test, redshifts and angular separations have to be converted into into co-moving distances using a test cosmology. A power spectrum can then be calculated from the resulting positions and the result compared to a template power sectrum calculated from theory using a reference cosmology.

As this is very time consuming, this process can, as shown in Koehler et al. (2007) be emulated by scaling only the resulting power spectrum. The ratio in scale length s is a function of the ratios in parallel and perpendicular direction

$$s\left(\Omega_\Lambda, \Omega_\mathrm{m}, w\right) = \frac{1}{3}\frac{\hat{x}_\parallel}{x_\parallel} + \frac{2}{3}\frac{\hat{x}_\perp}{x_\perp} , \tag{3.14}$$

with $\hat{x}$ denoting the length scale of a test cosmology and x being the length scale of the reference cosmology. Although the parallel and perpendicular modes can not, due to the braking of the distant observer relation, be directly translated into the Cartesian coordinates of the Fourier transform, Wagner et al. (2008) has shown that the difference can be neglected in HETDEX like surveys, where only a small fraction of the sky is sampled.

The power spectrum itself scales with the scale ratio s both in frequency and in amplitude

$$P(k, s) = s^{-3} P\left(k\, s\right) . \tag{3.15}$$

As the power spectrum declines at higher k values like a power law with a slope of -1.8, the scaling relation is dominated by the decline in amplitude and not by the phase contribution.

3.5.2   Growth Test

Cosmological parameters like w are also affecting linear structure growth, by delaying or amplifying the growth rate, and thus the amplitude of the power spectrum, independently from the geometrical effect. To measure both the geometrical part and the linear growth D, we include the ratio of structure growth factors

$$d\left(\Omega_\Lambda, \Omega_\mathrm{m}, w\right) = \frac{\hat{D}}{D} , \tag{3.16}$$

where the reference growth factor D is determined by the reference cosmology and $\hat{D}$ by the cosmological parameters of the test cosmology. D can be approximated (see Linder & Jenkins, 2003) like

$$D\left(\Omega_\Lambda, \Omega_\mathrm{m}, w\right) = \frac{5}{2}\Omega_\mathrm{m}\, H(z) \int_0^a da\left[a\, H(z)\right]^{-3} , \tag{3.17}$$

with the scale factor $a \propto (1 + z)^{-1}$. The power spectrum itself scales with the growth ratio d only in amplitude

$$P(k, d) = d^2\, P(k) . \tag{3.18}$$

In this case, the scaling relation is purely determined by amplitude.

3.5.3 Test Setup

Combining both the geometrical and growth scaling relations we end up comparing a test power spectrum

$$P^*(k, s, d) = d^2 \, s^{-3} \, P\left(k\,s\right) \, , \tag{3.19}$$

with the reference power spectrum $P(k)$ depending on various cosmological parameters, including biasing b, structure growth D, matter and dark energy content Ω_{m},Ω_{Λ}, and the normalization σ_8 in addition to some other parameters.

We use a simple χ^2-test to calculate to goodness of fit between the reference and the test cosmology. The power spectral template is computed using CAMB (Lewis et al., 2000).

Table 3.2: CAMB Parameters

Par	Value	Description
w	-1	equation of state of DE
Ω_{Λ}	0.75	dark energy density
Ω_{m}	0.25	matter density
Ω_{b}	0.045	baryon density
H100	0.72	normalized Hubble constant
z	0.0	final redshift
σ_8	0.79	power normalization
τ	0.087	optical depth

The errors and the effects of the various selection functions are estimated to first order by computing the error $\sigma(k)$ according to Eqn. 3.1 an multiplying them by the ratio of simulated to computed error given in Fig. 3.10 to include errors introduced by non linearities, redshift-space, and of course the selection functions. The covariance matrix cov_{ij} is calculated using $\mathrm{cov}_{ij} = \sigma_i \, \rho_{ij} \, \sigma_j$.

The final χ^2 test is then computed using:

$$\chi^2 = [P^*(k) - P(k)]^{\mathrm{T}} \, \mathrm{cov}^{-1}(k, k') \, [P^*(k') - P(k')] \, . \tag{3.20}$$

We only include the modes to a maximum k value k_{max}. It is suggested by Jeong & Komatsu (2006) that at redshift two, the maximum reliable wavelength is $k_{\mathrm{max}} = 0.3, h\,\mathrm{Mpc}^{-1}$.

The test proceeds as follows:

- Calculate reference power spectrum

- Convolute reference power spectrum

- Chose test parameters s and d

- Calculate test power spectrum

- Calculate errors of the test spectrum

- Simulate data with errors using Monte Carlo methods

- Calculate χ^2 value

In our test we rescale the errors to match those of a survey with a volume of $2.7\,h^{-3}\,\mathrm{Gpc}^3$ and 0.8 million particles at a redshift of 2.5 and a mean biasing of 2.5. For simplicity, and because we only want to differentially compare the impact of the selection function, we neglect the fact that the particle densities in a flux limited survey are distributed inhomogeneously and that biasing might change with redshift.

3.5.4   Results

Figure 3.12 shows the result of the cosmological test using the undistorted power spectrum. The two parameters d and s are highly degenerate. Fortunately the impact of w on the scale ratio s and the growth ratio d is, according to most sensible models (see Linder & Jenkins, 2003) correlated in their effect on the power spectrum. When w goes up, the scale ratio goes down, which leads to an increase of amplitude. A more positive w also leads to an increase in the growth ratio and thus an increase of amplitude.

The gray area in Fig. 3.12 illustrates what parameter combinations are allowed with current models of dark energy. With s and d being antiproportional in most current models of dark energy, the resulting model curve in the s-d-plane with a negative slope breaks the degeneracy of the test and leads to well constrained values.

To better show the effect of the selection functions on the test, we fix one parameter at 1, which actually underestimates the accuracy because of the anti correlation of the parameters given a model of dark energy, and vary the other. In Figure 3.13 the χ^2 values for the undistorted, the z-selected, the IFU-selected, and the 10% zero-point errors are given. In the top panel of the figure, the results for the scale parameters s are given, while the lower panel shows the results for the growth parameter d. The scale parameter is much better constrained than its counterpart in growth, as the power spectrum has a phase dependence on s and a stronger amplitude dependence than d.

While the impact of the z-selection on both parameters is barely visible, the IFU-selection decreases the accuracy of the test by a relative 16%. The 10% zero-point errors decrease the accuracy by 23% relative to the undistorted case.

To test whether the increase of absolute error, or the correlations introduced by the selection functions have the most impact on accuracy, we conduct two additional cosmological tests:

- the best errors (undistorted) and the worst correlation matrix (10% zero-point errors)

- the worst errors (10% zero-point) and the best correlation matrix (undistorted)
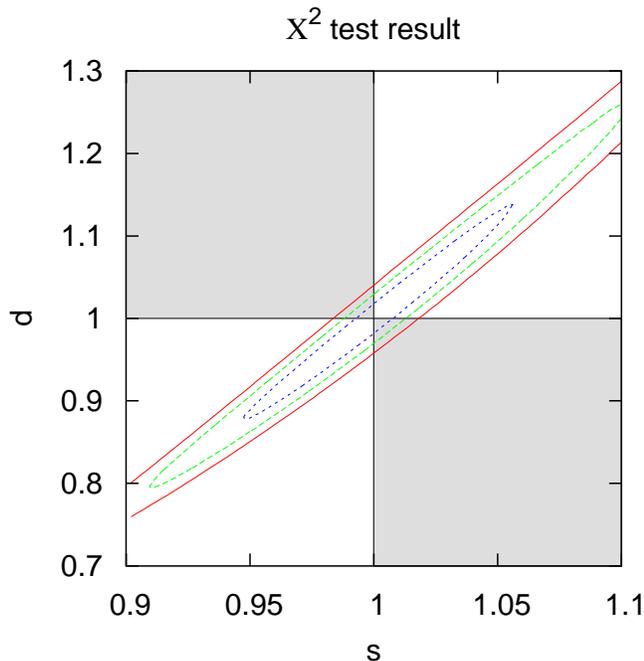
Figure 3.12: Results of the χ^2 test from the undistorted sample with one (short dashed line) two (long dashed line) and three (solid line) sigma contours. The gray regions mark parameter combinations allowed by current models.

Figure 3.14 shows the results. One can clearly see, that the effect of the correlations is about two times as strong as the absolute increase in error. This means that sparse sampling will impose a definitive limit on future very accurate surveys as the error introduced by correlations approaches the scale of the random measurement errors.

Finally we want to asses the impact of systematic errors on the test. We add four of the most basic types of systematic errors to the data sample with 10% zero-point errors:

- systematic overestimation of the true power spectrum by 5%

- systematic underestimation of the true power spectrum by -5%

- a k dependent under to over estimation from -10% to 10% leading to a shallower slope compare to the true power spectrum

- a k dependent over to under estimation from 10% to -10% leading to a steeper slope compared to the true power spectrum

Figure 3.15 gives a visual representation of the applied errors. In all three cases the summed relative deviation from the true power spectrum is the same. However, as different data points have different random errors, the systematic errors will not have the same weight. The effect on the χ^2 test are shown in Fig. 3.16. One can clearly see, that in
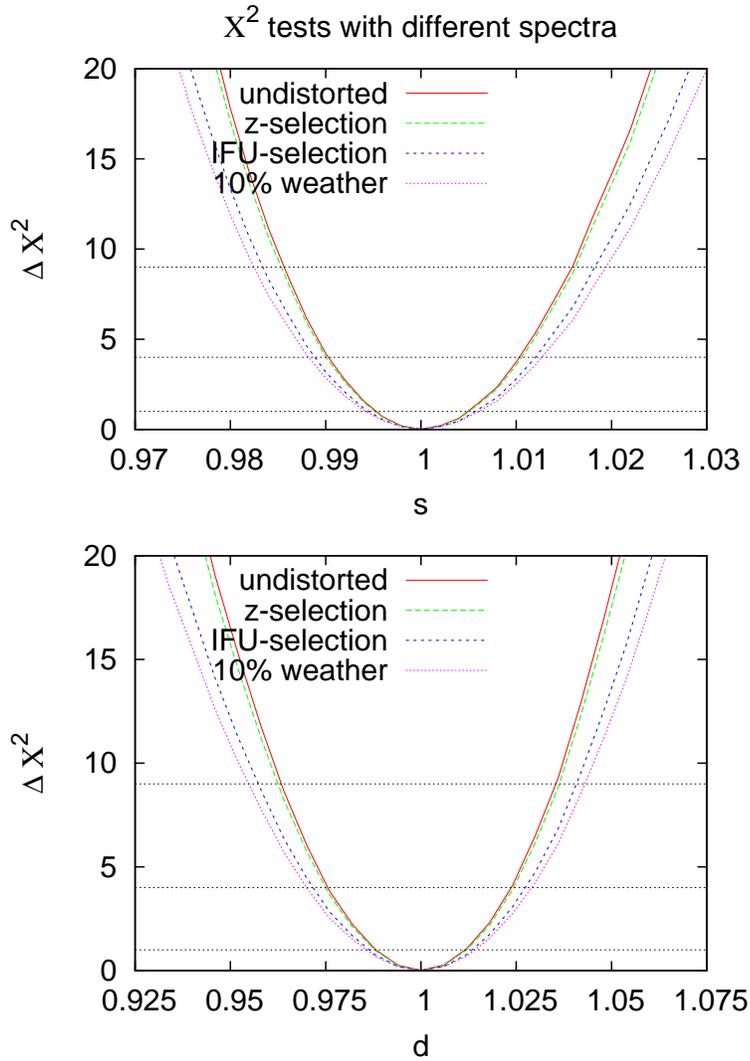
Figure 3.13: Results of the χ^2 test with d parameter fixed at 1 (upper panel) and the s parameter fixed at 1 (lower panel). The solid line denotes the results of the undistorted sample, the long dashed line of the z-selected sample, the short dashed line of the IFU-selected sample, and the dash-dotted line of the sample with 10% zero-point effects. Horizontal dashed lines give the one, two, and three sigma levels for the goodness of fit.
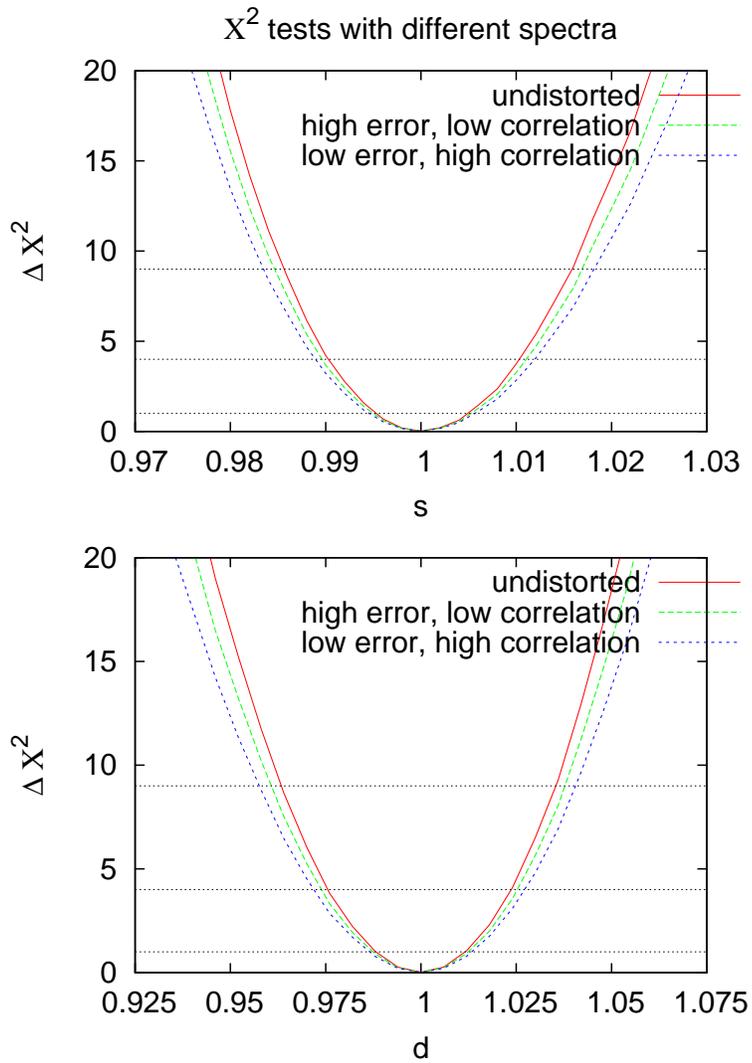
Figure 3.14: Results of the χ^2 test with d parameter fixed at 1 (upper panel) and the s parameter fixed at 1 (lower panel). The solid line shows the results of the undistorted sample, the long dashed line of a test using the worst error of all samples and the best correlation matrix, the short dashed line gives the results for the best errors and worst correlation matrix. Horizontal dashed lines denote the one, two, and three sigma levels for the goodness of fit.
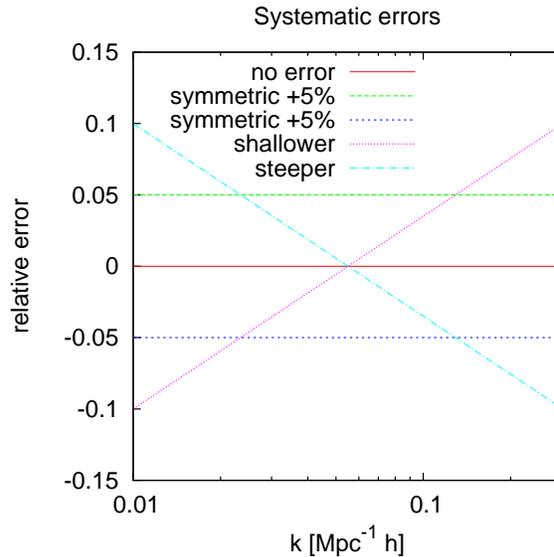
Figure 3.15: The applied systematic errors of a relative 5% (long dashed), -5% (short dashed), k dependent shallower slope (dotted) and k dependent steeper slope (dashed dotted).

both cases the pure over or under estimation leads to the stronger deviations of 2.5% in the estimation of d and 1% in the estimation of s. This is expected as the retrieval of both the scale factor s and the growth factor d relies heavily on amplitude information.

The effect is somewhat weaker in the determination of s, as this fit also uses phase information, compared to d that is determined purely on amplitude. In the case of steeper and shallower slope, the situation is reversed. Here s is, compare to pure under and over estimation, more affected than d. This effect would get stronger if the under and over estimation would follow a power law. In this case it would be degenerate with changes in the phase of the power spectrum and be able to maximally distort this part of the information. For HETDEX, the relative systematic errors should be brought to below 1%, in order to not heavily distort the recovered values for the scale factor and growth factor.

3.6   Conclusions

We simulate 400 cubes with a base line of $500\,h^{-1}\,\mathrm{Mpc}$ at a redshift of 2, to test the effects of various selection functions on the accuracy of a χ^2-test to constrain cosmological parameters like the equation of state of dark energy w.

We show, that the effect of the angular- and z-selection functions on the power spectrum can be corrected for by convoluting the calculated respective window functions with the undistorted power spectrum. This works especially well for zero-point effects if the error distribution is known to better than 10% of the absolute zero-point error. Correcting for the distortions introduced by the radial and angular selection functions works within the errors of the simulations, which are, however, not small enough to be acceptable. Further
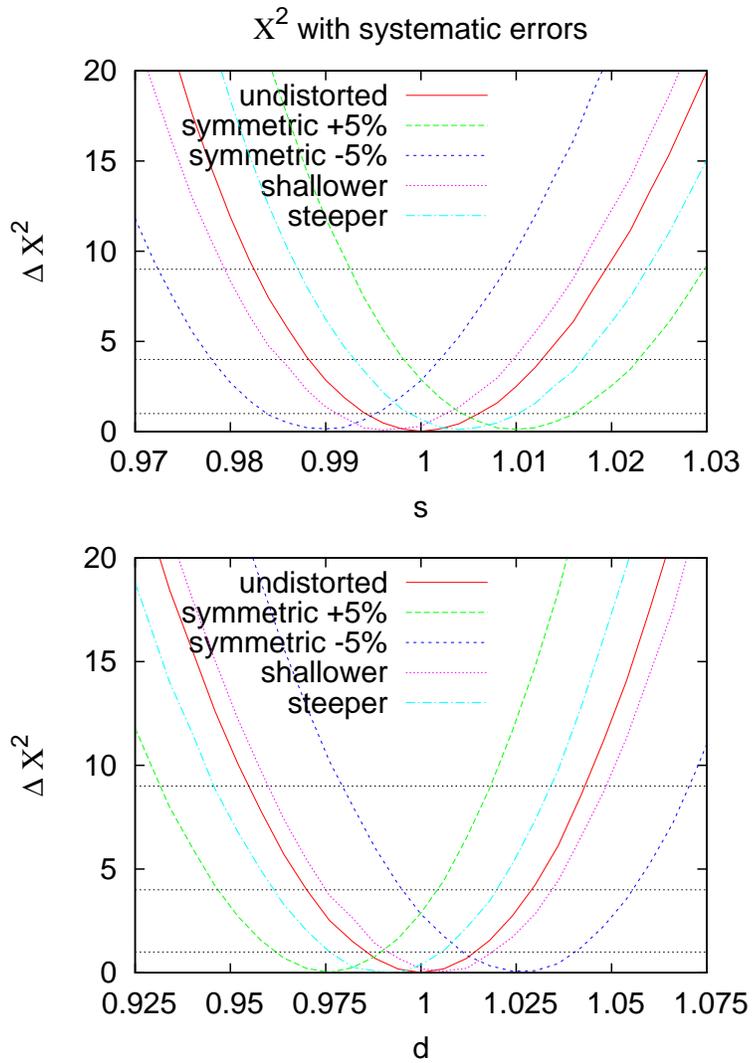
Figure 3.16: Results of the χ^2 test with d parameter fixed at 1 (upper panel) and the s parameter fixed at 1 (lower panel). The values for a test with applied systematic errors of a relative 5% (long dashed), -5% (short dashed), k dependent shallower slope (dotted) and k dependent steeper slope (dashed dotted) are given.

testing is needed, while there are, at least in principle, no conceptual problems to calculate the convolution to arbitrary accuracy.

We calculate the increase in absolute errors, introduced by the selection and correlations. The impact on the errors is small, only a relative 5-10% in the worst case. The effect on a sample cosmological test is a decrease in relative accuracy of about 10%.

The effect of selection on the correlations of the errors is investigated, with the result, that correlations are introduced on the order of 5%, at all length scales and of 15% at very small range scales which are, however, not relevant for the HETDEX. The introduced correlations have a noticeable effect on the accuracy of the cosmological test, reducing the accuracy by a relative 20%.

Neither the window function, nor the zero-point errors introduce systematic offsets in the recovered s and d values, assuming perfect knowledge of the window function and zero-point error distribution. In order to avoid errors from systematics, HETDEX should correct the various distorting effects to an accuracy of about 1%.

Chapter 4

Estimating galaxy bias in pencil-beam surveys with low point densities

Note: this chapter will be submitted as Koehler 2009. The work was performed under the supervision of Niv Drory.

4.1   Introduction

The properties of high redshift galaxies are, though under investigation form various groups (see Delliou et al., 2006; Ouchi et al., 2008; Gawiser et al., 2006), still poorly understood. Large samples of Lyman-alpha emitters (LAEs) have only recently been detected (see Ouchi et al., 2008) and make statistical analysis of their characteristics, like clustering lengths, luminosity functions, and biasing possible. High redshift galaxies are important not only for our knowledge about the evolution of galaxies and structures in the early universe, but also work as tracers of the underlying dark matter density field to be used in cosmological tests. LAEs have been proposed as one of these ideal tracers (Steidel et al., 1996, 1999; Hu et al., 1998), as their emission line of these early star forming galaxies can be easily detected over large distances. However, pinpointing the statistical properties of these tracer particles is paramount to the reliability for cosmological test using them as probes for geometry and growth. As cosmological tests get more and more accurate, uncertainties in properties like galaxy biasing introduce systematic errors of the order of the statistical fluctuations due to the size of the survey (cosmic variance) and discreteness effects. It is thus important to look into statistical effects of properties of LAEs on the galaxy power spectrum to judge the performance of future experiments like HETDEX (Hill et al., 2008) using LAEs as tracer particles for cosmological tests.

The distribution of Lyman alpha emitters was measured using narrow band imaging (see Ouchi et al., 2008; Gawiser et al., 2007) or long slit spectroscopy of gravitationally lensed regions (see Santos et al., 2004). Orsi et al. (2008) investigated the resulting angular

correlation functions to infer LAE characteristics like the luminosity function and biasing from these surveys. One of the main drawbacks of narrowband studies is the extremely small redshift range of the observed sample. This means that predictions are not only limited to a small redshift range, but also renders the volume covered very small and thus introduces large cosmic variance errors in the resulting sample. Long slit spectroscopy of gravitationally lensed regions on the other hand can not be used for the majority of the available volume on the sky and thus faces the same problems. Recently (Hill et al., 2008) are using integral field spectroscopy in the VIRUS-P survey to sample large volumes at a wide range of redshifts ($1.8 < z < 3.8$) for LAEs. The volume of the final sample will be about 8 times bigger than the volume of the two biggest surveys to date, the MUSYC (Gawiser et al., 2006) and the SXDS survey (Ouchi et al., 2008, 2005), combined.

In this work we will investigate the effect of a pencil-beam geometry on the ability to measure the amplitude of both power spectrum and correlation function, to recover biasing. Sect. 4.4 we will calculate the difference in accuracy between statistical analysis of a super sample as a whole and the combined analysis of sub samples. In Sect.4.5 the effect of extremely low point densities on the power spectrum and correlation function will be explored. These calculations and simulations will be used to discuss the effect of pencil-beam survey geometries on realistic data in Sect. 4.6, where we look at the shapes of both correlation function and power spectrum, as well as their errors and correlation matrices. Finally a simple χ^2 test will be conducted to investigate the ability of the data recoverd by the three estimators to constrain basing.

4.2   Survey Layouts

The main difference between narrow-band surveys and surveys utilizing integral field spectroscopy units (IFUs) is the geometry of the covered volume. From an observational point of view, three different basic geometries are possible. A cube, a z-slice, and a pencil-beam. All three of them require different methods of analyzing the statistical information content in correlations and impose different problems on the computation.

Cube   A cube can be used when a survey covers large volumes both in angular as well as in radial direction. Imaging surveys using spectroscopic redshifts like SDSS (Tegmark et al., 2004, 2006; Percival et al., 2001b) or 2dF (see Percival et al., 2001a; Cole et al., 2005) produce these kind of galaxy catalogs. In the case of a volume limited survey a cube is the natural geometry to compute correlations using Cartesian coordinates. It allows the calculation of all modes in the volume in all three dimensions, with an optimal window function and minimal error correlations. Compact, cube like surveys have also the advantage of offering the highest information content of two-point statistics per volume, with their large number of correlations growing with a power of 2 with distance.

z-slice   A z-slice is produced by narrow band surveys, that restrict their redshift range by relying on filters with a narrow z-range, or the radiometers probing for the cosmic

microwave background radiation, that is only emitted in a very narrow z-range. As a result, all the detected information lies in the selected z-range. Examples are the MUSYC (see Gawiser et al., 2007) and SXDS (see Ouchi et al., 2008) or the CMB surveys like the WMAP experiment (see e.g. Komatsu et al., 2008) or its predecessor COBE (see e.g. Smoot et al., 1997). Angular power spectra or correlation functions are able to exploit all the information in a z-shell. As they throw away all the correlations of the density field in the radial direction, they are ideal for computing survey volumes which are of very small thickness. z-slices provide the second highest information content within the compared geometries, with the number of correlations growing linearly with survey extent in the mostly two dimensions of the survey.

Pencil-Beam   A pencil-beam is produced by IFU surveys today but have been produced in the past using photometric redshifts (see e.g. Broadhurst et al., 1990; Fernandez-Soto et al., 1999; Heidt et al., 2004) or spectroscopic redshifts using long-slit or multi-object spectroscopy (see e.g. Cowie et al., 2004; Noll et al., 2004). As modern IFUs tend to have small fields of view compared to imaging surveys, the angular dimensions of the resulting surveys are small. On the other hand, they detect LAEs spectroscopically in a wide range of redshifts, giving an IFU survey a very elongated geometry. A pencil-beams information can not be calculated using angular correlation functions or power spectra as most of the information lies within the radial correlations of the density field, that are thrown away by purely angular statistics. One dimensional statistics are best suited to this survey geometry. Pencil-beams offer the lowest information content for two point statistics, as the number of correlations stays constant with distance in the pencil-beams largest dimension.

All three of the survey techniques mentioned are able to trade survey volume against point density and vice versa by either making longer exposures (and thus detect fainter galaxies) to increase point density, or making shorter but more numerous exposures to increase the survey volume. The function translating time into point density and volume is highly non-linear and often has sweet spots where longer or shorter exposures are not desireable. The maximum volume is limited by the luminosity of the brightest galaxies, while the point density is limited by the number of galaxies and the slope of the luminosity function.

Sparse sampling techniques (see Kaiser, 1986; Colombi et al., 1998; Blake et al., 2006) allow for a direct trade of volume for point density, but also allow a survey to vary geometry. A narrow band survey can use different filters to extend its size in the z-direction, while an IFU survey can set its exposures further apart to increase the size in angular dimension. The change from one geometry to another does not only change the information content by altering the cosmic variance and point density parameters, but also influences the choice of estimator. In principle each survey geometry would need its own tailored estimator like a Karhunen-Loeve Eigenvalue estimator (see e.g. Tegmark et al., 1997) as implemented by Tegmark et al. (2002).

4.3   Estimators

In this work, we will investigate the performance of three different estimators to measure clustering strength and difference between dark matter and galaxy clustering, the galaxy bias. Two of the estimators are different approaches to measure the power spectrum, the other one is the two point auto correlation function.

4.3.1   Power Spectrum

We measure using the FKP (see Feldman et al., 1994) algorithm to compute the dark matter power spectrum $P_{\mathrm{DM}}(k)$ of overdensities in $\delta_g(r)$:

$$
\begin{aligned}
P_{\mathrm{DM}}(k) &= \frac{(2\pi)^3}{V_k} \int_{V_k} d^3k\, \delta_D^3(\vec{k} - \vec{k'}) \, < |\delta(\vec{k})\, \delta^*(\vec{k'})| >\,, &\quad (4.1)\\
&= \frac{1}{N_k} \sum_{k<|\vec{k}|<k+\delta k} |n(k) - \bar{n}(k)|^2 - P_{\mathrm{shot}}\,. &\quad (4.2)
\end{aligned}
$$

with n_k, $\bar{n}_k$ being the Fourier transformed galaxy density field and average density and N_k the number of modes averaged in a k-shell. P_{shot} is the shot noise due to discreteness effects.

The observed power spectrum $P_{\mathrm{obs}}(k)$ is convolved with the window function $|w(k)|^2$ of the survey. It is the Fourier transform of the product of the the radial, line of sight window function introduced by the wave-length dependent throughput of the instrument and the angular window function due to the cone like observation window. The observed power spectrum is related to the true underlying dark matter power spectrum by the relation:

$$
P_{\mathrm{obs}}(k) = \left[b^2(k)\, P_{\mathrm{DM}}(k) \right] \otimes |w(k)|^2\,. \qquad (4.3)
$$

The biasing function $b(k)$, which connects the dark matter overdensities δ_m to the observed galaxy overdensities δ_g, can be expanded into a linear and scale dependent term:

$$
\delta_g(k) = b(k)\, \delta_m(k) = \hat{b}\, [1 + \tilde{b}(k)]\, \delta_m(k)\,. \qquad (4.4)
$$

We consider two ways to grid the data for the Fast Fourier Transform (FFT).

- (A) In a cube of the length of the pencil-beam survey

- (B) In a cuboid with the length of the pencil-beam survey width and height equal to the maximal expansion of the survey in each direction on the sky

The two methods are distinct in various ways: In case A, the fundamental mode of the Fourier transform will be of equal size in all dimensions, while in case B, the fundamental mode varies with dimension. This means, that the volume each mode occupies in the cuboid will also be elongated, leading to a non-standard sampling of k-space. The sampling in

the cube will be regular as normal. The cube will, however, be mostly empty, as only a small fraction of the volume is occupied by galaxies. This will lead to a extremely shallow window function and strong correlations in the errors and data. The cuboid will not suffer from these correlations, as the modes are, with the exception of edge effects due to the wedge like geometry, independent. As the shot noise is directly proportional to the filling factor, the cube will suffer from much higher shot noise than the cuboid, which will mostly retain the shot noise of the original sample.

To carry out the FFT we use the well tested FFT libraries FFTW (2008).

4.3.2 Correlation Function

The correlation function, the Fourier transform of the power spectrum, is measured using the algorithm proposed by (Landy & Szalay, 1993) that was shown to have almost Poissonian variance:

$$\xi(\vec{r}) = \int d^3k\, P(\vec{k})\, e^{i\vec{k}\cdot\vec{r}}, \tag{4.5}$$

$$= \frac{<DD(r) - 2DR(r) + RR(r)>}{<RR(r)>}. \tag{4.6}$$

$DD(r)$, $DR(r)$, and $RR(r)$ are the normalized data-data, data-random, and random-random pair number counts per distance r. We chose to use a logarithmic binning in this work to better sample the correlation function at small scales, but will increase bin size and thus decrease the error along r.

4.4 Analytical Error Analysis of 3D Power Spectra

Both the volume and the point density are critical for the estimation of the information content of a survey. The smaller a survey is, the higher the probability that the local realization of the overall density shows random deviations from the real mean of the whole sample. The lower the point density is, the higher the noise from the Poisson process distributing the discrete galaxies on the continuous density field gets.

The errors of an averaged 1- dimensional power spectrum without significant contribution from the window function can be estimated analytically using simple mode counting. In the case of a one dimensional power spectrum, the error is proportional to the number of modes in a k-shell of the original three dimensional power spectrum. The volume V_k of the shell depends only on the length k of the vector $\vec{k}$ and the thickness of the shell Δ,

$$V_k = 4\pi\, k^2\, \Delta. \tag{4.7}$$

To minimize correlation between the different modes, the binning should be at least of the size of the fundamental mode Δk:

$$\Delta k = \frac{2\pi}{L} \,. \tag{4.8}$$

In the following calculations we set the binning, and thus the thickness of a k-shell Δ equal to the fundamental mode Δk. The mode density m_k is then simply given by the size of the bins,

$$m_k = \frac{1}{(\Delta k)^3} \,. \tag{4.9}$$

The relative error $\sigma(k)$ of the mode average, the square root of the number of modes M_k per k-shell, becomes:

$$\sigma(k) = \frac{1}{\sqrt{M_k}} = \frac{1}{\sqrt{V_k \, m_k}} = \frac{\sqrt{2\pi}}{Lk} \,. \tag{4.10}$$

It is useful to calculate the difference in errors when computing a power spectrum from a full cube compared to computing the spectrum from small volume sub-samples of the cube and then combining them. To do this we divide the full cube with a length L_1 into x^3 smaller cubes with a length of $L_x = L/x$. As we are considering biasing we want to assess the ability of the measured data to constrain the amplitude of the power spectrum in a k range defined by $k_{\min} < k < k_{\max}$. Assuming the data is optimally binned in multiples of the fundamental mode Δk of the respective cube we can calculate the error of the amplitude estimation $\sigma_A(x)$, by adding up all data points in the relevant k region:

$$\sigma_A(x) = \frac{1}{(N_x - n_x)} \sqrt{\sum_{i=n_x}^{N_x} \left[\sigma \left(i \Delta k_x \right) \right]^2} \,, \tag{4.11}$$

with the first and last data point n_x and N_x given by

$$\Delta k_x = \frac{2\pi}{L_x} = \frac{x\,2\pi}{L_1} \,, \tag{4.12}$$

$$n_x = \frac{k_{\min}}{\Delta k_x} = \frac{n_1}{x} \,, \tag{4.13}$$

$$N_x = \frac{k_{\max}}{\Delta k_x} = \frac{N_1}{x} \,. \tag{4.14}$$

Using these definitions, Eqn. (4.11) can be further simplified,

$$\sigma_A(x) = \frac{1}{(N_x - n_x)} \sqrt{\sum_{i=n_x}^{N_x} \left[\frac{1}{i\,\sqrt{2\pi}} \right]^2} \,. \tag{4.15}$$

The same data points can be found in the smaller cubes as well as the larger cubes. They differ only in number, as well as in the value of the error. Using Eqn. (4.11) we can form the relation

$$\frac{1}{N_x - n_x} \sum_{i=n_x}^{N_x} \frac{1}{i} = \frac{x}{N_1 - n_1} \sum_{i=n_1}^{N_1} \frac{x}{i} = \frac{x^2}{N_1 - n_1} \sum_{i=n_1}^{N_1} \frac{1}{i} \, . \tag{4.16}$$

We can now combine all the smaller cubes, adding up their data points using simple Gaussian error propagation to the total error of all measurements Σ_x,

$$\Sigma_x = \frac{1}{\sqrt{x^3}} \sigma_x \, . \tag{4.17}$$

and combine Eqns. (4.15) - (4.16) to arrive at the relation

$$\Sigma_A(x) = \frac{\sqrt{x}}{N_1 - n_1} \sqrt{\sum_{i=n_1}^{N_1} \left[\frac{1}{i \sqrt{2\pi}} \right]^2} \, . \tag{4.18}$$

The total error for the amplitude $\Sigma_A(x)$ only depends on the k-range used and the ratio of cube lengths x. The error goes up with x which means that the smaller the cubes are the more information we lose, and the worse the amplitude of the power spectrum is constrained. This can readily explained by the loss of information which is represented by the correlations between the density fields in the different cubes that are caught by using a large cube, but are thrown away when using lots of smaller cubes.

In the case of a pencil-beam survey one can either separate the volume into smaller parts of cube like geometry or use the whole volume accepting a heavily distorting window function and strongly correlated errors. But dividing the whole volume into smaller fractions does not only neglect information, but also introduces the numerical challenge of very small samples.

In reality we are not dealing with smooth densities but rather with discrete representations of an underlying density field, like galaxies. This introduces a further source of errors, the shot noise. The more subtle fluctuations in a power spectrum are, the more points are needed to sample these differences. Thus the errors introduced by discrete sampling are indirectly proportional to the spectral power of that scale,

$$\sigma(k) = \frac{\sqrt{2\pi}}{Lk} \left[1 + \frac{1}{\bar{n} \, P(k)} \right] \, . \tag{4.19}$$

However, Eqn. (4.18) still holds as both the full sample and the sub sample will have the same point density. This will change when we examine a situation where the survey geometry is not cube like but elongated. In this case, the full sample itself put into a cube will have a much lower point density than the smaller subsamples.

4.5 Low Point Density Performance

In the case of VIRUS-P about 250 LAEs are expected in a volume of about $4 \times 10^5 \, h^{-3} \, \mathrm{Mpc}^3$. The average distance between emitters equals about 10 Mpc. If one was to divide the whole

volume into cubes of equal size, every cube would only contain a couple of LAEs. In this section we want to investigate the impact of extremely low point densities on widely used numerical implementations of both power spectrum and correlation function estimators.

We use a $500^3 \, h^{-3} \, \mathrm{Mpc}^3$ cube from the Hubble Volume simulation (see Sect. 4.6.2) at a redshift of 2.5 and divide it in 1000 subcubes with a length of $50 \, h^{-1} \, \mathrm{Mpc}$. Starting with a density of 0.02 particles per $h^{-3} \, \mathrm{Mpc}^3$ we gradually reduce the density and calculate both power spectra and correlation functions in seven steps. Each step halves the point density. Table 4.1 gives point densities, average particles per cube and average distance of particles in Fourier space to next particle.

Table 4.1: Point densities, shot noise, number of particles and average distance in k-space of seven samples consisting of 1000 realizations each from the Hubble Volume Simulation

Sample #	Point Density $[h^3 \, \mathrm{Mpc}^{-3}]$	Shotnoise $[h^{-3} \, \mathrm{Mpc}^3]$	Particles [per Cube]	k distance $[h \, \mathrm{Mpc}^{-1}]$
1	1.9×10^{-2}	5.3×10^1	2375	1.7
2	9.5×10^{-3}	1.1×10^2	1188	1.3
3	4.7×10^{-3}	2.1×10^2	594	1.1
4	2.4×10^{-3}	4.2×10^2	297	0.84
5	1.2×10^{-3}	8.4×10^2	148	0.67
6	5.9×10^{-4}	3.4×10^3	74	0.53
7	3.0×10^{-4}	6.7×10^3	37	0.42

The results for the Fourier transform are shown in Fig. 4.1. One can clearly see that with decreasing point density, the average power spectrum deviates more and more from the true spectrum at larger and larger k values, as the average distances between particles increases.

The results for the correlation function are shown in Fig. 4.2. In this case the correlations do not, within the error bars, deviate from the true correlation function.

This artificial bias should be taken into account when analyzing surveys with low point densities at small scales with Fourier transforms. Like shot noise, this sampling effect depends mostly on the scale probed and the power in the sample. Thus surveys like HETDEX with an expected 800,000 galaxies in a volume of $2.7 \mathrm{Gpc}^3 h^{-1}$ and a resulting particle density of 2.9×10^{-4} will have to cope with this effect on the percent niveau on k values bigger than about $0.3 \, h \, \mathrm{Mpc}^{-1}$ as their expected biasing is about 2, effectively multiplying the expected power by 4.

4.6   Realistic Data

After investigating the effect of low point densities in an artificial environment, we now want to test the effect of both low point densities and highly elongated surveys on the estimation of correlation function and power spectrum. We use the VIRUS-P survey currently
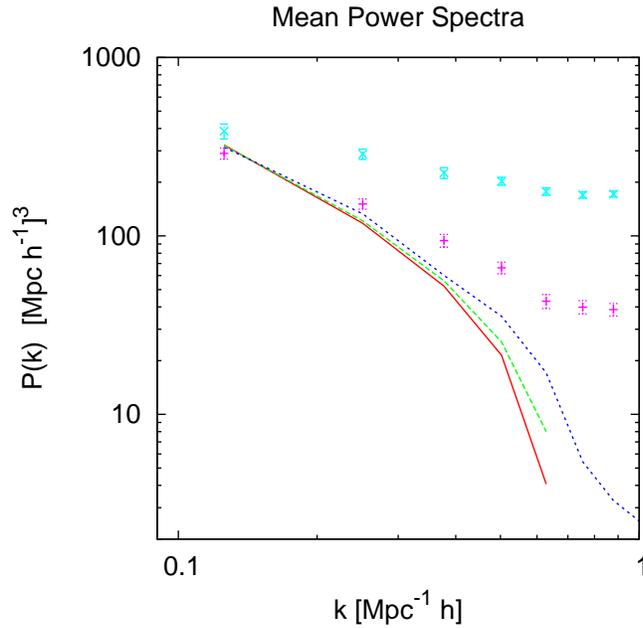
Figure 4.1: The mean power spectra of the 1000 cubes with a cube length of 50 h^{-1}Mpc and an average distance in k of (from bottom line to top data points) 1.7, 0.84, 0.67, 0.53, and 0.42 $h\,\mathrm{Mpc}^{-1}$. The latter two spectra include error bars derived from the sample to visualize the significance of the deviations.
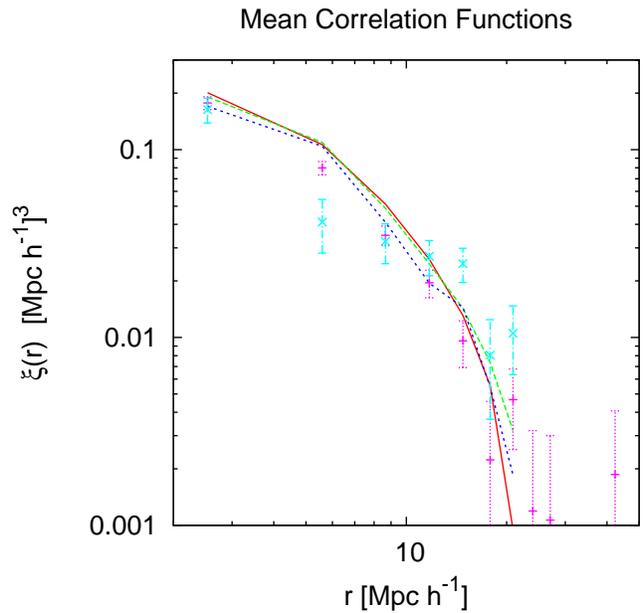
Figure 4.2: The mean correlation functions of the 1000 cubes with a cube length of $50\,h^{-1}$ Mpc and a point density of (analogous to Fig. 4.1) 1.9×10^{-2}, 2.5×10^{-3}, 1.2×10^{-3}, 5.9×10^{-4}, and 3.0×10^{-4} $h\,\mathrm{Mpc}^{-1}$. Error bars are given for the latter two samples.

underway as an example and simulate various geometries and point densities around this configuration.

4.6.1   The VIRUS-P survey

The VIRUS-P survey, a pilot survey for the planned HETDEX experiment, (see Hill et al., 2008), is using an a integral field spectroscopy unit with a field of view of $4.2\,\mathrm{arcmin}^2$ to detect LAEs within a z-range of 1.8 to 3.8. It is going to sample $250\,\mathrm{arcmin}^2$ in three continuous fields of two times 100 and 50 arcmin^2. In this way, the survey will cover a total volume of $4 \times 10^5\ h^{-3}\,\mathrm{Mpc}^3$. The average angular size of one survey field of about $25\ h^{-1}\,\mathrm{Mpc}$ is roughly 60 times smaller than its expansion in the line of sight direction of $1400\ h^{-1}\,\mathrm{Mpc}$.

The radial selection function of VIRUS-P is given in Fig. 4.3. This function includes the throughput of the telescope and decreased detection probabilities at the position of various sky lines.
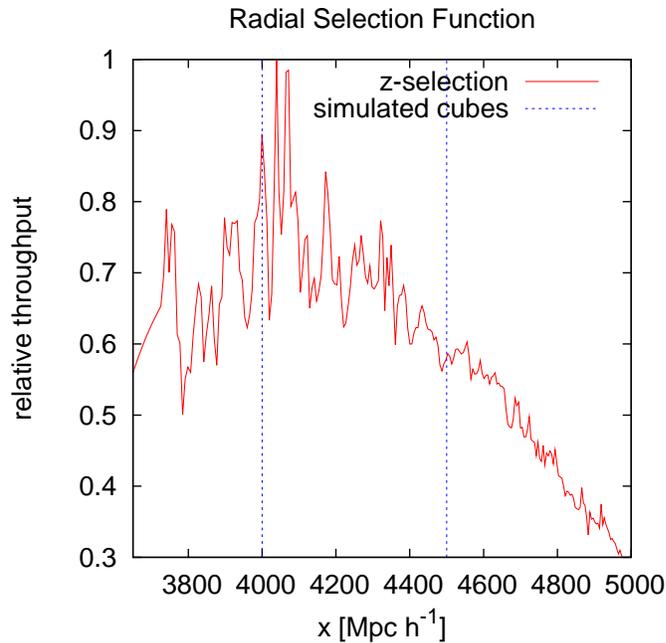
Figure 4.3: The radial selection function includes throughput of the instrument and loss in detection performance due to sky lines.

4.6.2   The Simulation

We are using the Hubble Volume Simulation (see Evrard et al., 2002) to simulate the VIRUS-P survey. It was already used before to simulate pencil-beam redshift surveys (see Yoshida et al., 2001). The Hubble Volume simulation offers light cone data of 10 square degrees at the required redshifts. The light cone data includes the velocities of dark

matter particles to simulate redshift space, proper variations in redshift with distance to the observer, and non-linear growth. Table 4.2 gives the parameters of the length of the simulation box L, the dark energy content Ω_Λ, the dark matter content Ω_m, the Hubble constant h, the amplitude normalization σ_8, and the mass per particle m_p.

Table 4.2: Cosmological Parameters of the Hubble Volume Simulation

Model	L	Ω_Λ	Ω_m	h	σ_8	$m_p(M_\odot/h)$
ΛCDM	3000	0.7	0.3	0.7	0.9	2.25×10^{12}

To investigate the effects of geometry and sampling on the ability to recover the power spectrum and correlation function we generate 5 different samples consisting of 100 realizations of a survey geometry.

A subsample of three samples share the same point density and vary in angular diameter, while another subsample of three samples varies in point density and shares the same angular diameter. All the samples versus the same extension in line of sight direction, resembling the VIRUS-P redshift range between $1.8 < z < 3.8$. See Tab. 4.3 for detailed information.

Table 4.3: Geometrical properties and point densities of the 5 samples consisting of 100 realizations extracted from the Hubble Volume Simulations

Sample #	Angular Size [arcmin^2]	Point Density [$h^3\,\mathrm{Mpc}^{-3}$]	Elongation [$h^{-1}\,\mathrm{Mpc}$]
1	25	2.8×10^{-2}	1370
2	400	2.8×10^{-2}	1370
3	100	2.8×10^{-2}	1370
4	100	2.8×10^{-3}	1370
5	100	5.6×10^{-4}	1370

The samples were generated by dividing the whole Hubble Volume light cone into 100 realizations with the appropriate geometry. All samples were selected using the VIRUS-P window function in radial direction.

As the samples do not contain galaxies, but only dark matter halos, the bias of all samples equals 1. Interactions between biasing and redshift space are not included in this analysis.

4.6.3 Shapes

We compute power spectra using both FFT methods and correlation functions of all samples given in Tab. 4.3. To explore the effect of a different survey geometry, we first use samples 1-3, which vary in angular diameter. Point density is at the constant value of $2.8 \times 10^{-2}[h^3\,\mathrm{Mpc}^{-3}]$ for all the samples. Figure 4.4 shows the mean power spectra of 100 realistions using the cube FFT. The smaller the angular diameter, the shallower the

window function, and the stronger the deviation from the true power spectrum. However, the effect is not most prominent on large scales, as the window function leads to a strong migration of power from larger scales to smaller scales.

The same behavior is found in the estimated power spectrum using the cuboid FFT, as shown in Fig. 4.5. In this case, the smaller scales are not affected by the artificial window function introduced by the elongated geometry, although the larger scales still see the window effect of the wedge like geometry and the selection function. These effects dominate, as expected, in the more elongated samples, as the power spectrum propagates closer to the 1 dimensional case. This was already investigated by (Kaiser & Peacock, 1991). They give the one dimensional power spectrum of a cone-like survey as within the limit of the elongated dimension being much larger than the angular dimension as:

$$P_{\text{obs}} \propto \frac{1}{2\pi} \int_k dk \, k P(k) \,. \tag{4.20}$$

This leads to a projection of small scale power to larger scales and thus a suppression of the power at small k-values. The exact amount of projection depends on the power spectrum itself. Park & Gott (1990) give examples on how large scale power is affected depending on the shape of the spectrum.

The exponential fall-off of the input power spectrum compared to a linear theory spectrum at scales smaller than $0.4 \, h \, \text{Mpc}^{-1}$ can be explained by the initial glass-distribution of the Hubble Volume simulation (see Smith et al., 2003), that leads to less power and a smaller variance at small scales. Furthermore redshift space effects are dampening the power at smaller scales due to the Finger of God effect.

The mean correlation functions for samples 1-3 are plotted in Fig. 4.6. Here the effects of the window functions on larger scales are more prominent. While in the case of the power spectrum the window function lead mostly to a strong correlation along all modes (see Sect. 4.6.6), the effect is more intuitive in the correlation function. The smaller the angular diameter of the pencil-beam, the more correlations are suppressed at larger scales. Another noteable feature is the lack of correlation at very small scales, in this case smaller than $3 \, h^{-1} \, \text{Mpc}$, where the correlation function exhibits a slope of about 0. This is due to the resolution of the Hubble Volume simulation and the afore mentioned glass effect of the initial distribution.

To compare the effect of low point densities on both statistics we compare the samples 3 - 5. All have the same angular diameter of 10 arcmin and vary in point density by a factor of 1/10th and 1/50th. The effect investigated in Sect. 4.5 is also present in the current data set. Figure 4.7 using the FFT cube and Fig. 4.8 using the FFT cuboid both show the same behavior as the test samples without a window function, plotted in Fig. 4.1. Depending on k-value, the power systematically deviates from the true power spectrum. This deviation exhibits the same dependence, even quantitatively, on point density, when compared to the earlier tests.

The results of the correlation function, shown in Fig 4.9, are also consistent with the previous simulations. Within the error bars, no deviations from the true correlation function are found. The functions of the realizations with very low point densities, shown as
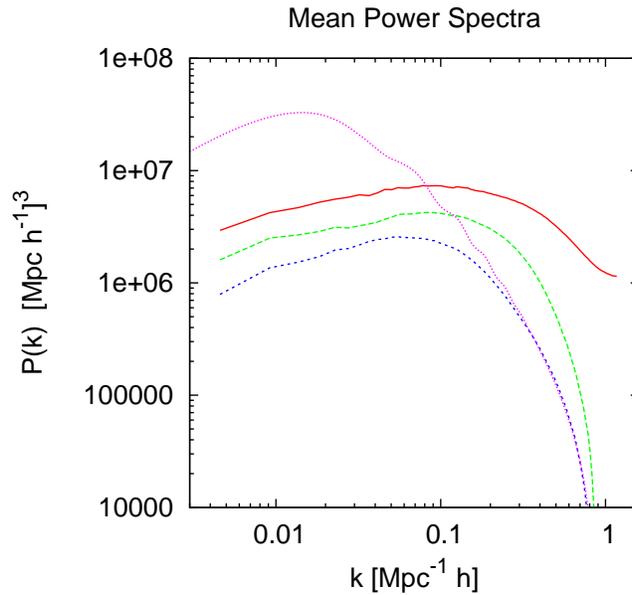
Figure 4.4: Power spectra with different angular diameter using a FFT cube for estimation. The angular diameter size varies from 5 arcmin (about 6 h^{-1} Mpc at a redshift of 2.5, solid line), to 10 arcmin (12 h^{-1} Mpc, long dashed line), and 20 arcmin (24 h^{-1} Mpc, short dashed line). The dotted line gives the re-scaled input power spectrum.
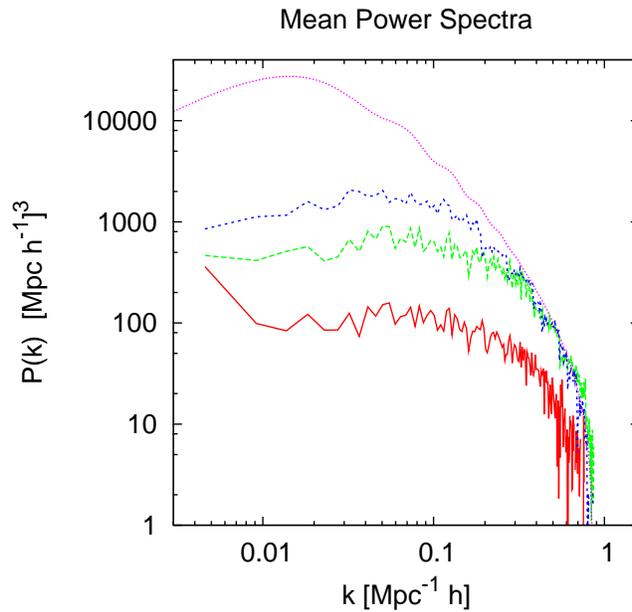
Figure 4.5: Power spectra with different angular diameter using a elongated FFT cuboid for estimation. The angular diameter size varies from 5 arcmin (about 6 h^{-1} Mpc at a redshift of 2.5, solid line), to 10 arcmin (12 h^{-1} Mpc, long dashed line), and 20 arcmin (24 h^{-1} Mpc, short dashed line). The dotted line gives the input power spectrum.
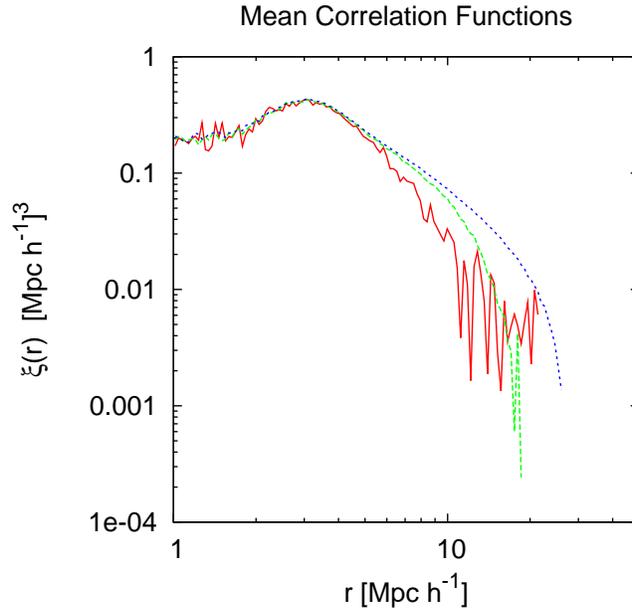
Figure 4.6: Correlation functions with different angular diameter. The angular diameter size varies from 5 arcmin (about 6 h^{-1} Mpc at a redshift of 2.5, solid line), to 10 arcmin (12 h^{-1} Mpc, long dashed line), and 20 arcmin (24 h^{-1} Mpc, short dashed line).
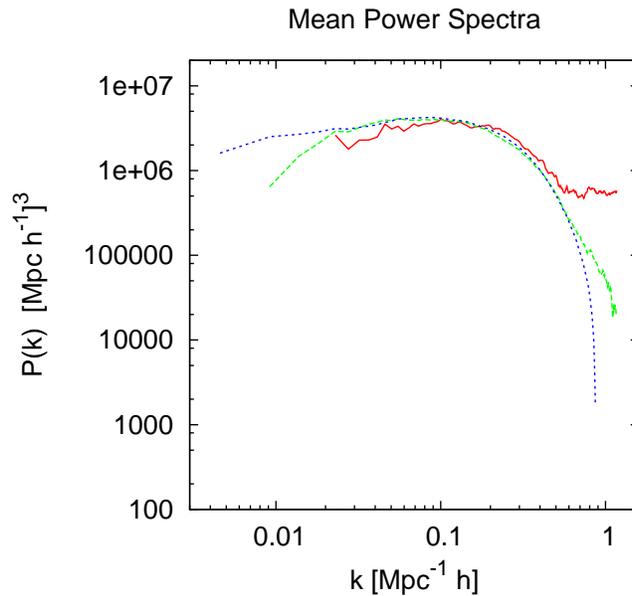
Figure 4.7: Power spectra of the FFT cube with different point densities. The point density varies from 5.6×10^{-4} (solid line), to 2.8×10^{-3} (long dashed line), and 2.8×10^{-2} h^3 Mpc^{-3} (short dashed line).
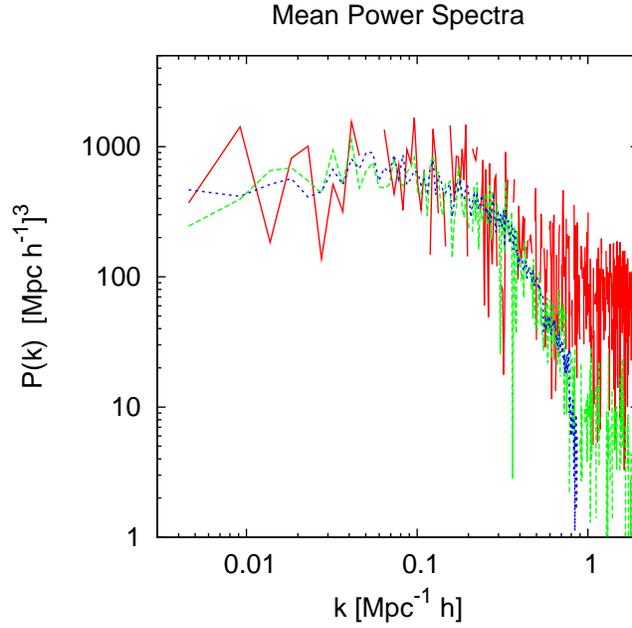
Figure 4.8: Power spectra of the elongated FFT cuboid with different point densities. The point density varies from 5.6×10^{-4} (solid line), to 2.8×10^{-3} (long dashed line), and $2.8 \times 10^{-2} \, h^3 \, \mathrm{Mpc}^{-3}$ (short dashed line).
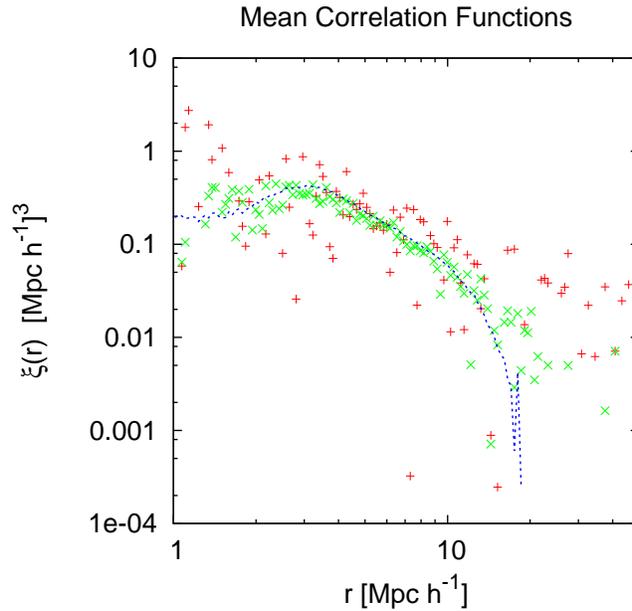
Figure 4.9: Correlation functions with different point densities. The point density varies from 5.6×10^{-4} (+), to 2.8×10^{-3} (x), and $2.8 \times 10^{-2} \, h^3 \, \mathrm{Mpc}^{-3}$ (short dashed line).

dots, seem to have a bias towards higher correlations at larger scales, but this is only an effect of the logarithmic plotting.

4.6.4   Errors

We estimate the errors of the power spectra and correlation functions from the variation of the 100 realizations of each sample itself. The accuracy of the errors given 100 realizations is 7% of the errors. Figure 4.10 gives the relative errors for the samples 1-3.
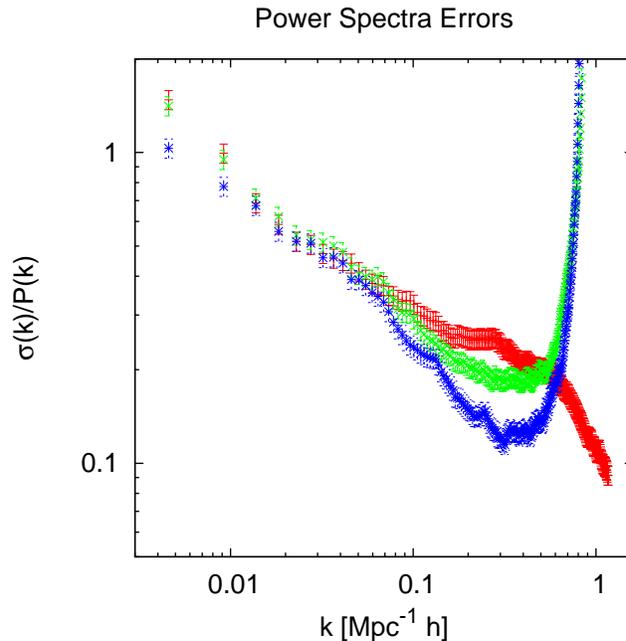
Figure 4.10: The relative power spectral errors derived from the simulations for the samples with (from top to bottom) 5 arcmin, 10 arcmin, and 20 arcmin angular dimension using FFT cubes.

The error behavior of the power spectrum can be roughly divided into two stages: In the first stage the number of modes stays constant as they now see the elongated structure of the survey geometry. In the second stage the number of modes increases quadratically again as the related scales are now small enough to fit into the small dimensions of the volume. In the case of the FFT cube, this transition is continuous, as all three dimensions are sampled with equal resolution. The number of modes in the cube is proportional to the autocorrelation function of a homogeneous sample with a geometry of the survey shown in Fig. 4.16.

In the case of the FFT cuboid, the different dimensions are not sampled with equal resolution. In this case, all the larger k values are sampled an equal number if times, till the k value exceed the fundamental mode of the smaller dimensions. At this point, the transformation instantly sees the two other dimensions and the number of modes in a shell starts to increase with a power of 2. This behavior is shown in Fig. 4.11. In the case of the 20 and 10 arcmin survey, the sudden drop in errors at the size of the smaller dimensions
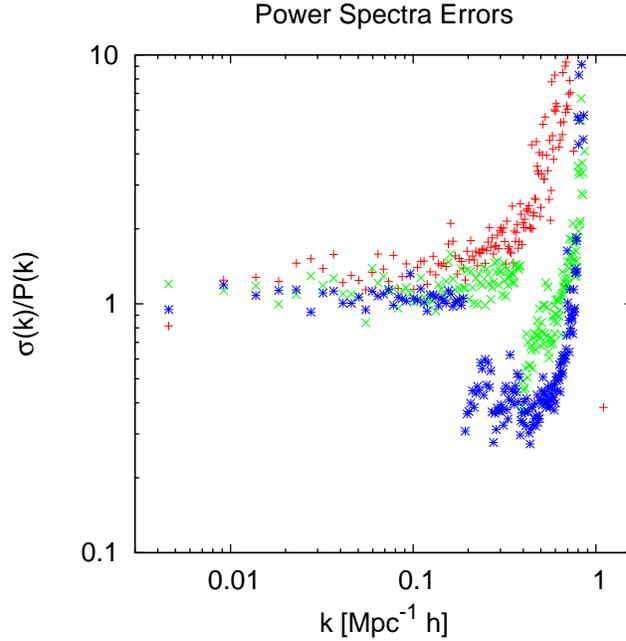
Figure 4.11: The relative power spectral errors derived from the simulations for the samples with (from top to bottom) 5 arcmin, 10 arcmin, and 20 arcmin angular dimension using FFT cuboids.

is clearly visible. In the case of the 5 arcmin sample, the drop is not visible due to the overlay with the onsetting shot noise and exponential fall off of the input power spectrum.

The errors for the correlation function of samples 1-3 are plotted in Fig. 4.12. The aforementioned behavior of two stages can also be found in this representation. However, this time not in the relative errors, but in the absolute measurement errors. On small scales the errors decrease rapidly, as they see all three dimensions of the cube and the number of pairs restricted by the discrete sampling increases. Intermediate scales see only one dimension and thus have mostly the same error. And large scales increase in error, as the number of pairs goes down again. The absolute errors of the correlation function are inversely proportional to the square root of the correlation function of a homogeneous sample (see Fig. 4.16) of the same geometry, as it just gives the number of correlations per radial bin and thus the Poissonian error. Note that the decrease in the errors in this works correlation function is steeper than expected, as we are using logarithmic binning. In this case the bin gets bigger depending on radius and thus adds up more correlations resulting in smaller errors. Using linear binning, the number of correlation pairs in an homogeneously distributed cube grows with the power of 2, while it grows with approximately the power of 3 using logarithmic binning.

Figure 4.13 compares the power spectral errors of sample 1 using the FFT cube (1_{cube}), and the FFT cuboid ((1_{cuboid}) with the calculated errors (using Eqns. 4.18 and 4.19) of two artificial samples:
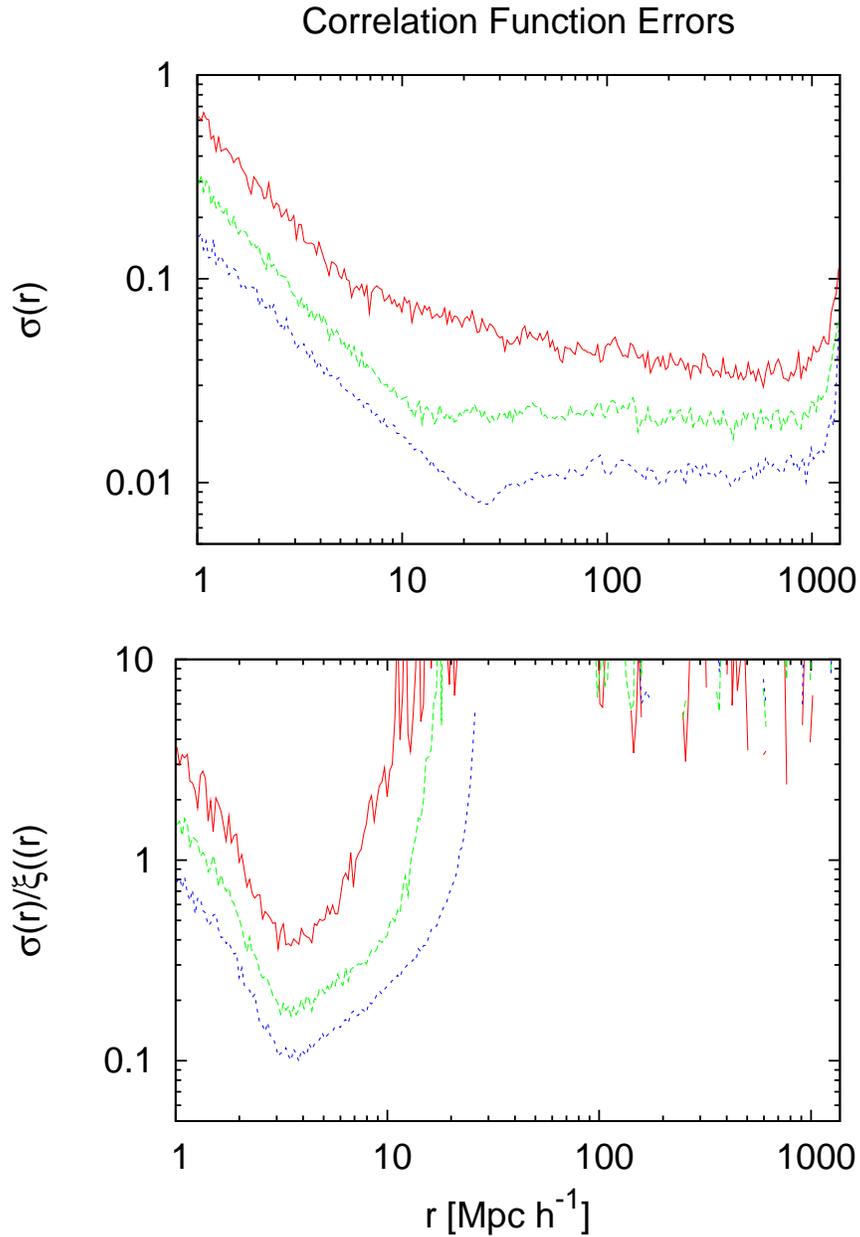
Figure 4.12: The absolute (upper panel) and relative (lower panel) errors derived from the simulations for the samples with (from top to bottom) 5 arcmin, 10 arcmin, and 20 arcmin angular dimension.

Sample A , filling the whole FFT cube of sample 1 with a length of $1370 \, h^{-1} \, \mathrm{Mpc}$, and the same FFT point density, and thus the same number of particles, as sample 1. The point density is much smaller, as the volume is much bigger than the volume of sample 1.

Sample B , 60 FFT cubes with the same angular diameter as the simulated sample (about $22 h^{-1} \, Mpc$), and a total volume and number of particles as sample 1.

Both samples A, and B have the optimal window function of a FFT cube.
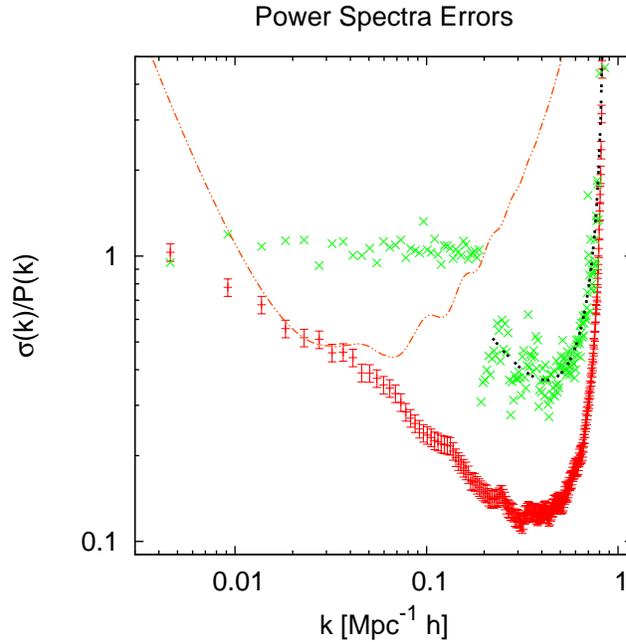
Figure 4.13: Sample errors of sample 1 (points with error bars) and calculated errors for sample A (dashed dotted line), and sample B (dotted line). See text for sample description.

Although sample A and 1_{cube} share the same technical shot noise, meaning that they have the same point density averaged over the whole FFT cube, they do not have the same "turn around" scale, where the increase in error due to higher shot-noise to spectral power ratio compensates for the decrease in error due to more modes contributing to the power spectrum at smaller scales. The "effective" shot noise of sample 1_{cube} is much closer to the value computed by only using the actual sample volume and neglecting the sourrounding empty cube used for the FFT. Both samples 1_{cube} and 1_{cuboid} start with an error of 1 at their fundamental mode, a clear indication that only one dimensional information is available. A three dimensional power spectrum using a bin-size of the fundamental k-mode would already see 8 independent modes. The different slope at large scales due to the different resolutions of sample 1_{cube} and sample 1_{cuboid}, is also clearly visible. The errors of sample 1_{cuboid} agree perfectly with the predictions of sample B, as the fundamental mode of the two smaller dimensions of the former,is equal to the fundamental mode of the latter. Shot noise is also equivalent, so the power spectrum of the cuboid is in principle the sum of a number of uncorrelated Fourier cubes with a length of the smaller dimension.

Comparing samples A on the one hand and B and 1_{cuboid} on the other, shows the effect of trading shot noise for volume when one dimension of the geometry has a fixed length, like in an IFU survey. Both samples have the same number of particles, but vary in volume. As expected, the smaller scales are much better constrained by the sample with the lower volume but higher point density (B), and larger scales are better constrained by the survey with lower point density and higher volume (A). This would be visible even more if the glass distribution would force the power sectrum to fall off exponentially on that scales and thus lead to an exponential increase in relative error. The exact transition point depends on the spectral power of the sample and the total number of particles. In the case of an IFU survey, this trade off is also restricted by the minimum shot noise of the average number of particles found in the volume of one reasonably deep exposure.

Looking at the absolute error value of all the samples, it is instantly clear that the errors of sample 1_{cuboid} have to be highly correlated to be so small, as the should never be smaller than the errors of both sample A and B. They can, however, be smaller than B at larger scales, because of the better sampling, and A at small scales, because of the smaller effective shot noise.

4.6.5   Window Function

The effect of the sample geometry on the shape of the power spectrum and the correlation of the errors is best examined by looking at the window function of the survey. Figure 4.14 shows the integrated window functions of all three tested geometries using the cube FFT. All three samples show a window function with a very shallow slope at small k values, leading to very strong correlations. Depending on the actual survey geometry, the window function falls off at the scale of the angular diameter of the sample. The fall off is clearly visible for the 20, 10, and 5 arcmin samples, at the k values of 0.2, 0.4, and 0.8 $h\,\text{Mpc}^{-1}$, corresponding to the angular diameter scales. In the case of the 20 arcmin sample, the higher order peaks of the Fourier transformed rectangle are also visible.

The integrated window function of the cuboid FFTs is plotted in Fig. 4.15. The slope falls off much faster than in the case of the FFT cube estimation. Three effects prevent the window function from falling off exponentially: The window function introduced by the line of sight selection function, the window function introduced by the cone like survey geometry, and the oversampling in one dimension with respect to the two smaller dimensions, leading to a projection of power from smaller to larger scales. The fundamental mode of the elongated dimension is much smaller than the fundamental modes of the angular dimensions, which leads to elongated volumina of the sampled modes, and correlations on the scale of the fundamental modes of the smaller dimensions.

The window function of the correlation function is simply the cuboid itself. However, it is instructive to calculate the auto correlation function of all geometries, populated with a homogeneous particle distribution. The result is simply the number of pairs available within a given separation interval. This value is antiproportional to the square of the error. The autocorrelation function of the homogeneous samples are shown in Fig. 4.16. As in the previous discussion, the samples see the three dimensions at small scales, where the
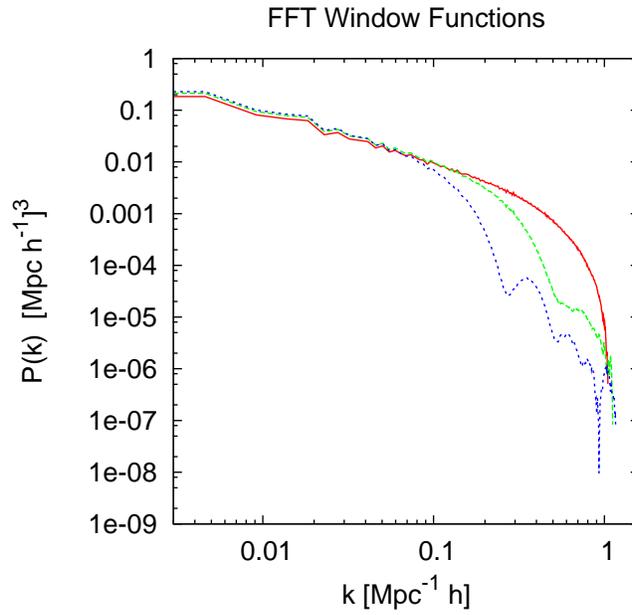
Figure 4.14: Window functions for the samples with (from top to bottom) 5 arcmin, 10 arcmin, and 20 arcmin angular dimension using the FFT cube.
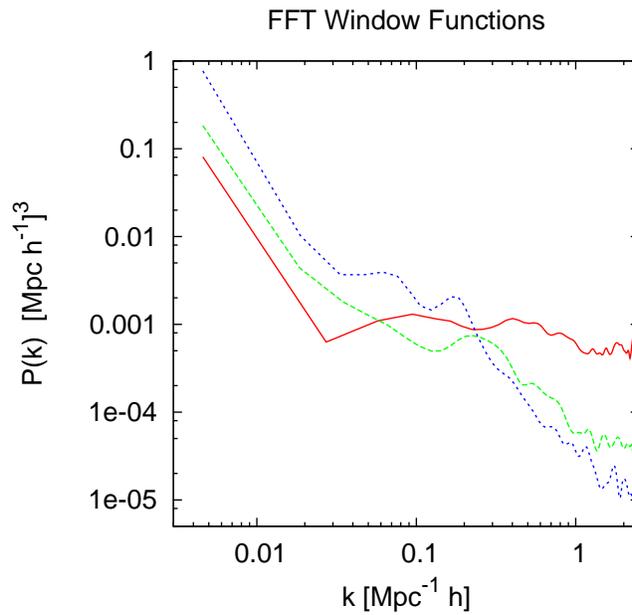
Figure 4.15: Window functions for the samples with (from top to bottom) 5 arcmin, 10 arcmin, and 20 arcmin angular dimension using the FFT cuboid.

number of correlating particles increases with the surface of a sphere around every particle, and see only one dimension at large scales, where the number of correlating particles stays constant. In this work, we use logarithmic bins and thus add a factor of about 1 to the slope. The autocorrelation function for a homogeneous cube with a side length equal to the length of all surveys is also plotted. It sees all three dimensions at every scale and thus has the same slope throughout the whole r range.
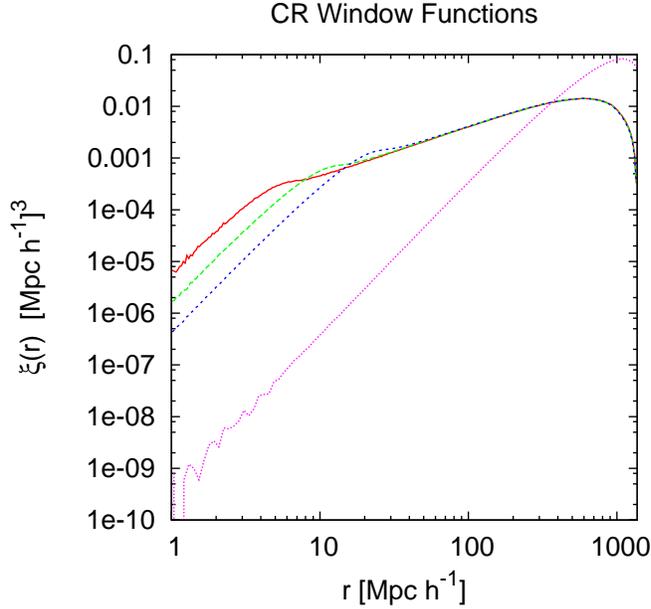
Figure 4.16: Autocorrelation functions for a homogeneous point distribution within the geometries of the samples with 5 arcmin (solid line), 10 arcmin (long dashed line), 20 arcmin (short dashed line) and a full cube (dotted line).

4.6.6   Correlation Matrix

The correlation matrix gives the correlation of a mode k with another mode k', with the case of total correlation normalized to 1 and total anticorrelation normalized to -1. We calculate the mean correlation matrix $\rho(k, k')$ by calculating the mean power $\langle P(k) \rangle$ and the error $\sigma(k)$ from the $N = 100$ realizations, compute the correlation matrix for one realization, and average over all the realizations i using

$$\rho(k, k') = \frac{1}{N} \sum_{i=1}^{N} \frac{(\langle P(k) \rangle - P_i(k)) \, (\langle P_i(k') \rangle - P(k'))}{\sigma(k) \, \sigma(k')} \, . \tag{4.21}$$

The results for sample 1-3, that vary in angular diameter but not in point density, are shown in Fig. 4.17. We only plot the correlation matrices of the power spectrum estimated with the FFT cube and the correlation function. The correlation matrices estimated using
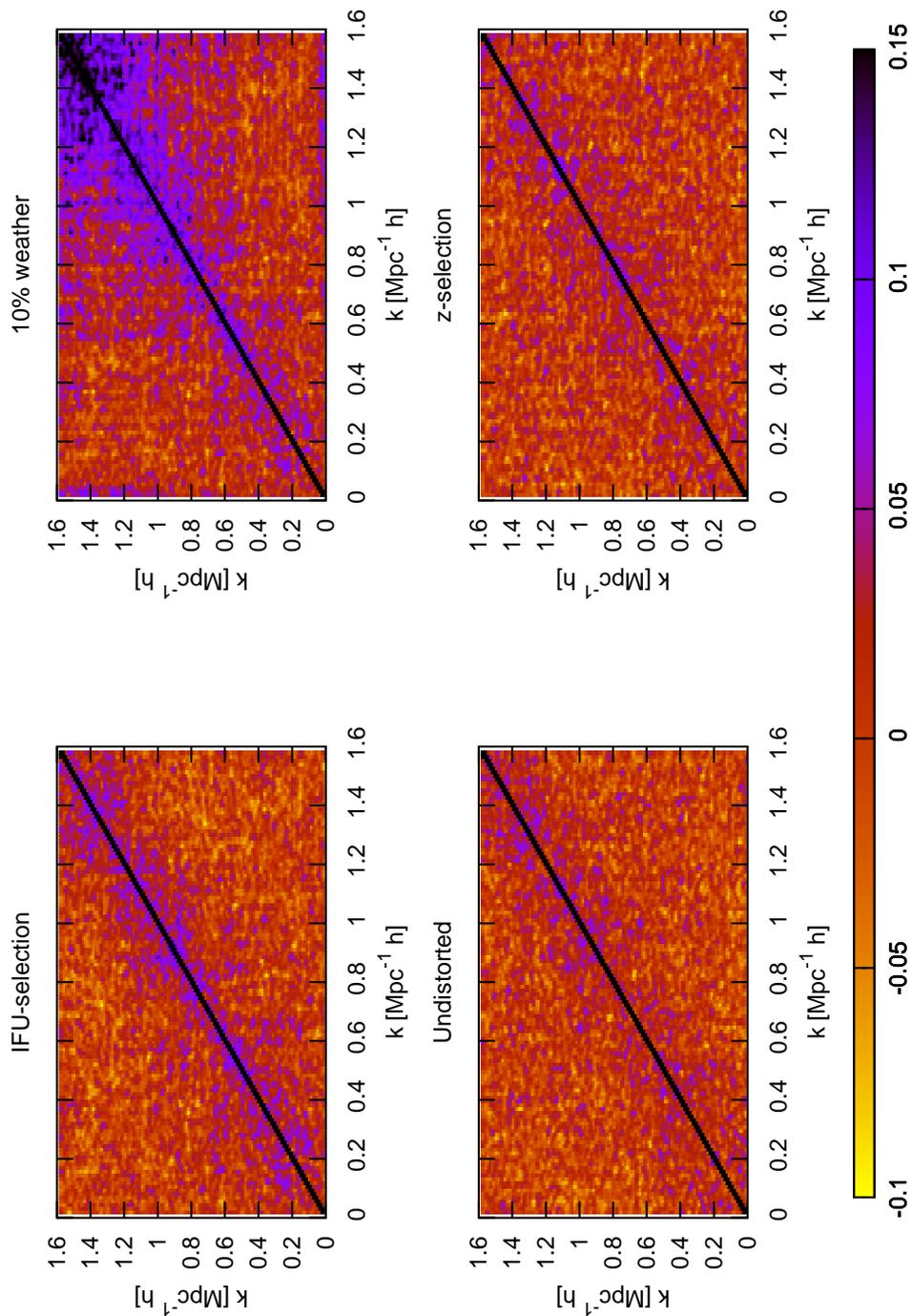
Figure 4.17: Correlation matrices for correlation functions (left column) and power spectra using FFT cubes (right column). The samples 1-3 are organized from top to bottom starting with the 20 arcmin diameter sample, followed by 10 and 5 arcmin. All samples share the same point density of $2.8 \times 10^{-2} \, h^3 \, \mathrm{Mpc}^{-3}$.
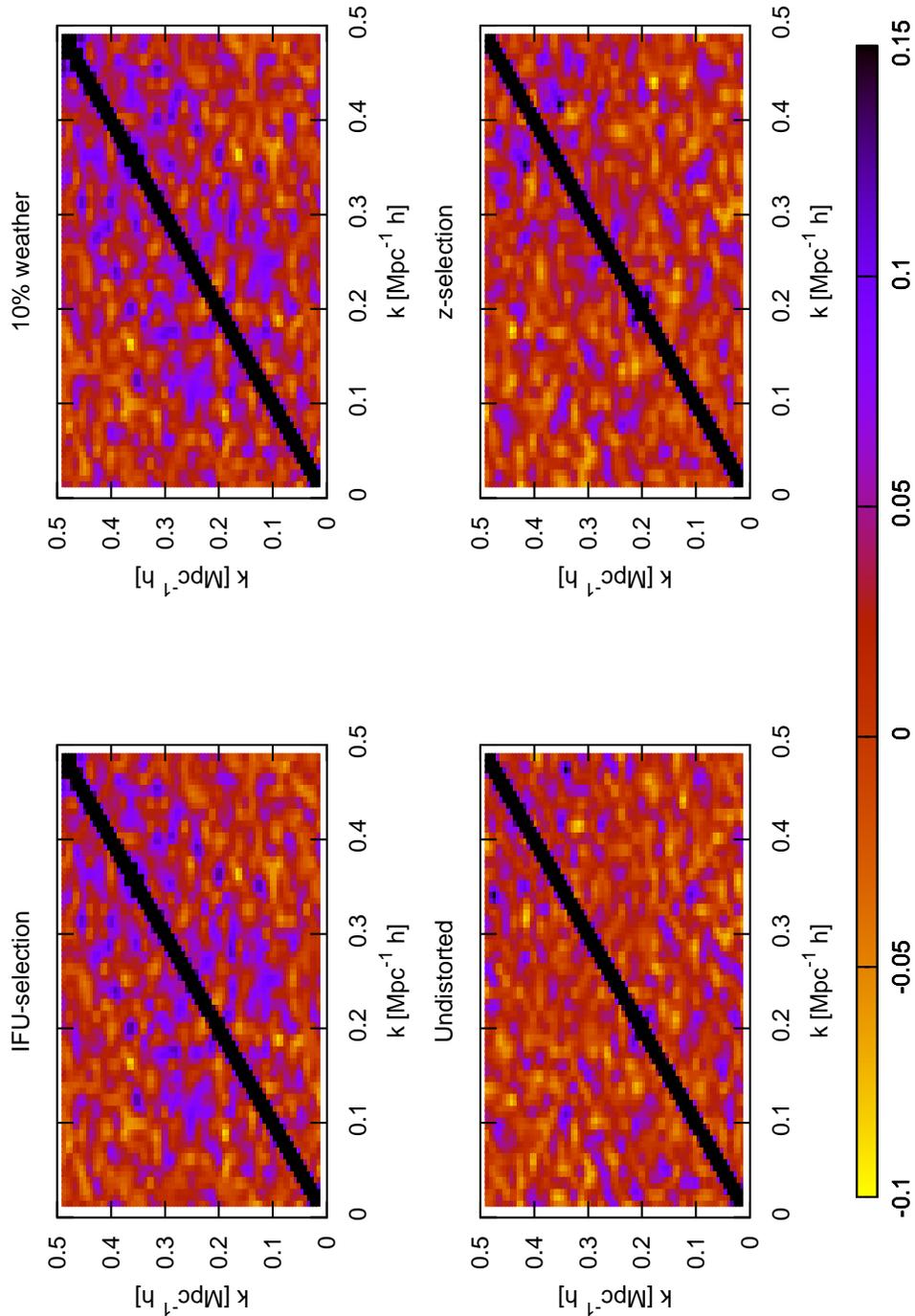
Figure 4.18: Correlation matrices for correlation functions (left column) and power spectra (right column). The samples 3-5 are organized from top to bottom starting with the 2.8×10^{-2} $h^3 \, \mathrm{Mpc}^{-3}$ sample, followed by 2.8×10^{-4} and 5.6×10^{-4} $h^3 \, \mathrm{Mpc}^{-3}$. All samples share the same angular diameter of 10 arcmin.

the FFT cuboid from the 100 realizations are, within the errorbars of $\pm\,0.15$, in agreement with the unit matrix, meaning no measureable correlations.

The correlations for the power spectrum estimated using the FFT cube are extremely strong and evident throughout all three samples. As predicted by the window function, the samples with smaller angular diameter are correlated stronger than samples with larger angular diameter. In the limit of a vanishing window function one would expect the power spectrum to not show any correlations at all, while the correlation function is expected to show considerable correlations. In the samples examined in this work, the correlation function displays correlations that are an order of magnitude smaller than the power spectrum. Furthermore, the correlations decrease with smaller angular diameter of the sample. This can be explained by that fact that the less dimensions a sample has, the fewer correlation bins a single particle contributes to, and thus the weaker the correlation between the bins.

Figure 4.18 gives the results of samples 3-5, that vary in point density but not in angular diameter. In both cases, power spectrum and correlation function, the correlation between the different bins decreases with decreasing point densities. This can be explained by the impact of increasing shot noise, which works independently on various scales. As shot noise increases the error budget becomes dominated by the uncorrelated discreetness errors and the correlations go down.

4.7   Bias Determination

We will now examine the impact of the survey geometry on the ability to recover the amplitude of both correlation function and power spectrum and thus its accuracy in constraining amplitude dependent effects like galaxy bias.

Biasing works on the amplitude of the underlying dark matter distribution and thus on the respective dark matter power spectrum, $P_{\mathrm{DM}}(k)$, and correlation function, $\xi_{\mathrm{DM}}(r)$, to generate the observable galaxy power spectra, $P_{\mathrm{obs}}(k)$, and correlation functions, $\xi_{\mathrm{obs}}(k)$. However, the galaxy power spectrum and correlation function itself is distorted by the survey window function to produce the observed quantities:

$$
\begin{aligned}
P_{\mathrm{obs}}(k) \;&=\; \left[\, b^2\, P_{\mathrm{DM}}(k) \,\right] \,\otimes\, |\,w(k)\,|^2 & (4.22)\\
&=\; b^2 \left[\, P_{\mathrm{DM}}(k) \,\otimes\, |\,w(k)\,|^2 \,\right], & (4.23)\\
\xi_{\mathrm{obs}}(r) \;&=\; b^2\, \xi_{\mathrm{DM}}(r)\, w^2(r)\,. & (4.24)
\end{aligned}
$$

In this case we are only looking at the overall scale independent biasing factor b. Both the convolution with the window function and the related multiplication in real space are associative with biasing. We can thus simply compute the accuracy in the measurement of amplitude in the distorted samples and do not have to deconvolve them.

4.7.1 χ^2-Test

We compute the ability of the power spectrum to constrain the amplitude, and thus scale independent biasing, using a simple χ^2 test:

$$\chi^2 = [P_{\mathrm{m}}(k') - P_{\mathrm{O}}(k')]^{\mathrm{T}} \, \mathrm{cov}^{-1}(k, k') \, [P_{\mathrm{m}}(k) - P_{\mathrm{O}}(k)] \, . \tag{4.25}$$

The covariance matrix $\mathrm{cov}(k, k')$ is calculated from the simulations. To fully recover the effects of any correlations we use the sample of 100 estimated power spectra as the observed power spectra $P_{\mathrm{O}}(k)$. The model power spectrum $P_{\mathrm{m}}(k)$ is simply the mean power spectrum of the 100 realizations multiplied by the tested amplitude.

We restrict the test to a k range of $0 < k < 0.8 \, h \, \mathrm{Mpc}^{-1}$ in the case of the power spectrum and an r range of $3 < r < 12 \, h^{-1} \, \mathrm{Mpc}$ in the case of the correlation function. The maximum k value is motivated by the fact, that the input power spectrum, computed from the whole simulation light cone at the appropriate redshift values, turns negative at k values bigger than 0.8, because of the initial glass distribution (see Fig. 4.4 and 4.5). The minimum r range is motivated by the fact that correlation function starts to deviate from the predicted monotonously decreasing power law behavior at that value (see Fig. 4.6). The maximum r range was picked due to the exponential increase in relative error of the correlation function at that value (see Fig. 4.12).

4.7.2   Results

We calculate χ^2-values for all 5 samples estimated with the FFT cube, FFT cuboid and correlation function. We plot the results of all tests in Fig. 4.19, grouping them depending on estimator, geometry comparison, and point density comparison.

Comparing the results of the samples with varying geometry and the results with varying point density, we find that point density is affecting the results much more strongly. This is the case, because the error is reduced only by the the cube root of the volume, but goes down linearly with increasing point density, if the shot noise is much bigger than the power in the sample. This is true for the two low-density samples, where the density is decreased by a factor of 1/10th and 1/50th, while the volume only varies by a factor of 1/4th and 1/16th in the samples with variable geometries. It is thus advisable to trade volume for point density till the shot noise falls below the amplitude of the power spectrum.

Looking at the different estimators, we see that the FFT using the cuboid and the correlation function generate roughly the same results. The results of the correlation function are somewhat worse than the FFT cuboid results at higher point densities and better at low point densities. Furthermore, the results of the correlation function show a bias towards higher amplitudes at lower point densities. This could be an indication that even the correlation function is biased towards higher amplitudes at small scales. The power spectrum does not show this effect, as the maximum k value is small enough to not include the observed bias at high k values.

The FFT using the cube however is far off. The correlations introduced by the extremely shallow window function and the additional shot noise introduced by the mostly
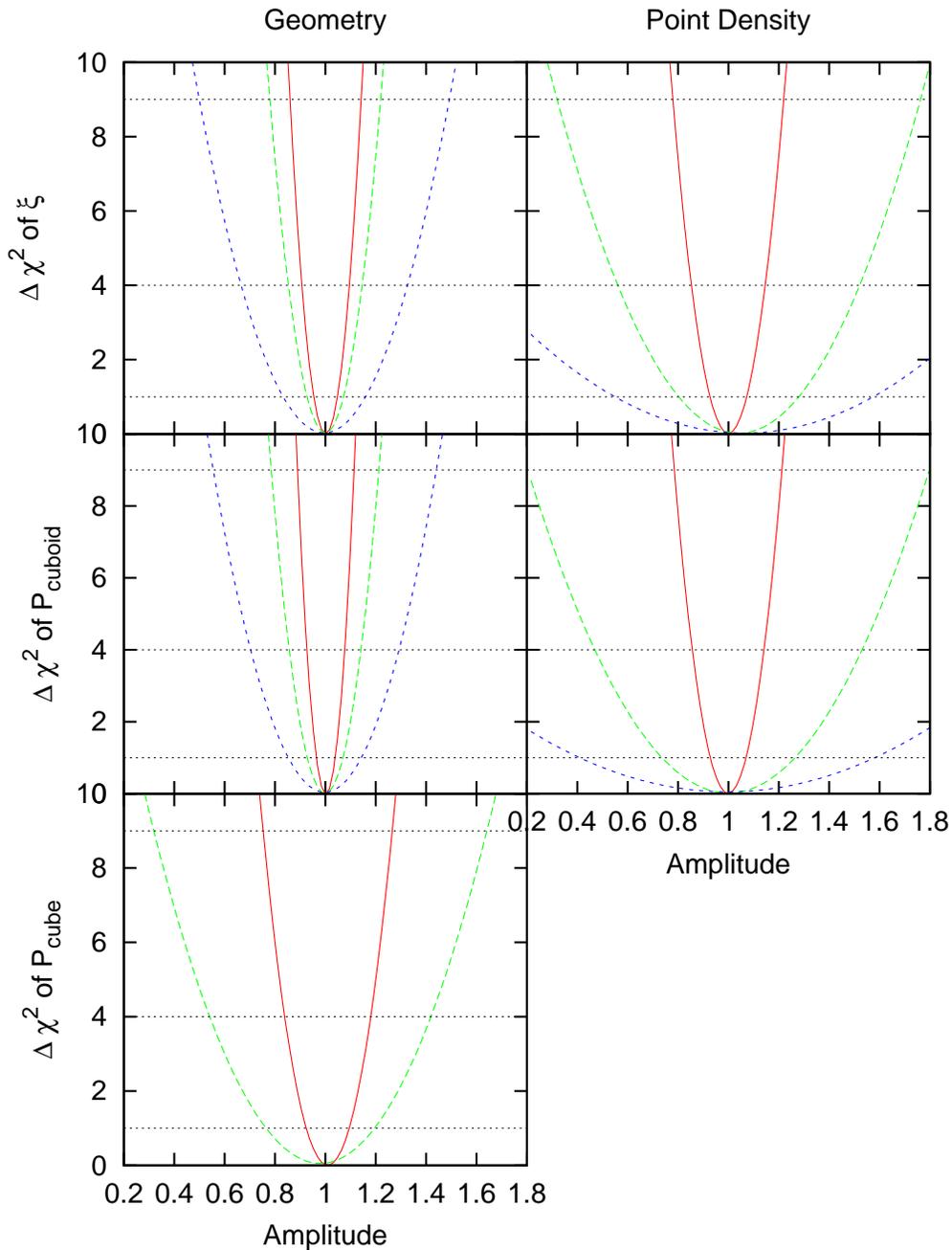
Figure 4.19: χ^2 test results for the estimators correlation function (first row), power spectrum using an FFT cuboid (second row), and power spectrum using an FFT cube (third row) comparing samples with different geometry (first column, samples 1-3) and different point density (second column, samples 3-5). Samples in the first column are 1 (solid line), 2 (long dashed line), and 3 (short dashed line). Samples in the second column are sample 3 (solid line), 4 (long dashed line) and 5 (short dashed line).

empty cube destroy much of the information. In the case of the 5 arcmin diameter sample and the low point density samples, the χ^2 test using the FFT cube estimates is even numerically extremely unstable and thus not plotted. The overall behavior of the FFT cube is numerically very unstable, especially in the case of matrix inversion, regardless of algorithm used. We are thus not able to recover the information content of the samples to a degree comparative with the FFT cuboid estimation of correlation function estimation.

Table 4.4: $1\,\sigma$ intervals for the amplitude measurement of all 5 samples using the 3 different estimators. The amplitude is normalized to 1.

Sample	$\xi(r)$	$P_{\text{cuboid}}(k)$	$P_{\text{cube}}(k)$
1	0.95 - 1.05	0.96 - 1.04	0.92 - 1.10
2	0.93 - 1.07	0.97 - 1.07	0.76 - 1-20
3	0.83 - 1.16	0.85 - 1.14	-
4	0.80 - 1.23	0.74 - 1.26	-
5	0.55 - 1.58	0.41 - 1.58	-

The exact values for the $1\,\sigma$ interval of the amplitude measurement accuracy are given in Tab. 4.4. The VIRUS-P survey is best modeled by sample 4. Although the shot noise in the VIRUS-P survey is best modeled by sample 5, the expected biasing of about 2 increases the power versus shot noise ratio by a factor of 4, which is modeled by sample 4. The VIRUS-P survey can expect the measure the amplitude with one survey volume to about 25% and the biasing to about 12% in the k-range of 0.01 to 0.8 h^{-1} Mpc. The inclusion of smaller scales will increase accuracy, but not considerably, as the shot-noise versus power ratio is already very high at these small scales, and both power spectrum and correlation function estimators show possible systematic errors at higher k-values.

4.8   Conclusions

We have calculated the effect of various survey geometries on the ability of three two-point estimators to recover the information content. As expected by the short analytical calulations made, the information content decreases as the the survey gets less and less compact in its geometry.

We use the Hubble Volume Simulation to show that the FKP power spectrum estimator shows a strong bias at very low point densities and see hints that the Landy-Szalay correlation function estimator shows similar behavior.

We use three dimensional estimators on various pencil-beam like survey geometries to retrieve power spectra and correlation functions. We find that the inclusion of the two smaller dimensions in VIRUS-P like surveys carries considerable advantages compared to using the one dimensional pendants as all estimators "see" the two smaller dimensions at some point and let them contribute to the overall information budget.

We show that the Fourier transform pairs, power spectrum and correlation function, yield, within edge effects and numerical accuracies, the same results in constraining the

amplitude and thus biasing of the investigated samples. The investigated samples show a strong dependency on point density in the ability to constrain biasing and a weaker dependency on survey volume and elongation vs. width ratio. This can easily explained by the high shot noise and its domination of the budget.

We are not able to probe both very small scales smaller than 3 h^{-1} Mpc and large scales of the order of 1500 h^{-1} Mpc as the Hubble Volume Simulations do not have enough resolution for this. Future investigations should also implement a realistic biasing model to probe for scale dependent biasing and interactions between redshift and biasing.

The predictions are, however, accurate enough to predict an estimated accuracy on the biasing measurement for the completed VIRUS-P survey of about 12% in the k range of 0.01 to 0.8 h^{-1} Mpc for one of their two 100arcmin^2 fields.

Chapter 5

The CURE Pipeline

This chapter describes the CURE pipeline developed in association with Niv Drory with help from Claus Gössl, and Helena Relke.

5.1 Introduction

In contrast to most imaging instruments that can make use of various generic reduction pipelines, integral field spectroscopy units need reduction pipelines which are tailored to fit the needs of a specific instrument (see Scodeggio et al., 2005; Sanchez, 2006; Modgiliani et al., 2007).

The CURE pipeline is designed to reduce the data coming from ≈ 200 integral field spectroscopy units (IFUs) of the VIRUS (see Hill et al., 2008) spectrograph and detect line emission objects, in this case Lyman-alpha emitters (LAEs), almost in real time and without human intervention. As most emission lines will be just above the detection limit and/or only visible in one fiber, CURE has to use robust error propagation to deliver detection probabilities together with the detections themselves.

The data CURE will be reducing is, contrary to imaging and alike spectroscopy, extremely inhomogeneous: The angular information is distributed along the y-direction of the chip. It is projected in the form of various fiber beams that guide the light from the different angular positions from the fiber-head through the grating to the camera. The line-of-sight (LOS) direction is distributed along the x-axis of the chip, in the form of different wavelengths, dispersed by the grating. The transformation between the coordinates is in principle only a geometrical one. However, the projection from fibers and wavelength to x and y coordinates on the chip is not uniform. Furthermore, the profile of the fiber itself, and thus the point spread function (PSF) has to be taken into account. After the geometrical mapping, the sky has to be subtracted from the science data. This task has to be especially thorough, as sky lines could easily be mistaken for emission lines, and the total flux of a line emission object is very small. Finally, CURE has to detect line emission objects, combining various dithers taken to reach a fill factor of one, and pinpoint their position on the sky.

5.2 Generic Reduction

The general philosophy of the CURE pipeline is to modify science frames as less as possible and avoid re-sampling to other coordinate systems as much as possible. In this way, errors are not distorted and correlations are kept to a minimum. This is important, as we do not only want to optimally detect various sources, but also give each detection a robust detection probability. Looking at the error statistics is also a good tool to check for consistency and the right treatment of the data.

Fig 5.1 show a flow diagram of the reduction and object detection process. The first step is to median the relevant calibration exposures (Arc Lights, Traces, and Biasing). In our case we separate evening and morning calibration exposures (see Sect. 5.3.3). Bias is subtracted from all calibration exposures and science exposures. Finally, an error frame is constructed for all exposures and the respective photon noise is added, as all the flux should now be from photons. We use the tools developed by Gössl & Riffeser (2002) to calculate error frames along with the reduction and track errors with Gaussian error propagation throughout every step of the reduction. Every frame is thus related to its corresponding error frame.

Afterwards the deformer and deformer2 processes construct the distortion model (see 5.3.1) and fiber model (see 5.3.2), that describes the projection of the wavelength and fiber space onto the chip. A corrected trace, science exposure specific trace is also constructed.

With the artificial models and the corrected trace, the sky is subtracted (see 5.4) from all science exposures.

Finally the detection routine searches the sky subtracted science exposures for emission lines (see 5.5).

5.3 Modelling Fibers

To model the projection of the light of the sky onto the chip, we can separate two different projections. The transformation of wavelengths and fiber coordinates onto the chip, and the profile of the fibers itself.

5.3.1 Modelling Fiber Positions

The wavelengths, w, and positions in fiber space, f, are projected onto the CCD with x and y coordinates depending on the geometry of the camera and grating. To model the distortion we use an N-dimensional polygon with cross terms:

$$F = \sum_{j,i=0}^{N} A_{ij} x^i y^j \ . \tag{5.1}$$

If the coefficients A_{ij} are known, the f and w coordinates can be calculated for every point (x, y) on the chip:
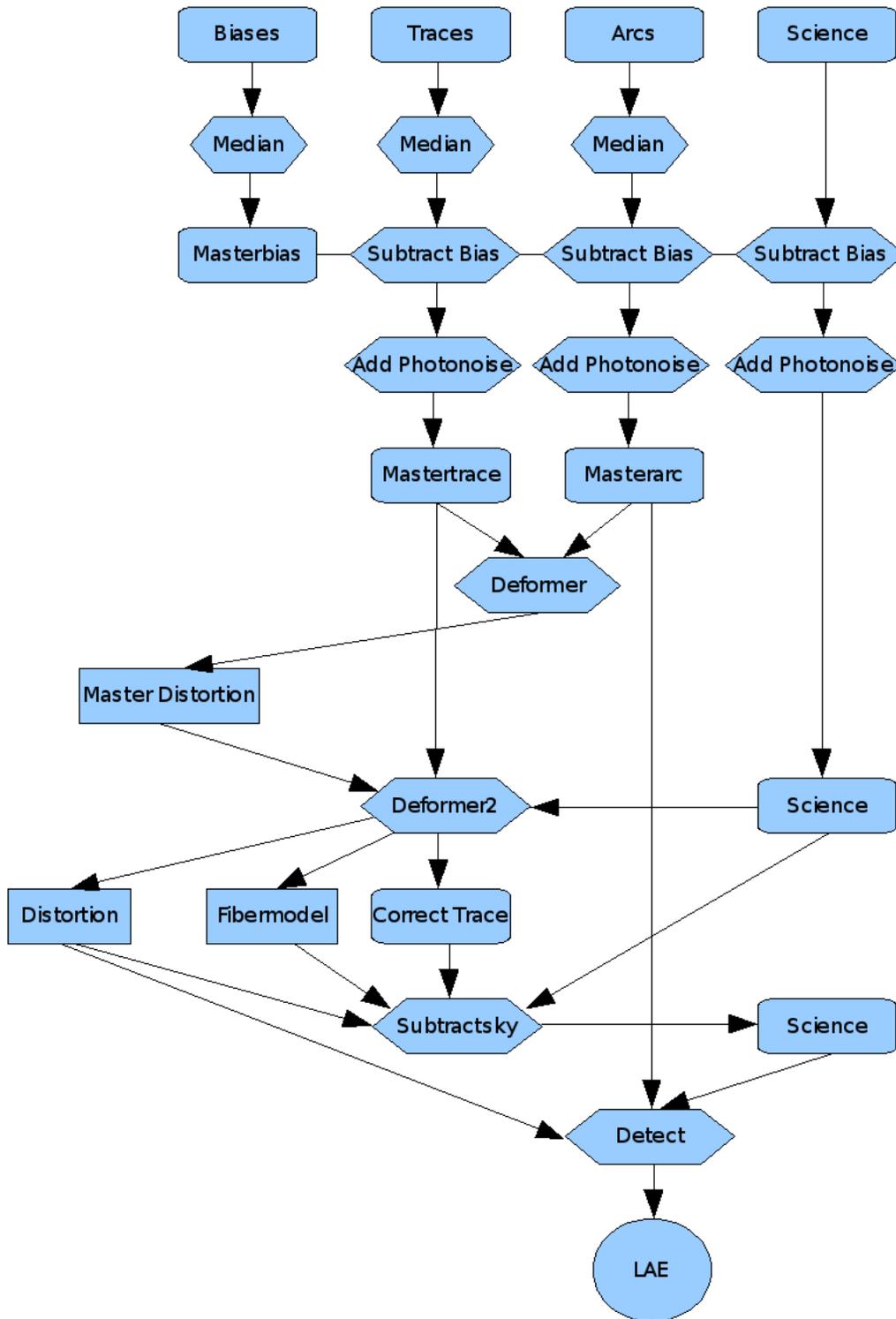
Figure 5.1: The flow diagram for the reduction and object detection process of CURE. Cuboids with round corners represent fits data files, cuboids represent text data files and hexagons represent processes.

$$f = F_f(x, y), \tag{5.2}$$
$$w = F_w(x, y). \tag{5.3}$$

We measure (w, x, y) triples by fitting seven parameter (amplitude, x-position, y-position, x-sigma, y-sigma, rotation angle, offset) Gaussians to the calibration lines of arclight exposures. Wavelengths are supposed to be constant at the same arc line. However, the w values have to be corrected, as the non uniform infall-angle for the fiber beams into the grating introduces a different w-offset in x direction per fiber. We measure this offset over a set of especially well constrained arc lines and apply it to the $w(x, y)$ solution.
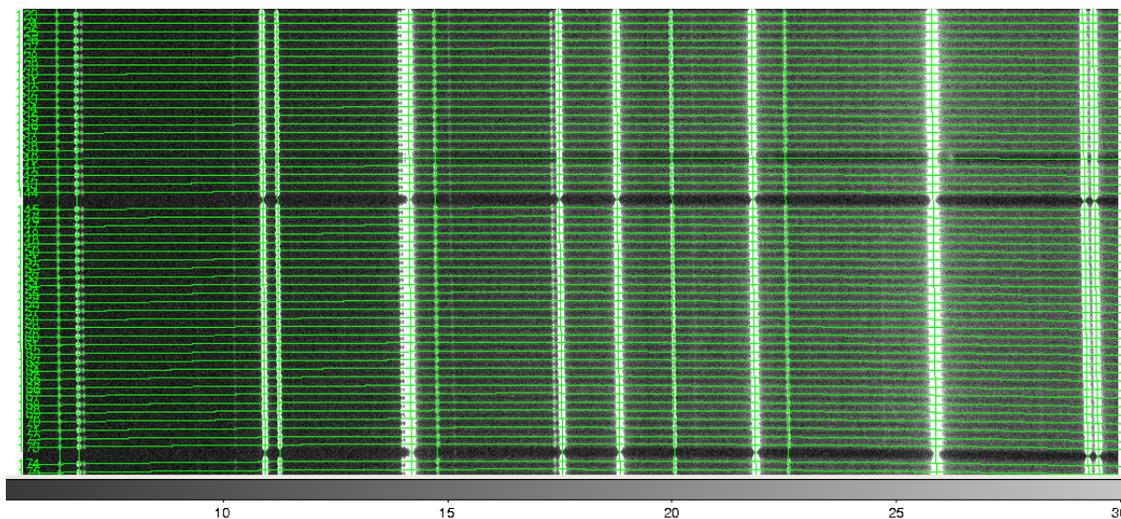
Figure 5.2: An arc frame used for calibration. The green lines show the position of fiber peaks and the position of the arc lights from the measured distortion model.

Figure 5.2 shows a cutout of an arc frame. The fitted wavelength solution at the positions of the arc lights is overlay-ed in green.

The (k, x, y) triples are also known from two dimensional Gaussian fits to arc lights, as the peak of the Gaussian is also the peak of the fiber at that wavelength, and by one dimensional Gaussian fits along the y-direction to the peak positions of the fibers. We arbitrarily construct the f-coordinates to match the y-coordinates at a certain reference wavelength. Fiber coordinates are constant along the peak flux of a fiber.

Figure 5.3 shows a cutout of a trace frame. The fitted fiber-peak solution is overlay-ed in green. Note that even very dark fibers are traced all the way from blue to red.

We also calculate coefficients for inverse and hybrid transformations, $x(w, f)$, $y(w, f)$, $y(f, x)$, and call the sum of the transformations the distortion model. We can get an estimate of the accuracy of the transformations by looking at the standard deviation of the fitted triples and by doing a forward and backward transformation for all the pixels on the chip.
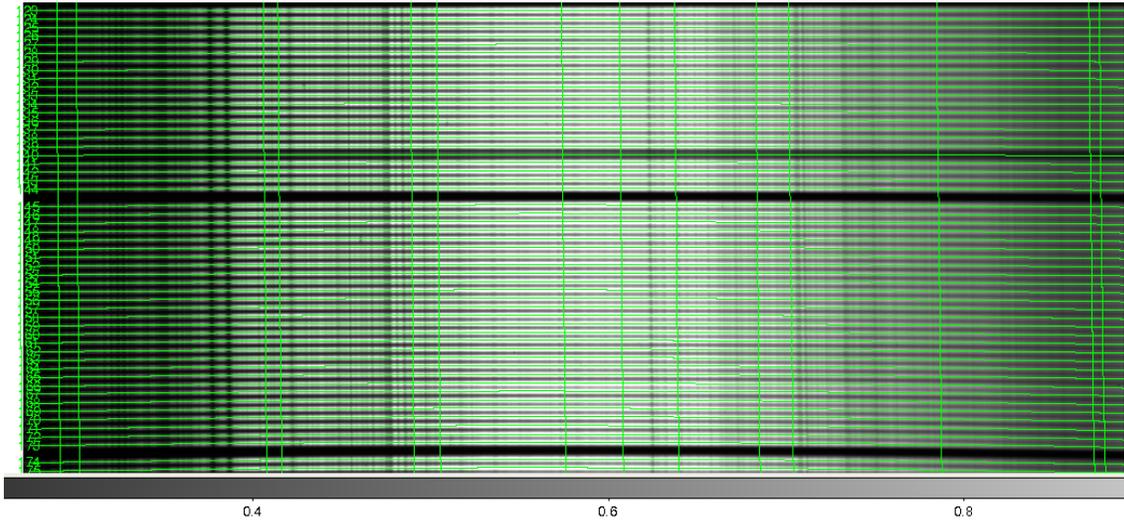
Figure 5.3: A trace frame used for calibration. The green lines show the fitted positions of fiber peaks from the measured distortion mode.

The transformations using trace and arc-light data $(f(x, y), y(f, x))$ are typically correct to about 0.02 pixel sizes, while wavelength transformations, using arc-light data only, are accurate to about 0.2 pixel sizes.

5.3.2 Modelling Fiber Profiles

To model the profile of the fibers on the chip, we fit first-order Gauss-Hermites $(GH(y))$ to cuts along the y-direction of the chip:

$$GH(y) = A \exp\left[-0.5\,\frac{y^2 - \mu^2}{\sigma^2}\right] \left[1 + h_2\left(\frac{y^2 - \mu^2}{\sigma^2} - 1\right)\right], \qquad (5.4)$$

$$Flux(y) = GH_{F+1}(y) + GH_F(y) + GH_{F-1}(y). \qquad (5.5)$$

The starting position of the peak in y-direction μ is given by the distortion solution acquired earlier, leaving us with the standard deviation σ and the first-order Hermite coefficient h_2.

Because of the overlapping fiber fluxes, we fit three adjacent fibers $(F, F+1, F-1)$ simultaneously, using the result as starting point for next fiber triple and keeping the fits for the fibers next and right to the probed fiber fixed. This procedure is iterated until the GH parameters converge. At the red parts of the chip, the possible parameter space has to be constrained as the fit tries to exploit the degeneracy of the data by alternating the sign of h_2.

The parameters of the Gauss-Hermite (especially y and σ) vary over the chip. We again fit a polynomial with cross terms to the measured Amplitude A, width σ, and first Hermite coefficient h_2:

$$A \;=\; F_A(x,y)\,, \tag{5.6}$$

$$\sigma \;=\; F_\sigma(x,y)\,, \tag{5.7}$$

$$h_2 \;=\; F_{h_2}(x,y)\,. \tag{5.8}$$

Unfortunately, the fiber profiles are overlapping quite a bit in some regions of the chip, leading to degeneracies between the different parameters of he Gauss-Hermite model. Figure 5.4 shows the fiber profile of three adjacent fibers in the blue, green and red spectrum on the chip. One can instantly see the broadening of the fiber profile and the resulting overlap. It is also obvious that it is much harder to estimate the profile of a single fiber in the case of heavily overlapping fiber profiles, where the cumulative fiber profile (solid lines in Fig 5.4) and the single fiber profile (dashed lines in Fig 5.4) deviate so much.

5.3.3   Models and Reality

Reaching a model accuracy of about 1% on the chip is not an easy task. While the fiber peaks have relatively high fluxes and are thus easy to constrain, the fiber edges, which cover more area, are much fainter (see Fig. 5.4). In the blue, where the fibers have only negligible overlap, the flux difference is large. In the red, the flux difference is no so big, but the strong overlap reduces the accuracy of the fit. In both cases, the model has to be much more accurate than 1% at the peak of a fiber to reach this accuracy at the edges of the fiber.

We also find, that both the distortion patterns and fiber models vary within a night. Figure 5.5 shows the deviations of both the fiber positions in y-direction and the σ-values of the model from calibration exposures at the start and the end of a night. This behavior is most likely explained by changes in temperature and the resulting breathing of the instrument. The impact of this effect is strong, especially because the fiber profiles are quite steep and a small change in position or width results in a large change in flux at a given position. This forces us to correct distortions and models acquired from calibration frames taken at the start and the end of the night for science exposures taken during the night.

To correct the distortion model, we keep the $w(x,f)$ transformation and correct the $f(x,y)$ transformation by measuring the peak positions (and thus f) from the sky flux in the science exposures. The fiber model is also corrected, by measuring σ and h_2 directly from the science exposures, using the Gauss fitting technique mentioned earlier (see Sect. 5.3.1 and 5.3.2).

However, due to degeneracies in the fit, and because the Gauss-Hermite is probably not the best fitting model, a trace constructed from the fiber model of a science exposure is not as accurate as measured trace from earlier or later that night. Thus we also calculate a trace frame that is corrected for the change in the distortion model by re-sampling it to the correct fiber peak positions, but is not corrected for changes in the fiber model (especially the width of the profile). The re-sampling is not that much of a problem here,
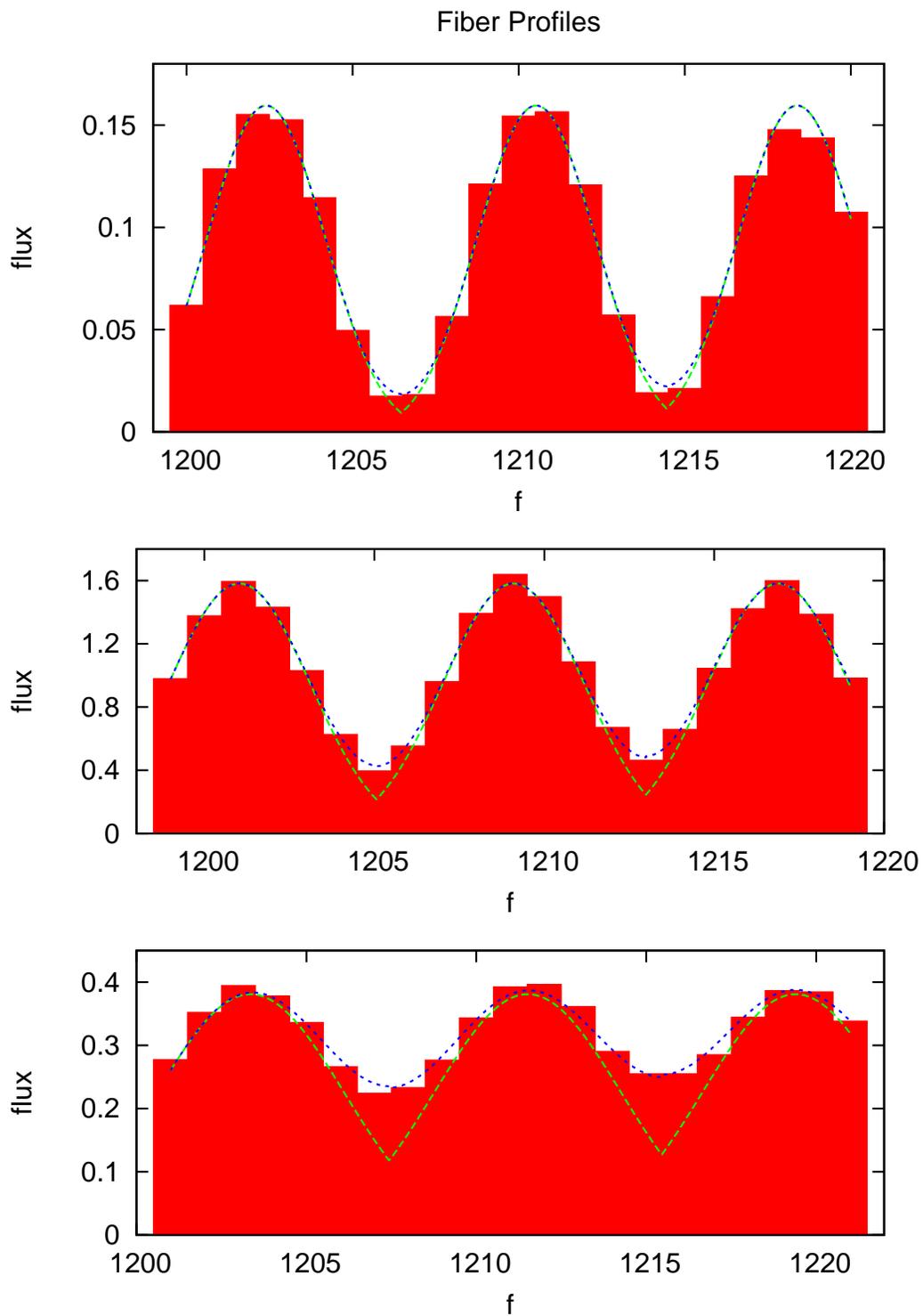
Figure 5.4: Fiber profiles from the blue (top panel), green (middle panel), and red (bottom panel) of the chip. The filled red areas give the data points in the trace frame. The long-dashed line is the fitted artificial fiber model summed over all fibers. The short-dashed line gives the artificial fiber model for the nearest fiber only.
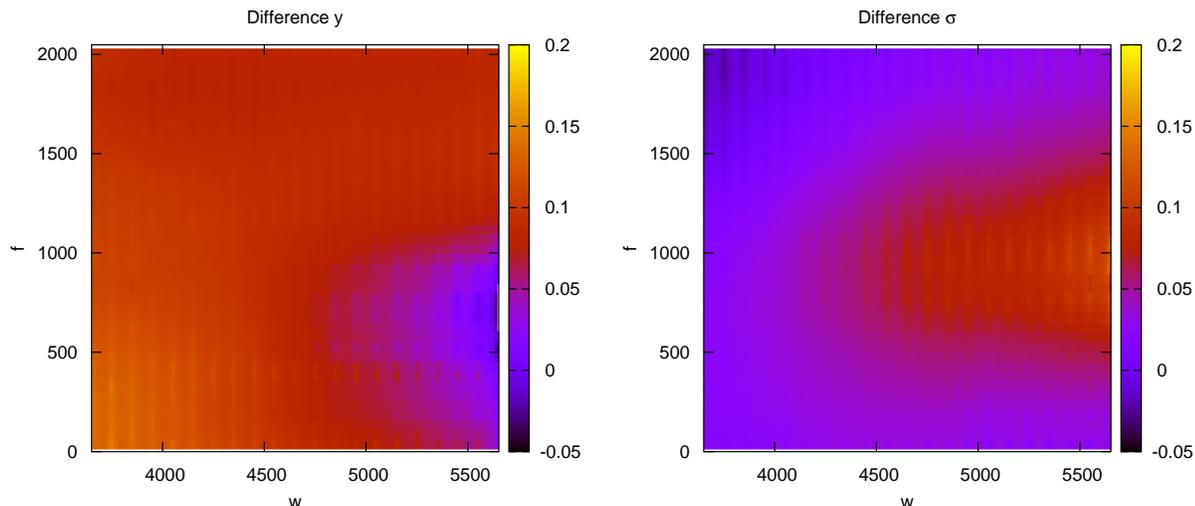
Figure 5.5: Differences in the measured y-positions of fiber peaks (left) and width of the fibers (right) of two trace frames at the start and end of a night.

as the trace is only used as a weight for sky subtraction. We call this fits file the corrected trace. The corrected trace can only be used in cases, where the cumulative flux of all fibers is looked at, as it can not reproduce fiber profiles of single fibers. This is needed in the sky subtraction to eliminate continuum sources. For these purposes, CURE uses the fitted fiber model.

5.4   Subtracting Sky

HETDEX is a blind survey looking for emission line galaxies. Because of this, most of the observed area, both in angular direction as well as in wavelength direction, will by empty and thus contain only sky emission. By exploiting this fact, CURE is able to take the sky for sky subtraction directly out of the science exposure.

In order to subtract the sky from the science frame, we first divide the whole science frame by the corrected trace. In this way, all fluxes of one wavelength will be brought to the same level. We then construct the sky by looking at the flux per wavelength (in this case the pixels size in wavelength space) of each pixel on the CCD and assign it the respective wavelength. Because of the distortions introduced by the camera, the sky is sampled with a much higher resolution than the size of the pixels in Angstrom, as every pixel gets light from a slightly different wavelength.

Only pixels of a certain window around a reference fiber for which the sky is constructed are used, because the sky changes slightly over the chip due to instrument distortions. The resulting sky-flux - wavelength pairs are interpolated using a b-spline, clipping all values with an offset of a certain value over the estimated spline in a second step to get rid of cosmics and possible continuum sources.

Figure 5.6 shows the performance of the spline fit. The upper panel gives the sky of a
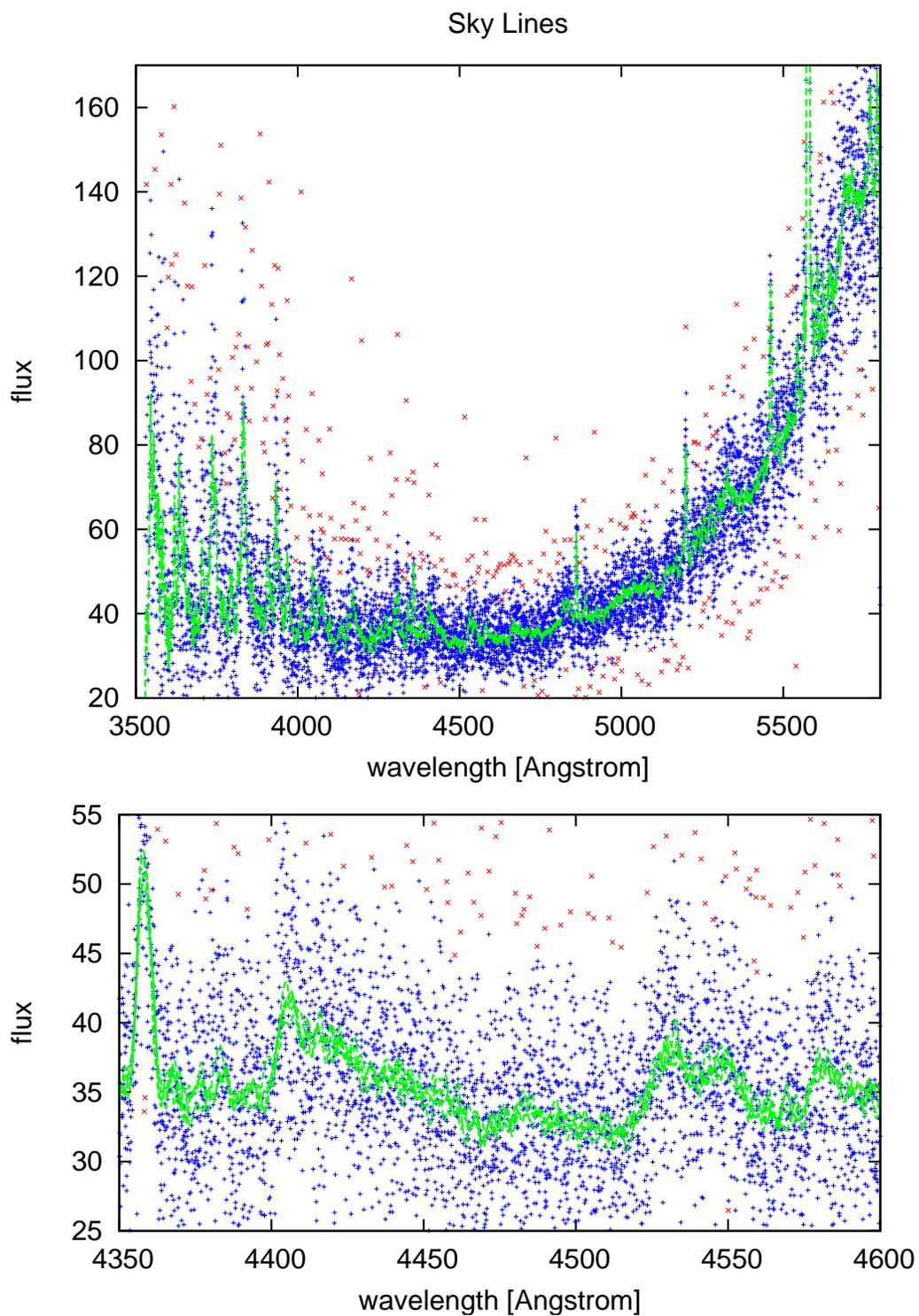
Figure 5.6: Examples of sky construction with CURE. The top panel shows the full sky with the reconstructed sky (solid line), while the bottom panel shows a cutout of the sky with reconstructed sky spectra from 4 areas of the chip. Both panels show contributing pixels (+) and rejected pixels (x).

100 pixel window around a fiber in the middle of the chip, while the lower panel shows a
small section with sky fits from four different fibers equally distributed on the chip.

The sky construction is robust until about 50% of the available chip space is contaminated with continuum sources, as CURE automatically detects continuum sources and
ignores them during sky construction.

After the construction of the sky, the wavelength for every pixel of the science exposure
is calculated and the sky flux at the appropriate wavelength, multiplied by the corrected
trace, is subtracted.
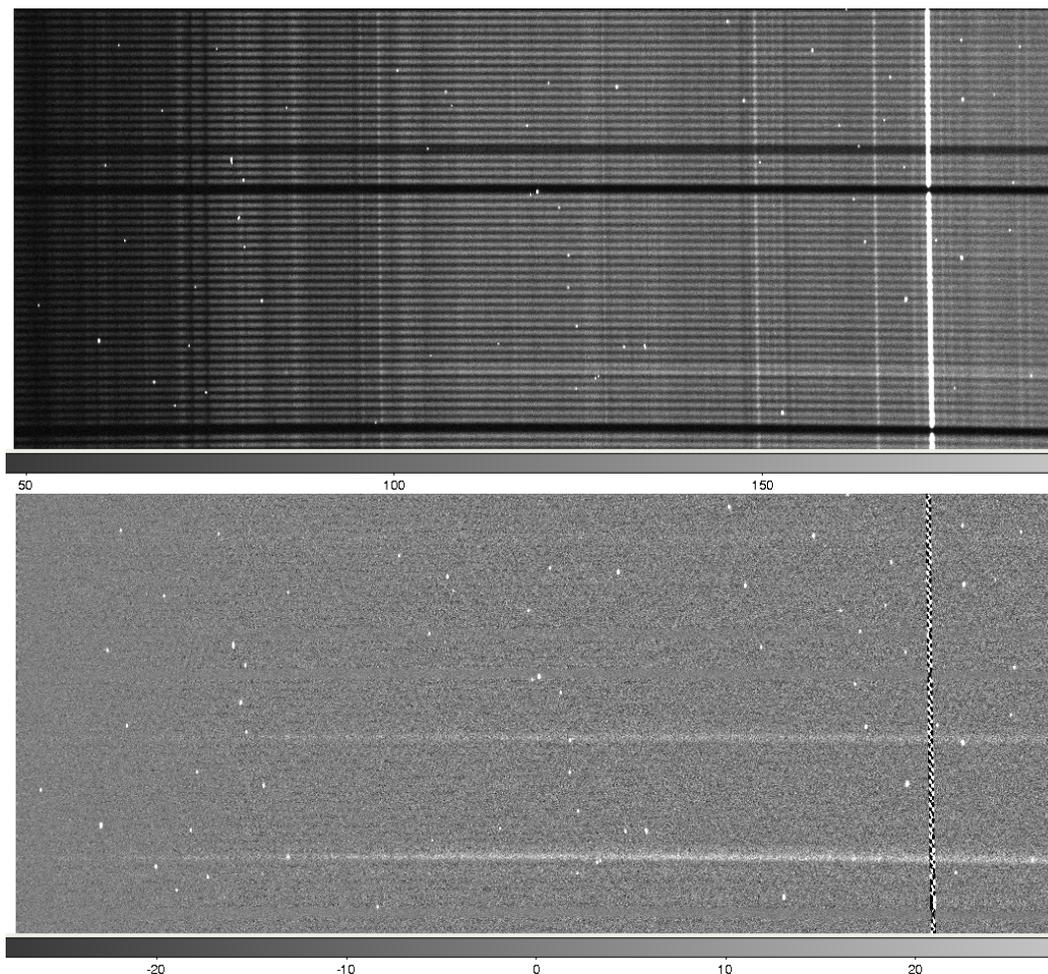
Figure 5.7: Science frame before and after sky subtraction. The residuals of the 5577 Å sky-line
and the higher fluctuations due to higher photon noise on the fiber peaks are clearly visible.

Figure 5.7 show the difference between a section of a science exposure before and after
the sky subtraction. Although one can still see the fibers in the sky subtracted sky, this
is due to the higher absolute photon noise in the fiber peaks compared to the fiber edges.
Sky lines, except for the 5577, are eliminated to the point where they vanish in the noise.
Two fibers do now clearly show continuum, which is filtered out by the sky subtraction

routine when constructing the sky. If the continuum sources were not filtered out for sky construction, the estimated sky around these fibers had a higher flux leading to an oversubtraction in this area. Note that the science exposure is not corrected for cosmic rays, as the sky subtraction routine also automatically neglects them when constructing the sky.

After sky subtraction, the science frame should have a uniform distribution around zero with deviations given by the error frame. In the case of inaccurate sky subtraction fiber profiles will not completely vanish which has a severe impact on statistical detection of line emission sources. To check for accurate sky subtraction we look at various statistical properties of resolution elements on the chip. Each resolution element is centered on a fiber peak. The dimensions in fiber and wavelength space are given by the size of the point spread function at that position, with a total pixel number n.

We measure the histogram of the average sum of fluxes, f, in the resolution element:

$$sum = \frac{1}{n} \sum_{i=1}^{n} f(i) \,, \qquad (5.9)$$

the significance, σ, of all the pixels within a resolution element,

$$\sigma = \sum_{i=1}^{n} \frac{f(i)}{e^2(i)} \,, \qquad (5.10)$$

the root-mean-square, rms, of the pixels within a resolution element,

$$rms = \sqrt{\frac{1}{n} \sum_{i=1}^{n} f^2(i)} \,, \qquad (5.11)$$

and the reduced χ^2-value of the resolution element,

$$\chi^2 = \frac{1}{n} \sum_{i=1}^{n} \frac{f^2(i)}{e^2(i)} \,. \qquad (5.12)$$

The histograms for the statistics of a typical sky-subtracted frame are shown in Fig. 5.8. The sum of all the pixels in a resolution element of a mostly empty frame should be 0, with a Gaussian distribution and a standard deviation of the error of one pixel divided by the square root of the number of pixels. The error of one pixels is the rms of the pixels in a resolution element. One can read of two peaks in the rms histogram that represent the errors of mostly empty pixels at the edge of a fiber, where read noise dominates, and pixels in a fiber, where read noise and photon noise are added up. The error of fiber pixels is fitted with a value of 11. Having 64 pixels per resolution element, the standard deviation of the sum should be 88, which agrees well with the fitted 85. The histogram of the significance σ should resemble a Gaussian distribution with mean of 0 and a standard deviation of exactly 1. The fit shows a good agreement with this prediction. Finally, the χ^2 histogram should follow a χ^2 distribution with $k = 64$ degrees of freedom. We fit such a distribution with an additional offset term, e, to the histogram. The offset we measure,
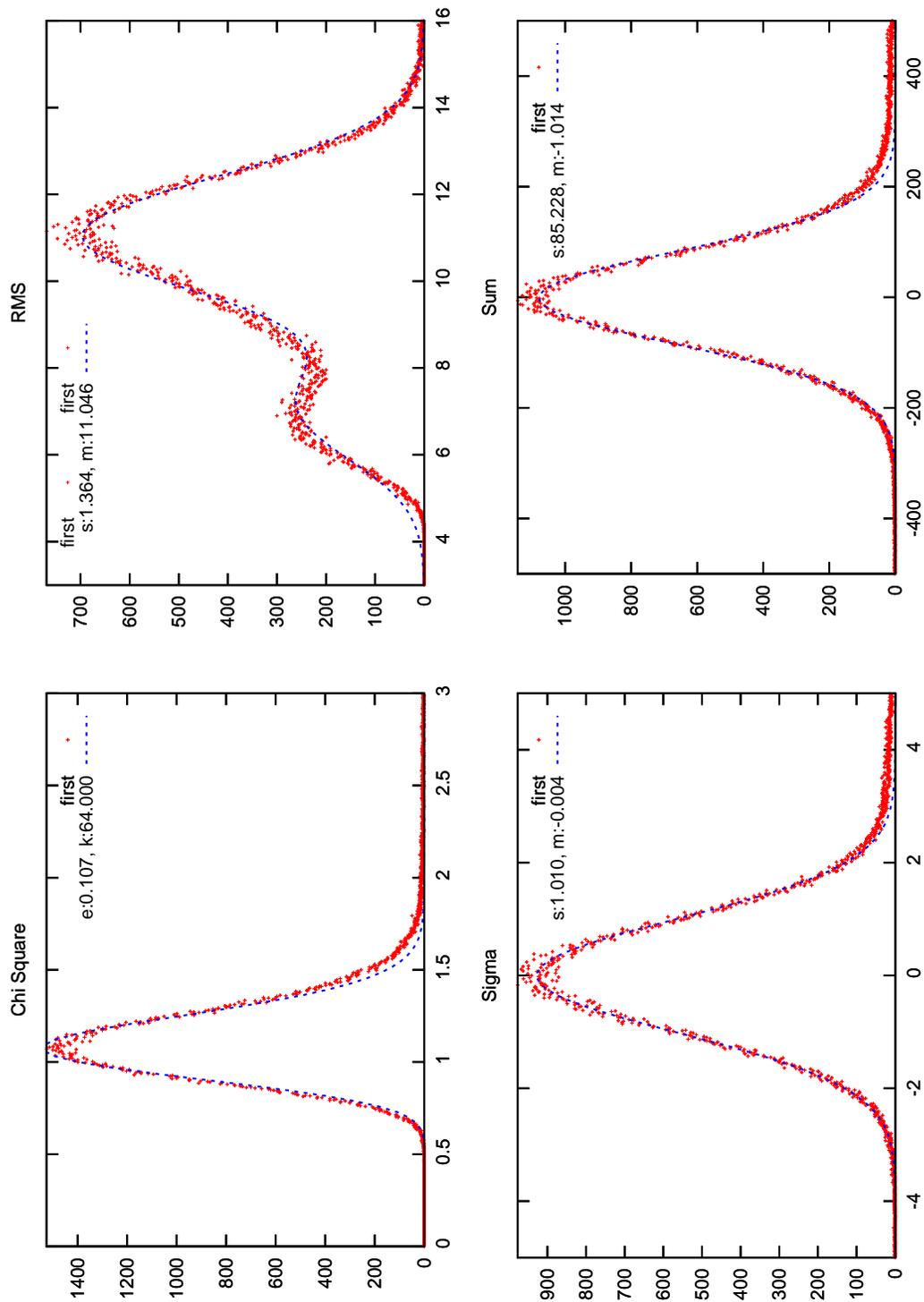
Figure 5.8: Histograms of the *sum* (lower left), σ (lower right), *rms* (upper right), and χ^2 (upper left) values of the resolution elements in a sky subtracted frame. Dashed lines denote Gaussian fits (*sum*, *rms*, σ) and χ^2-distribution fits (χ^2) to the histogram data.

$e \approx 0.1$, is very small, and the χ^2 distribution fits the histogram very well. We notice an excess of resolution elements with higher χ^2 values, which are due to non empty resolution elements (cosmics, continuum sources, line emission).

All in all the statistics look consistent and match the predicted values. Although the offset in χ^2 shows, that there are still some minor systematic errors in the sky subtraction, that introduce asymmetries in the error weighted fluxes.

5.5   Detecting Emission Lines

After sky subtraction, CURE is ready to detect LAEs on the science exposures. All science exposures have to be binned into bundles containing all the dithers of one observation field. CURE does not need to have multiple science exposures at the same dither positions for cosmic ray elimination, as they are rejected at a later time.

Arc-light frames are also needed for every science exposure. They are used to calculate the point spread function (PSF) by fitting a seven-dimensional Gaussian to the arc lights. The resulting parameters are then fitted by a polynomial to describe the evolution of the PSF over the chip. The quality of this fit is visualized in Fig. 5.9. The strong variations of the PSF's standard deviation are captured well by the polynomial fit.

The detection routine starts to sample the sky covered by the observation (α, δ) and the wavelength range observed (w) on a three-dimensional grid. It calculates the flux in each fiber of the available science exposures using the geometry (position and size of the fibers) of the IFU and the seeing of the related science exposure. The fluxes are then projected on the chip using the fiber and wavelength information together with the distortion solutions.

The possible LAE positions (it can be more than one, as the same LAE can appear in different fibers in one dither and in different dithers) is then checked for signal on the chip. Only if the cumulative singal to noise in the expected positions is above a defined threshold, the position together with the expected flux and signal to noise ratio is stored.

After all positions on the grid were checked for singals, the resulting positions in $\alpha-, \delta-, w$-space with enough signal to noise are segmented, looking for roughly spherical regions of high signal to noise. Elongated segments represent continuum sources and are not probed for LAEs.

Suitable segments are probed for possible LAEs. The stored fluxes on the possible LAE are distributed according to the PSF on the stored spots on the chip and compared to the observed data using a simple χ^2-test. All regions in the various science frames with a minimum expected flux are taken into account.

The detection algorithm produces a list of χ^2-values, and signal to noise ratios for every probed position in the segment.
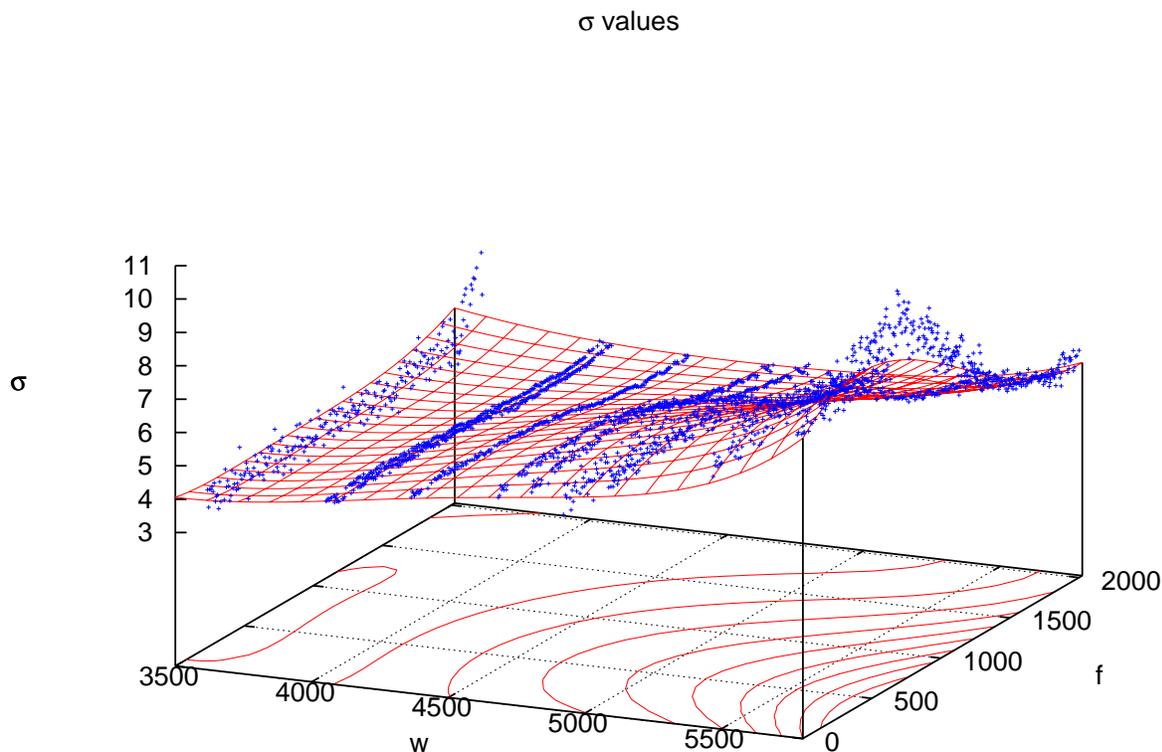
Figure 5.9: Standard deviations fitted to the point spread function over the chip. Crosses show the measured standard deviations, while the grid visualizes the fitted polynomial plane.

5.6   Testing the Pipeline

To test the pipeline we use a set of observations from VIRUS-P, the prototype instrument for HETDEX. We simulate a number of LAEs into the set of observations by using the detection template for the χ^2-test and put these directly into the unreduced science frames adding photon noise on top of it. We then put the science frame through the whole reduction and detection process and see how well we can reproduce the input parameters of various LAE sources.

Figure 5.10 shows the detection result of an artificial Lyman alpha emitter. Only the result with the minimum χ^2 value is plotted. To check for the detection performance, we simulate a couple of hundred sources with a signal to noise below and above to the expected detection limit and some very bright sources. We find that $\sim 97\%$ of the generated sources are found. The 3% missing sources are lost due to the following reasons:

- part of the source source is not fully projected onto the chip (edge of chip of IFU)
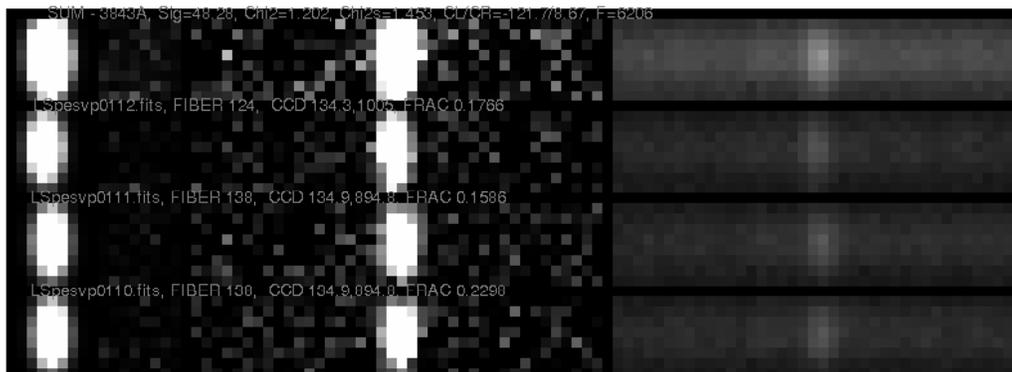
Figure 5.10: A Lyman-alpha emitter detected in three fibers distributed over three dither steps. The data shows, from left to right, the flux of the predicted LAE model, the values per pixel that contribute to the χ^2 test, the sky-subtracted data, and the corresponding errors.

- part of the source is projected onto a comic ray

- part of the source is projected onto the 5577 sky line

None of the sources were missed without an obvious reason. Together with the detection of generated sources a lot of fake sources were detected. Figure 5.11 shows the reduced χ^2 versus σ distribution of the generated sources (x) and fake detections (+).

Within the given σ-range, the generated sources do not show a strong dependency on chi^2. This is expected, as the generated sources should, by construction, all have a reduced χ^2 value of ~ 1. Any deviations apart from the usual random fluctuations should be coming from the underlying flux distribution due to systematic errors in the reduction process. By looking at the χ^2-statistics of the reduced sky spectra we already know that only a small offset to higher chi^2-values exists.

Fake sources are much more numerous at lower σ values, as the probability for random fluctuations increases with decreasing absolute total flux in the fluctuation. Cosmic rays are almost never mistaken for LAEs, as their chi^2-values are very high and are thus not shown in the plot.

The reduced χ^2-value of fake sources seems to be proportional with the σ-value. This is also expected, because higher randomly distributed flux increases the χ^2 value when compared to a specific model. A composite threshold of $\sigma > 4$ and $\chi^2 < 1.3$ seems to exclude virtually all fake sources, and only generated sources with very low flux. Further investigation is needed to infer better exclusion functions.

When CURE is searching for a line emission, it looks at one segment of connected regions (in angular-redshift-space) of high singal to noise and tries to get the best fitting position of a LAE. This is represented by the smallest χ^2-value in the segment. We use the related angular and redshift position as best fit for the actual location of the detected LAE. When comparing the actual input position with the recovered position after reduction and detection we can asses the accuracy of the CURE pipeline.
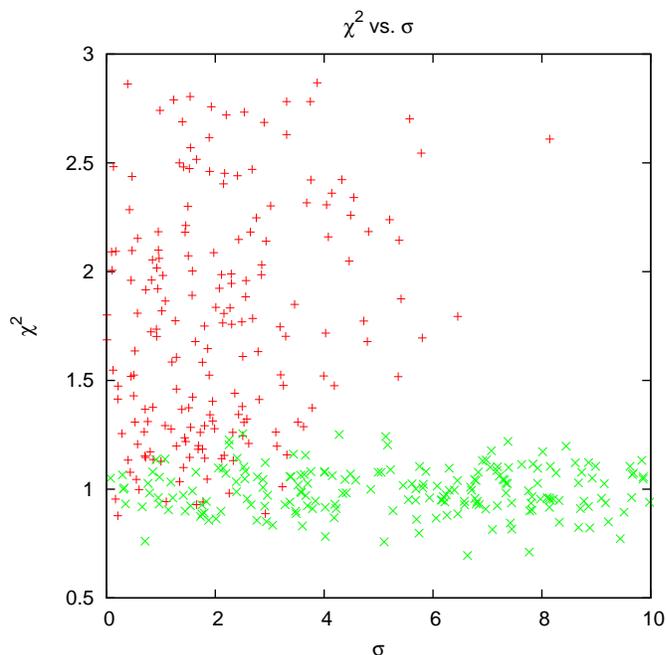
Figure 5.11: Comparison of χ^2 and σ values between generated sources (x) and fake detections (+).

Figure 5.12 shows the accuracy we reach modelling VIRUS with a fiber size of 1.2 arcsec and an average fiber separation of 2.1 arcsec using three dither positions to achieve a fill factor of 1. The recovered plane resembles a Gaussian with a standard deviation of about 1 arcsec in angular space and $10^{-4} z$ in redshift-space. Both values are centered around 0 and show no signs of systematic errors. This accuracy was reached by using a very small initial search grid, both in angular and in redshift coordinates. Making the initial grid much more coarse results in less accuracy and in systematic offsets.

5.7   Conclusions

We develop a pipeline that is able to reduce large quantities of IFU data automatically and robustly. This is achieved mostly by consistent use of Gaussian error propagation throughout the whole reduction process and by avoiding the re-sampling of actual science data as much as possible. CURE is able to accurately model the projection of fiber ,f, and wavelength, w, coordinates to x and y coordinates on the chip and correct for wavelength offsets due to variation in the infall angle of fiber beams. Fiber profiles are modeled to a satisfactory degree using first order Gauss-Hermite terms. This could be improved further, as a good fiber model is imperative to fit parameters in an environment where fibers
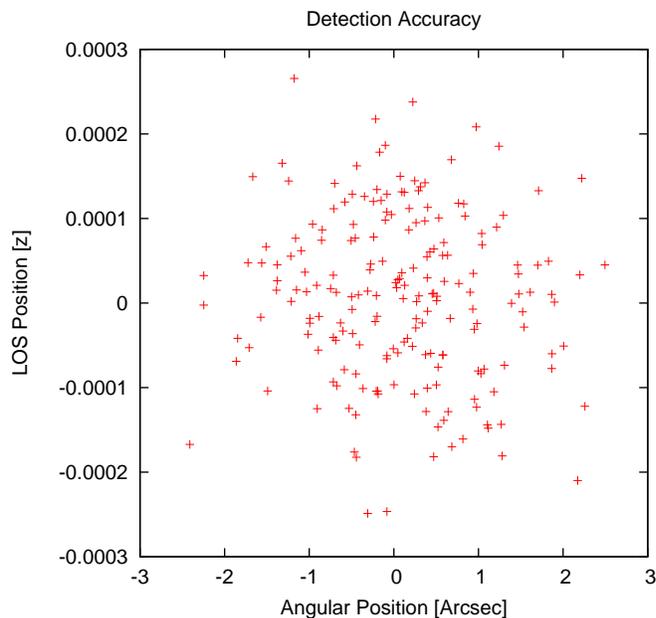
Figure 5.12: Accuracy (position of generated source minus position of detected source) of line emission detections in angular coordinates and redshift coordinates.

overlap heavily and a fit carries the risk of exploiting degeneracies. We show that CURE is subtracting sky, using sky information from science frames only, robustly and unbiased on the percent niveau. This is the case even in the presence of continuum sources of up to 50% and lots of cosmic rays in the science frame. CURE is able to detect artificially generated LAEs reliably and with a good detection accuracy of the order of the fiber diameter in angular coordinates and to the resolution limit of the spectrograph in wavelength coordinates. The Gaussian error propagation enables CURE to calculate robust statistics which help to discriminate between true and fake detections.

Chapter 6

Conclusions and Outlook

6.1 Conclusions

During the course of this work we have shown that the various physical and technical problems that could arise in the course of using a Lyman-alpha emitter power spectrum to constrain cosmological parameters can be corrected for or neglected.

We have shown that Baryonic Acoustic Oscillations can be extracted from the overall power spectrum without any further assumptions about the specific composition of the universe using a parametric fit to the non-oscillating dark matter part of the power spectrum. We provide a robust method to apply a theoretical template to the transfer function calculated from the power spectrum. The Oscillations themselves have, although representing only a small part of the overall information content contained in the power spectrum, prooven to be a very robust tool to measure cosmological parameters. The phases of the BAOs are, within the parameters of a survey like HETDEX, not affected by effects like non-linear growth, biasing or redshift-space distortions. However, recent studies (see Sanchez et al., 2008) have shown that in the sub percent regime, which is not accessible to simulations like the Hubble Volume used in this work, non-linear growth and redshift-space effects do indeed change the phase of the oscillations. This has to be taken into account for future, even more accurate surveys.

Because of their robustness and low systematics BAOs could even be used as a tool to check for the consistency of a survey, as the amplitude of the oscillations should behave exactly like the amplitude of the power spectrum itself. In principle cosmological tests probing for growth could be enhanced by using both the amplitude of the BAOs and the amplitude of the whole power spectrum.

Using the whole shape of the power spectrum greatly increases the power of a survey to constrain cosmological parameters, but also introduces more systematic errors and much stronger demands on the survey design, instruments and statistical treatment of the data. The power spectrum has to be modeled to the sub percent niveau including non-linear

growth, redshift-space effects, and biasing. Jeong & Komatsu (2006) and Jeong & Komatsu (2008) have made good progress on that. The requirements to the statistical treatment of realistic data remain.

We have shown that effects introduced by sparse sampling, especially the highly inhomogeneous window function and the additional zero-point effects, can be corrected for. The correlations introduced by sparse sampling techniques are not strong enough to destroy the additional information content gained by the additional survey volume and render sparse sampling a usefull tool in the observational arsenal. A cosmological test can either be completed by recovering the amplitude and shape of the input power spectrum using decomposition or the theoretical template itself can be convolved.

We used the realistic dataset to try and recover both the geometrical stretching factor and the amplitude of the power spectrum that can then be used to constrain cosmological models. The simulations show that all instrumental effects are under controll and do not decrease the accuracy much beyond the theoretical prediction of an ideal data set. They furthermore do not introduce strong systematic errors if the distribution of random errors is known well enough. This is especially important as larger surveys that are planned today, reduce statistical errors to a point where resulting estimates are heavily dominated by systematic errors.

Getting good estimates of the biasing is essential in predicting the accuracy and designing the parameters of an effective power spectrum survey. The higher the biasing, the more point density can be traded for area to increase the accuracy of the power spectrum estimation. However, the statistical data available for estimating the biasing is not very information rich.

We have investigated the available possibilities to recover the biasing from a pencil-beam like survey, that will be produced by the VIRUS prototype. We have found that the standard FKP estimator is, because of numerical effects, biased towards higher amplitudes. The correlation function seems to show this effect, too, though not as strongly developed. Using Fast Fourier Transforms of non-equal dimension size, we found that both the power spectrum and the correlation function give the same accuracies on a biasing estimate. This proves the robustness of both measures, as relative error bars and the amount of correlations in the recovered datasets vary greatly with both estimators.

The accuracy of the estimate is, however, strongly dependent on the geometry of the survey. More compact surveys yield much better results than elongated ones. This is true for surveys with one, two or three dimensions. For three dimensional surveys, we give a short formula to calculate the loss of information.

Finally, we describe the algorithms and accuracies of the CURE pipeline, both in reducing future VIRUS data and detecting line emission objects like LAEs. CURE is able to reduce the data provided by VIRUS and detect LAEs almost in real time and automatically without human intervention. Statistical evaluation of the sky-subtracted data shows that the reduction is robust and statistically consistent, even in the presence of cosmic rays and continuum sources. Accuracy test confirm that we reach the demanded 5-σ limit on

detections and are able to recover the positions of the source to sub fiber-diameter accuracy using dithered exposures.

Remaining issues, like a more accurate fiber model, and the ability to correct for pixel-to-pixel variations will be solved, once the fiber-profile of VIRUS is measured in more detail. At the time of this writing the CURE pipeline is used to reduce real data from the VIRUS-P and has found its first Lyman-alpha emitter candidates.

6.2 Outlook

The HETDEX survey is scheduled to start in 2011 and is supposed to be completed until 2013. Up to that time, we should have fairly robust knowledge about the statistics of the LAE properties from the VIRUS-P survey. We expect to get a biasing estimate with an accuracy of about 30%, and a very good knowledge of the surface density and luminosity function. This will help to finalize the already well researched survey parameters. The group around Eiichiro Komatsu is working hard to further refine their methods to pinpoint the biasing in the final dataset, model non-linear biasing and correct for redshift-space effects. Further help could be provided by progress in re-normalized perturbation theory (see Crocce & Scoccimarro, 2008).

For the final HETDEX results an in depth numerical simulation with about 10-100 times the volume of the actual HETDEX survey will be needed to assess the exact nature of the various effect that work on the power spectrum, including all observational and instrumental effects. One volume of the simulation should have a resolution of better than $1\,h-1\mathrm{Mpc}$ (that equals about $6\,h\mathrm{Mpc}^{-1}$ in k space) and a size of more than $1500\,h-1\mathrm{Mpc}$, available in a light cone output to include the evolution of the large scale structure with redshift. Such simulations have just become possible on modern day supercomputers (see e.g. Angulo et al., 2008; HORIZON, 2008; Kim et al., 2008) and will hopefully be available in multiple volumes by the time of 2011-2013. A full end-to-end simulation can then be made using the final survey parameters, including:

- redshift-dependent non-linear structure growth

- redshift and scale dependent biasing

- exact angular window-function of the survey

- exact radial window function of the survey

- Lyman-alpha emitter luminosity function

- contamination rates

- CURE detection probabilities

Only then a robust error estimation can be guaranteed.

This is especially important as modern surveys get bigger and bigger, both in depth and in angular dimension, to minimize statistical errors on the observed samples. These surveys are more and more limited by the systematic errors of the used instruments and estimation methods. Especially supernova experiments and weak lensing surveys are heavily dependent on the robustness of their assumptions and reduction methods. While BAOs are relatively robust, using the whole amplitude and shape of the power spectrum also introduces this problem into power spectrum cosmology that uses galaxies as tracers.

However, the statistical methods of analyzing the power spectrum can not only be refined and made robust, but also expanded combining multiple parameter dependencies to constrain cosmological models. The equal scaling of the amplitude of the power spectrum itself and the BAOs could be exploited to check for systematic errors. Redshift-space effects can be measured by doing a full three-dimensional analysis of the power spectrum (see Shoji et al., 2008; Guzzo et al., 2008). The biasing dependence of redshift-space effects can be used to further refine the information on biasing.

A large number of surveys is currently planned to constrain the equation of state of dark energy, w, and look for evolution of this value. Be it ground based surveys like Wiggelz[1], HETDEX[2], BOSS[3], and Pan-STARRS[4], or satellite missions like EUCLID (formerly DUNE[5] and SPACE[6]), SNAP[7] or PLANCK[8]. Methods like power spectrum cosmology, weak lensing tomography, or supernovae light cure measurements are used, discussed and refined. The equation of state of dark energy, or whatever model it will turn out to be, is currently getting a lot of attention and will be heavily constrained ten years from now.

At the moment, the Cosmological Constant is the best fitting model to the data, where w has to be exactly -1. Only very few models can currently be ruled out and a large zoo of theories awaits to be falsified. The question has been asked how accurate we will have to measure w if it was really -1? One might argue that there is no reason to stop measuring w if it was known to 1 ± 0.001 as it could still be both a cosmological constant, phantom energy that has a w of smaller than -1 or any dark energy model with $w > -1$. However, the same question could be asked about a couple of other measurements, like the equality of inertial mass and gravitational mass. People simply stop measuring when it gets to expensive.

As from now, the future of dark energy seems to be a bright one. Being one of the topics that heavily shape the view of our world and universe, it will be interesting to follow and participate in the progress to determine its properties.

---

[1]http:/wigglez.swin.edu.au

[2]http://www.hetdex.org

[3]http://www.sdss3.org

[4]http://www.ps1sc.org

[5]http://www.dune-mission.net

[6]http://www.spacesat.info

[7]http://www.snap.lbl.gov

[8]http://www.rssd.esa.int/index.php?project=Planck

Bibliography

Albrecht, A., Bernstein, G., Cahn, R., Freedman, W. L., Hewitt, J., Hu, W., Huth, J., & Kamionkowski, M. (2006). Repor tof the dark energy task force. *arXiv:astro-ph/0609591*.

Angulo, R., Baugh, C., Frenk, C. S., Bower, R. G., Jenkins, A., & Morris, S. L. (2005). Constraints on the dark energy equation of state from the imprint of the baryons on the power spectrum of clusters. *Monthly Notices of the Royal Astronomical Society*, *362*, L25–L29.

Angulo, R., Baugh, C., Frenk, C. S., & Lacey, G. C. (2008). The detectability of baryonic accoustic oscillations in future galaxy surveys. *Monthly Notices of the Royal Astronomical Society*, *383*, 755–776.

Bardeen, J., Bond, J. R., Kaiser, N., & Szalay, A. S. (1986). The statistics of peaks of gaussian random fields. *The Astrophysics Journal*, *304*, 15–61.

Bardeen, J. M., Bond, J. R., & Efstathiou, G. (1987). Cosmic fluctuation spectra with large-scale power. *The Astrophysical Journal*, *321*, 28–35.

Bardeen, J. M., Steinhardt, P. J., & Turner, M. S. (1983). Spontaneuous creation of almost scale-free density perturbations in an inflationary universe. *Physical Review D*, *28*, 679–693.

Baugh, C. M., Gaztanga, E., & Efstathiou, G. (1995). A comparison of the evolution of density fields in perturbation theory and numerical simulatons- ii. counts-in-cells analysis. *Monthly Notices of the Royal Astronomical Society*, *274*, 1049–1070.

Benson, A. J., Cole, S., Frenk, C. S., Baugh, C. M., & Lacey, C. G. (2000). The nature of galaxy bias and clustering. *Monthly Notices of the Royal Astronomical Society*, *311*, 793–808.

Blake, C., Collister, A., Bridle, S., & Lahav, O. (2007). Cosmological baryonic matter densities from 600000 sdss luminous red galaxies with photometric redshifts. *Monthly Notices of the Royal Astronomical Society*, *374*, 1527–1548.

Blake, C., D. Parkinson, B. B., Glazebrook, K., Kunz, M., & Nichol, R. C. (2006). Universal fitting formulae for baryon oscillation surveys. *Monthly Notices of the Royal Astronomical Society*, *365*, 255–264.

Blake, C., & Glazebrook, K. (2003). Probing dark energy using baryonic oscillations in the galaxy power spectrum ans a cosmological ruler. *The Astrophysical Journal*, *594*, 665–673.

Bond, J. R., & Efstathiou, G. (1987). The statistics of cosmic microwave background radiation fluctuations. *Monthly Notices of the Royal Astronomical Society*, *226*, 655–687.

Bridges, M., Lasenby, A. N., & Hobson, M. P. (2006). A bayesian analysis of the primordial power spectrum. *Monthly Notices of the Royal Astronomical Society*, *369*, 1132–1130.

Broadhurst, T. J., Ellis, R. S., Koo, D. C., & Szalay, A. S. (1990). Large-scale distribution of galaxies at the galactic poles. *Nature*, *343*, 726–728.

Caldwell, R. R., Dave, R., & Steinhardt, R. J. (1998). Cosmological imprint of an energy component with general equation of state. *Physical Review Letters*, *80*, 1582–1585.

Caldwell, R. R., & Linder, E. V. (2005). Limits of quintessence. *Physical Review Letters*, *95*, 141301.

Carroll, S. M., Hoffman, M., & Trodden, M. (2003). Can the dark energy equation of state parameter w be less than -1? *Physical Review D*, *68*, 023509.

Charlot, S., & Fall, S. M. (1993). Lyman-alpha emission from galaxies. *The Astrophysical Journal*, *415*, 580.

Cole, S., hatton, S., Weinberg, D. H., & Frenk, C. S. (1998). Mock 2df and sdss galaxy redshift surveys. *Monthly Notices of the Royal Astronomical Society*, *300*, 945–966.

Cole, S., Percival, W. J., Peacock, J. A., Norberg, P., & Baugh, C. M. (2005). The 2df galaxy redshift survey: power-spectrum analysis of the final data set and cosmological implications. *Monthly Notices of the Royal Astronomical Society*, *362*, 505–534.

Colombi, S., Szapudi, I., & Szalay, A. S. (1998). Effects of sampling on statistics of large-sclae structure. *Monthly Notices of the Royal Astronomical Society*, *296*, 253–274.

Copeland, E. J., Sami, M., & Shinji, T. (2006). Dynamics of dark energy. *arXiv:hep-th/0603057v3*.

Cowie, L. L., Barger, A. J., Hu, E. M., Capak, P., & Songaila, A. (2004). A large sample of spectrsocopic redshifts in the acs-goods region of the hubble deep field north. *The Astronomical Journal*, *127*, 3137–3145.

Crocce, M., & Scoccimarro, R. (2008). Nonlinear evolution of baryon acoustic oscillations. *Physical Review D*, *77*, 023533.

Dekel, A., & Lahav, O. (1999). Stochastic nonlinear galaxy biasing. *The Astrophysical Journal*, *520*, 24–34.

Delliou, M. L., Lacey, C. G., Baugh, C. M., & Morris, S. L. (2006). The properties of ly-alpha emitting galaxies in hierarchical galaxy formation models. *Monthly Notices of the Royal Astronomical Society*, *365*, 712–726.

Efstathiou, G., Bond, J., & White, S. (1992). Cobe background radiation anisotropies and large-scale structure in the universe. *Monthly Notices of the Royal Astronomical Society*, *258*, 1P–6P.

Eisenstein, D., & White, M. (2004). Theoretical uncertainty in baryon oscillations. *Physical Review D*, *70*, 103523.

Eisenstein, D. J., & Hu, W. (1998). Baryonic features in the matter transfer function. *The Astrophysical Journal*, *496*, 605–614.

Eisenstein, D. J., Hu, W., & Tegmark, M. (1998). Cosmic complementarity: H 0 and omega m from combining cosmic microwave background experiments and redshift surveys. *The Astrophysical Journal Letters*, *504*, 57.

Eisenstein, D. J., Zehavi, I., Hogg, D. W., Scoccimaro, R., Blanton, M. R., Nichol, R. C., Scranton, R., & Seo, H. (2005). Detection of the baryon accoustic peak in the large-scale correlation function of the sdss luminous red galaxies. *The Astrophysical Journal*, *633*, 560–574.

Evrard, A. E., MacFarland, T. J., Couchman, H. M. P., Colberg, J. M., Yoshida, N., White, S. D. M., Jenkins, A., & Frend, C. S. (2002). Galaxy clusters in hubble volume simulation: Cosmological constraints from sky survey populations. *The Astrophysical Journal*, *573*, 7–36.

Feldman, H. A., Kaiser, N., & Peacock, J. A. (1994). Power-spectrum analysis of three-dimensional redshift surveys. *The Astrophysical Journal*, *426*, 23–37.

Fernandez-Soto, A., Lanzetta, K. M., & Yahil, A. (1999). A new catalog of photometric resdshifts in the hubble deep field. *The Astrophysical Journal*, *513*, 34–50.

FFTW (2008). *FFTW - The Fastest Fourier Transform in the West*. URL http://www.fftw.org

Frenk, C. S., White, S. D. M., & Davis, M. (1983). Nonlinear evolution of large-scale strcuture in the universe. *The Astrophysical Journal*, *271*, 417–430.

Gawiser, E., Francke, H., Kamson, L., Schawinski, K., Gronwall, C., Ciardullo, R., Quadri, R., & Orsi, A. (2007). Ly-alpha emitting galaxies at z = 3.1:l* progenitors experiencing rapid star formation. *The Astrophysical Journal*, *671*, 278.

Gawiser, E., van Dokkum, P., , Gronwall, C., Ciardullo, R., Blanc, G. A., Castander, F. J., Feldmeier, J., & Francke, H. (2006). The physical nature of ly-alpha emitting galaxies at z=3.1. *The Astrophysical Journal*, *642*, L13–L16.

Glazebrook, K., & Blake, C. (2005). Measuring the cosmic evolution of dark energy with bayonic oscillaitons in the glaxy power spectrum. *The Astrophysical Journal*, *631*, 1–20.

Gössl, C. A., & Riffeser, A. (2002). Image reduction pipeline for the detection of variable sources in highly crowded fields. *Astronomy and Astrophysics*, *381*, 1095–1109.

Guth, A. H. (1981). Inflatonary universe: A possible solution to the horizon and flatness problem. *Physical Review D*, *23*, 347–356.

Guzzo, L., Perleoini, M., Meneux, B., Branchini, E., Fevre, O., Marinoni, C., Garilli, B., & Blaizot, J. (2008). A test of the nature of cosmic acceleration using galxy redshift distortions. *Nature*, *451*, 541–544.

Hamana, T., Ouchi, M., Shimasaku, K., Kayo, I., & Suto, Y. (2004). Properties of host haloes of lyman-break galaxies and lyman-α emitters from their number densities and angular clustering. *Monthly Notices of the Royal Astronomical Society*, *347*, 813–823.

Heidt, J., Appenzeller, I., Gabasch, A., Jaäger, K., Seitz, S., Bender, R., Böhm, A., & Snigula, J. (2004). The fors deep field:field selection, photometric observations and photometric catalog. *Astronomy and Astrophysics*, *398*, 49–61.

Hill, G., Gebhardt, K., Komatsu, E., Drory, N., & MacQueen, P. (2008). The hobby-eberly telecope dark energy experiment (hetdex): Description and early pilot survey results. *arXiv:0806.0183*.

Hoekstra, A., Mellier, Y., van Waerbeke, L., Sembolini, E., Fu, L., Hudson, M. J., Parker, L. C., Tereno, I., & Benabed, K. (2006). First cosmic shear results from the canada-france-hawaii telescope wide synoptic legacy survey. *The Astrophysical Journal*, *647*, 116–127.

HORIZON (2008). *The horizon project*.
URL `http://www.project-horizon.fr`

Hu, E. M., Cowie, L. L., & McMahon, R. G. (1998). The density of ly alpha emitters at very high redshift. *The Astrophysical Journal*, *502*, L99.

Hu, W., & Sugiyama, N. (1996). Small-scale Cosmological Perturbations: An Analytical Approach. *The Astrophysical Journal*, *471*, 542–570.

Hütsi, G. (2006). Clustering of sz clusters on a past light-cone: Acoustic oscillations and constraints on dark energy. *Astronomy and Astrophysics*, *446*, 43–60.

Jeong, D., & Komatsu, E. (2006). Perturbation theory reloaded: Analytical calculation of nonlinearity in baryonic acoustic oscillations in the reals-space matter power spectrum. *The Astrophysical Journal*, *651*, 651–619.

Jeong, D., & Komatsu, E. (2008). Perturbation theory reloaded ii: Non-linear bias, baryon acoustic oscillations and millenium simulation in real space. *arXiv:0805.2632v2*.

Jing, Y. (2005). Correcting for the alias effect when measuring the power spectrum using fast fourier transform. *The Astrophysical Journal*, *620*, 559–563.

Kaiser, N. (1986). A sparse-sampling strategy for the estimation of large-scale clustering from redshift surveys. *Monthly Notices of the Royal Astronomical Society*, *219*, 785–790.

Kaiser, N. (1998). Weak lensing and cosmology. *The Astrophysical Journal*, *498*, 26.

Kaiser, N., & Peacock, J. A. (1991). Power-spectrum analysis of one-dimensional redshift surveys. *The Astrophysical Journal*, *379*, 482–506.

Kim, J., Park, C., Gott, R., & Dubinski, J. (2008). The horizon run n-body simulation: Baryonic acoustic oscilations and tolopogy of large scale structure of the universe. *arXiv:0812.1392*.

Knox, L. (2006). Precision measurement of the mean curvature. *Physical Review D*, *73*, 023503.

Koehler, R. S. (2005). *Baryonic acoustic oscillations as a csomological probe*. Master's thesis, Ludwig-Maximilians-Universität.

Koehler, R. S., Schuecker, P., & Gebhardt, K. (2007). Probing dark energy with baryonic acoustic oscillations at high redshifts. *Astronomy and Astrophysics*, *426*, 7–20.

Kolb, E. W., Matarrese, S., & Riotto, A. (2006). On cosmic acceleration without dark energy. *New Journal of Physics*, *8*, 322.

Komatsu, E., Dunkley, J., Nolta, M. R., Bennett, C. L., Gold, B., Hinshaw, G., Jarosik, N., & Larson, D. (2008). Five-year wilkinson microwave anisotropy probe (wmap) observations: Cosmological interpretation. *arXiv:0803.0547*.

Landy, S. D., & Szalay, A. S. (1993). Bias and variance of angular correlation functions. *The Astrophysical Journal*, *412*, 64.

Lewis, A., Challinor, A., & Lasenby, A. (2000). Efficient computation of cosmic microwave background anisotropies in closed friedmann-robertrson-walker models. *The Astrophysical Journal*, *538*, 473–476.

Linder, E. V., & Jenkins, A. (2003). Cosmic structure growth and dark energy. *Monthly Notices of the Royal Astronomical Society*, *346*, 573–583.

Ma, C., Caldwell, R. R., Bode, P., & Wang, L. (1999). The mass power spectrum in quintessence cosmological models. *The Astrophysical Journal*, *521*, L1–L4.

Modgiliani, A., Hummel, W., Abuter, R., Amico, P., Ballester, P., Davies, R., Dumas, C., & Horrobin, M. (2007). The sinfoni pipeline. *arXiv:astro-ph/0701297*.

Monaco, P., Theuns, T., & Taffoni, G. (2002). The pinocchio algorithm; pinpointing orbit-crossing collapsed hierarchical objects in a linear density field. *Monthly Notices of the Royal Astronomical Society*, *331*, 587–608.

Nojiri, S., & Odintsov, S. D. (2006). Introduction to modified gravity and graviatational alternative for dark energy. *arXiv:hep-th/0601213*.

Noll, S., Mehlert, D., Appenzeller, I., bender, R., Böhm, A., Gabasch, A., Heidt, J., & Hopp, U. (2004). The fors deep field spectroscopic survey. *Astronomy and Astrophysics*, *418*, 885–906.

Orsi, A., Lacey, C., Baugh, C., & Infante, L. (2008). The clustering of ly-alpha emitters in a lambdacdm universe. *arXiv:0807.3347*.

Ota, K., Iye, M., Kashikawa, N., Shimasuka, K., Kobayashi, M., Totani, T., Nagashima, M., & Morokuma, T. (2008). Reionisation and galaxy evolution probed by $z = 7$ lyman-alpha emitters. *The Astrophysical Journal*, *677*, 12–26.

Ouchi, M., Shimasaki, K., Akiyama, M., Sekiguchi, K., Furusawa, H., Okamura, S., Kashikawa, N., & Iye, M. (2005). The discovery of primeval large-scale structures with forming clusters at redshift 6. *The Astrophysical Journal*, *620*, L1–L4.

Ouchi, M., Shimasaki, K., Akiyama, M., Simpson, C., Saito, T., Ueda, Y., Furusawa, H., & Sekiguchi, K. (2008). The subaru/xmm-newton deep survey (sxds). iv. evolution of ly-alpha emitters from z=3.1 to 5.7 in the $1\,deg^2$ field: Luminosity functions and agn. *The Astrophysical Journal Supplement Series*, *176*, 301.

Park, C., & Gott, J. R. (1990). Simulated deep redshift surveys. *Monthly Notices of the Royal Astronomical Society*, *249*, 288–299.

Peebles, P. (1993). *Principles of Physical Cosmology*. Princeton: Princeton University Press.

Peebles, P., & Ratra, B. (2003). The cosmological constant and dark energy. *Reviews of Modern Physics*, *75*, 559–606.

Percival, W. J., Baugh, C. M., Bland-Hawthorn, J., Bridges, T., Cannon, R., Cole, S., Colless, M., & Collins, C. (2001a). The 2df galaxy redshift survey: the power spectrum and the matter content of the universe. *The Astrophysical Journal*, *327*, 645–664.

Percival, W. J., Robert, C., Eisenstein, D. J., Frieman, J. A., Fukugita, M., Loveday, J., Pope, A. C., & Schneider, D. P. (2001b). The shape of the sloan digital sky survey data release 5 galaxy power spectrum. *Monthly Notices of the Royal Astronomical Society*, *327*, 1297–1306.

Perlmutter, S., Gabi, S., Goldhaber, G., Goobar, A., Groom, D. E., Hook, I. M., Kim, A. G., & Kim, M. Y. (1998). Discovery of a supernova explosion at half the age of the universe. *Nature*, *391*, 51.

Riess, A. G., Filippenko, A. V., Challis, P., Clochiatti, A., , Diercks, A., Garnavich, P. M., Gilliland, R. L., & Hogan, C. J. (1998). Observational evidence from supernovae for an accelerating unvierse and a cosmological constant. *The Astronomical Journal*, *116*, 1009–1038.

Riess, A. G., Strolger, L., Casertano, S., Ferguson, H. C., Mobasher, B., Gold, B., Challis, P. J., & Filippenko, A. V. (2007). New hubble space telescope discoveries of tpye ia supernovae at $z >= 1$: Narrowing constraints on the early behaviour of dark energy. *The Astrophysical Journal*, *659*, 98–121.

Sahni, V., & Shtanov, Y. (2003). Brane world models of dark energy. *Journal of Cosmology and Astroparticle Physics*, *11*, 014.

Sanchez, A. G., Baugh, C. M., & Angulo, R. (2008). What is the best way to measure the baryonic accoustic oscillations. *Monthly Notices of the Royal Astronomical Society*, *390*, 1470–1490.

Sanchez, S. (2006). Techniques for reducing fiber-fed and integral-field spectroscopy data: An implementation on r3d. *Astronomische Nachrichten*, *327*, 850.

Santos, M. R., Ellis, R. S., Kneib, J. P., Richard, J., & Kuijken, K. (2004). The abundance of low-luminosity la-alpha emitters at high redshift. *The Astrophysical Journal*, *606*, 683–701.

Schuecker, P. (2005). New cosmology with clusters of galaxies. *astro-ph/0502234*.

Schuecker, P., Ott, H., & Seitter, W. C. (1996). The muenster redshift projet. i. simulations of power spectra and analytical correlations. *The Astrophysical Journal*, *459*, 467.

Scodeggio, M., Franzetti, P., Garilli, B., Zachinelli, A., Paltani, S., Maccagani, D., Bottini, D., & Brun, V. L. (2005). The vvds data-reduction pipeline: Introducing vipgi and the vimos interactive pipeline and graphical interface. *The Astronomical Society of the Pacific*, *117*, 1284–1295.

Seo, H., & Eisenstein, D. J. (2003). Probing dark energy with baryonic acoustic oscillations from future large galaxy redshift surveys. *The Astrophysical Journal*, *598*, 720–740.

Seo, H., & Eisenstein, D. J. (2005). Baryonic accoustic oscillations in simulated galaxy redshift surveys. *The Astrophysical Journal*, *633*, 575–588.

Shani, V., & Shtanov, Y. (2003). Braneworld models of dark energy. *Journal of Cosmology and Astroparticle Physics*, *11*, 014.

Shoji, M., Jeong, D., & Komatsu, E. (2008). Extracting angular diameter distance and expansion rate of the universe from two-dimensional galaxy power spectrum at high redhsifts: Baryonic acoustic oscillations fitting versus full modeling. *arXiv:0805.4238*.

Smith, R. E., Peacock, J. A., Jenkins, A., White, S. D. M., Frenk, C. S., Pearce, F. R., Thomas, P. A., Efstathiou, G., & Couchman, H. M. P. (2003). Stable clustering, the halo model and non-linear cosmological power spectra. *Monthly Notices of the Royal Astronomical Society*, *341*, 1311–1332.

Smoot, G. F., Bennett, C. L., Kogout, A., Wright, E. L., Aymon, J., Boggess, N., Cheng, E., & de Amici, G. (1997). Structure of cobe differential microwave radiometer first-year maps. *The Astrophysical Journal*, *396*, L1–L5.

Somerville, R. S., Lemson, G., Sigad, Y., Dekel, A., Kauffmann, G., & White, S. D. M. (2001). Non-linear stochastic galaxy biasing in cosmological simulations. *Monthly Notices of the Royal Astronomical Society*, *320*, 289–306.

Springel, V., White, S. D. M., Jenkins, A., Frenk, C. S., Yoshida, N. ., Gao, L., Navarro, J., & Thacker, R. (2005). Simulations of the formation, evolution and clustering of galaxies and quasars. *Nature*, *435*, 629–636.

Steidel, C. C., Adelberger, K. L., Giavalisco, M., Dickinson, M., & Pettini, M. (1999). Lyman-break galaxies at z¿4 and the evolution of the ultra violet luminosity density at high redshift. *The Astrophysical Journal*, *519*, 1–17.

Steidel, C. C., Giavalisco, M., Pettini, M., Dickinson, M., & Adelberger, K. L. (1996). Spectroscopic confirmation of a population of normal star-forming galaxies at redshift *z* 3. *The Astrophysical Journal*, *462*, L17.

Taniguchi, Y., Ajiki, M., Nagao, T., Shioya, Y., Murayama, T., Kashikawa, N., Kodaira, K., & Kaifu, N. (2005). The subaru deep field project: Lyman-alpha emitters at a redshift of 6.6. *Publications of the Astronomical Society of Japan*, *57*, 165–182.

Tapken, C., Appenzeller, I., Noll, S., Richling, S., Heidt, J., Meinköhn, E., & Mehlert, D. (2007). Ly alpha emission in high-redshift galaxies. *Astronomy and Astrophysics*, *467*, 63–72.

Tegmark, M., Blanton, M. R., Strauss, M. A., Hoyle, F., Schlegel, D., Scoccimarro, R., Vogeley, M. S., & Weinberg, D. H. (2004). The three-dimensional power spectrum of galaxies from the sloan digital sky survey. *The Astrophysical Journal*, *606*, 702–740.

Tegmark, M., Eisenstein, D. J., Strauss, M. A., Weinberg, D. H., & Blanton, M. R. (2006). Cosmological constraints from sdss luminous red galaxies. *Physical Review D*, *74*, 123507.

Tegmark, M., Hamilton, A. J., & Xu, Y. (2002). The power spectrum of galaxies in the 2df 100k redshift survey. *Monthly Notices of the Royal Astronomical Society*, *335*, 887–908.

Tegmark, M., Taylor, A. N., & Heavens, A. F. (1997). Karhunen-loeve eigenvalue problems in cosmology: How should we tackle large data sets? *The Astrophysical Journal*, *480*, 22.

Turner, M. S., & White, M. (1997). Cdm models with a smooth component. *Physical Review D*, *56*, 4439.

Wagner, C., Mueller, V., & Steinmetz, M. (2008). Constraining dark energy via baryon acoustic oscillations in the (an)isotropic light-cone power spectrum. *Astronomy and Astrophysics*, *487*, 63–74.

Weinberg, S. (1989). The cosmological constant problem. *Reviews of Modern Physics*, *61*, 1–23.

White, M. (2005). Baryonic oscillations. *Astroparticle Physics*, *24*, 334–344.

White, S. M. D. (2007). Fundamentalist physics: why dark energy is bad for astronomy. *Reports on Progress in Physics*, *70*, 883–897.

Yoshida, N., Colberg, J., White, S. D. M., Evrard, A. E., MacFarland, T. J., Couchman, H., Jenkins, A., & Frenk, C. S. (2001). Simulations of deep pencil-beam redshift surveys. *Monthly Notices of the Royal Astronomical Society*, *325*, 803–816.

Acknowledgments

A lot of people have helped me finishing this thesis.

First and foremost I want to thank Peter Schuecker, who also supervised my Diploma thesis and unfortunately died during the early stages of my work for this thesis in November 2006. I learned a lot from him and I would have never made it so far without his knowledge and support.

Then I want to thank Ralf Bender, who made it possible for me to finish this work. His daily efforts in managing the group result in one of the most desireable working groups, where science is not only work, but also a lot of fun.

I want to thank Niv Drory who took over after Peter was gone, as well as Ulrich Hopp and Stephanie Phleps for guidance and co-supervision. Thanks for the support and thanks Niv, for the insights into C++. If there is something I will take from this work it is some hard earned knowledge about the language, which will always come in handy.

I want to thank my mother for supporting me till I could stand on my own feet (and helping me after that point, still).

I also want to thank the ever changing crew of room 360, for lots of interesting discussions, experiments about sneezing and blessing, and THE infamous experiment that may yet prove that man can create life in a bottle of Punica (or at least contaminate room 360 with permanent bio-hazard).

I want to thank my brother, again, for not eating everything in the fridge and tolerating some changes in my (and his) life, which certainly made my life better.

I want to thank the SessionMOB crew for fragging them with style, the MunichRPG group for lots of fun, and all my friends for the good time. Life is not only work, and it is good to have good people around you in your free time.

Finally I want to thank my fiance, for bringing lots of changes into my life, that make it more colorful and interesting. Life would be much poorer without you.

Curriculum Vitae

Ralf Stephan Koehler

15.10.1979	Geburt in München
08.1986 - 06.1990	Besuch der Grundschule an der Ostpreußenstraße
08.1990 - 06.1999	Besuch des Wilhelm Hausenstein Gymnasiums
06.1999 - 03.2000	Wehrdienst bei den Gebirgspionieren Brannenburg
10.2000 - 10.2005	Studium der Physik an der Ludwig-Maximilians-Universität München
10.2005 - 01.2009	Promotion in Physik an der Ludwig-Maximilians-Universität München