

CHAPTER 4B.

Cosmic Web Reconstruction



*Florent Leclercq working on
the reconstruction of the Universe.*



*Adi Nusser paying close attention,
looking for an opportunity to pose an incisive question.*

Big Data of the Cosmic Web

Francisco-Shu Kitaura

Leibniz Institute for Astrophysics (AIP),
An der Sternwarte 16, 14482 Potsdam
email: kitaura@aip.de

Abstract. One of the main goals in cosmology is to understand how the Universe evolves, how it forms structures, why it expands, and what is the nature of dark matter and dark energy. Next decade large and expensive observational projects will bring information on the structure and the distribution of many millions of galaxies at different redshifts enabling us to make great progress in answering these questions. However, these data require a very special and complex set of analysis tools to extract the maximum valuable information. Statistical inference techniques are being developed, bridging the gaps between theory, simulations, and observations. In particular, we discuss the efforts to address the question: What is the underlying nonlinear matter distribution and dynamics at any cosmic time corresponding to a set of observed galaxies in redshift space?

An accurate reconstruction of the initial conditions encodes the full phase-space information at any later cosmic time (given a particular structure formation model and a set of cosmological parameters). We present advances to solve this problem in a self-consistent way with Big Data techniques of the Cosmic Web.

Cosmology is experiencing a golden era. A large number of galaxy surveys are planned to produce an enormous avalanche of data. These aim at understanding the nature of dark matter and dark energy, two still unknown components, which make up about 95% of the whole energy budget in the Universe. The scientific goal is to unveil the accelerated expansion of the Universe and the hidden mechanisms of cosmic structure formation, which ultimately led to the place we occupy in the cosmos.

Answering these questions from analyzing vast amounts of data will demand complex data mining techniques able to extract the maximum cosmological information. In particular we aim at doing a global analysis of the data to break all possible degeneracies making the least possible assumptions. This implies using as input data in such a joint analysis the closest form to the raw data. Moreover, one would like to include as many data sets as possible, to combine them in a self-consistent way. These would range from the cosmic microwave background (CMB), over the 21 cm line, the Lyman alpha (and beta) forest, the Lyman alpha emitters, the distribution of quasars, galaxies, and clusters, to the corresponding lensing maps throughout cosmic history. Focusing for instance on just the galaxy distribution, a number of issues needs to be considered. These can be related to observational systematics, such as, the survey geometry; the completeness on the sky; the photometric calibration; the photometric redshift uncertainty (for photometric surveys); the stellar contamination; etc. Other systematic effects can be due to intrinsic physical aspects, such as, the nonlinear, nonlocal, stochastic, and assembly (luminosity dependent) galaxy bias; the coherent and dispersed peculiar velocities; the gravitational mode coupling; the baryonic effects; etc. The complexity of the problem scales dramatically with the volume and resolution we need to achieve, pushed by the requirements of the new generation of surveys covering increasingly larger volumes and fainter objects.

New galaxy surveys aim at going deeper in redshift by not only using longer exposure times to determine the redshifts with absorption spectra, but also exploiting the characteristic OII doublet seen in emission for strongly star forming galaxies (see, e.g., Dawson *et al.* 2013). Such objects are correspondingly called emission line galaxies (eLGs). As a consequence a large variety of galaxies tracing different density regimes of the large-scale structure will be available. In addition to full gravity calculations, effective theories based on analytical models of structure formation become necessary to shed light on the physical problem and save computational costs inherent to N -body based computations. We will discuss below how such models enable us to make a Big Data analysis of the cosmic web.

0.1. *The legacy of Zel'dovich*

The year 2014 was the commemoration of the 100th birth anniversary of Yakov Borisowitsch Seldowitsch, also known in English as Zel'dovich. While his contributions range from chemistry, over hydrodynamics, atomic nuclei, elementary particles to astrophysics, we want to highlight one of his main contributions to cosmology: the formation of the so-called Zel'dovich pancakes, i.e., the cosmic filamentary network. Zel'dovich proposed in 1970 an elegant solution to cosmological structure formation based on what we nowadays call linear Lagrangian perturbation theory, or Zel'dovich approximation, in which matter tracers move along the paths defined by the initial displacements at early times, or equivalently high redshifts. This approximation is able to describe remarkably well the quasi-nonlinear regime of structure formation and in particular the formation of the cosmic web. However, this picture led to the top-down scenario, where large structures form first and then are fragmented to form smaller ones, which is disfavored by observations. Simulations based on N -body solvers during the 80s and 90s, describing the interaction between matter tracers forming virialised structures, helped to develop the current bottom-up paradigm, in which smaller structures merge to form larger ones. After the accuracy of gravity solvers was found to be crucial to understand structure formation and computational progress made it possible to perform ever larger N -body simulations, the Zel'dovich approximation was confined during some period of time to academic and rather historical studies. Nevertheless, the need to understand the large-scale structure in rapidly increasing volumes pushed by the development of large galaxy redshift surveys starting during the 2000s with the Sloan Digital Sky Survey, has drawn the attention back to approximate gravity solvers, and analytical models. In fact only the order of a dozen large volume N -body simulations have been done, which do not achieve the resolution required for eLGs. Luckily some part of the astrophysics community never ceased investigating Zel'dovich's legacy and set the basis for a whole branch of methods, which are turning out to be very useful to analyze and understand observations of the large-scale structure.

Let us list here some of the methods which rely on the Zel'dovich approximation and find modern applications:

- **setup of initial conditions for N -body simulations**

The Zel'dovich approximation has turned out to be very useful to setup N -body simulations by relating the primordial fluctuations of the Universe to the initial velocities of matter particles (Springel *et al.* 2005). While starting at high enough redshifts ($\gtrsim 100$) the plane Zel'dovich approximation is still being used to setup initial conditions, more sophisticated versions have been developed including second order Lagrangian perturbation theory, i.e., tidal field corrections, which however follow the same idea introduced by Zel'dovich. These have become standard and a number of codes are publicly available (Crocce *et al.* 2006; Jenkins 2009), including primordial non-Gaussianities (Scoccimarro *et al.* 2012).

- **mock galaxy catalogues**

Full N -body simulations are necessary to enable an accurate understanding of structure formation, and, in particular, the distribution of galaxies. However, they are too expensive to be massively produced to compute covariance matrices, and, hence, derive the error bars corresponding to the measured galaxy clustering from observations. Nevertheless, these calculations are necessary to obtain reliable reference halo catalogues. One can aim at including all known physical processes in structure formation simulations, however covering very limited volumes (see, e.g., Illustris simulations, Vogelsberger *et al.* 2014). A number of techniques can convert dark matter halo catalogues into simulated galaxy ones, such as semi-analytic models (e.g., White & Frenk 1991, Kauffmann *et al.* 1993, Cole *et al.* 1994, Somerville & Primack 1999, Cole *et al.* 2000, Croton *et al.* 2006, De Lucia & Blaizot 2006, Benson 2012), halo abundance matching (HAM, e.g., Kravtsov *et al.* 2004, Tasitsiomi *et al.* 2004, Vale & Ostriker 2004, Conroy *et al.* 2006, Kim *et al.* 2008, Guo *et al.* 2010, Wetzel *et al.* 2010, Behroozi *et al.* 2010, Trujillo *et al.* 2011, Leauthaud *et al.* 2011), or halo occupation distribution (HOD, e.g., Berlind *et al.* 2002, Kravtsov *et al.* 2004, Zentner *et al.* 2005, Zehavi *et al.* 2005, Zheng *et al.* 2007, Skibba & Sheth 2009, Ross & Brunner 2009, Zheng *et al.* 2009, White *et al.* 2011) techniques. A number of methods have pioneered the efforts of producing halo catalogues with approximate gravity solvers, such as the peak-patch formalism (Bond & Myers 1996), Pinocchio (Monaco *et al.* 2002), and PThalos (Scoccimarro & Sheth 2002). These approaches were however far from achieving percentage accuracy in reproducing the clustering statistics, and therefore more recent methods have been developed, such as PATCHY (Kitaura *et al.* 2014) and EZmocks (Chuang *et al.* 2015). The latter ones rely on a set of effective bias parameters, which can be calibrated with reference catalogues to obtain the desired high accuracies in the different 1-, 2-, 3- point statistics, aiming at being precise at the 4-point statistics (covariance matrices). All these methods rely in some form or the other on the Zel'dovich approximation.

- **speed-up of N -body solvers**

Another recent idea relies on exploiting the accuracy of Lagrangian perturbation theory to efficiently solve gravity, and compute the remaining mode coupling term with an N -body solver, saving hereby a considerable amount of computational effort (Tassev *et al.* 2013). A similar idea is applied using the spherical collapse model for the short range force to yield approximate extremely efficient structure formation solvers using augmented Lagrangian perturbation theory (ALPT, Kitaura & Heß 2013).

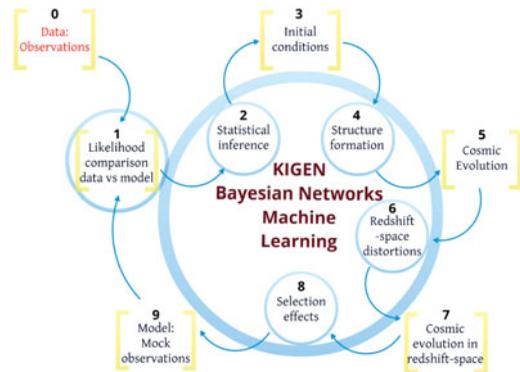
- **modification of cosmological parameters in an N -body simulation**

The cosmology of an N -body simulation can be modified for many purposes to great accuracy by using the Zel'dovich approximation to trace the dark matter particles to higher redshifts and then back to low redshifts, changing the cosmological parameters in this process (Angulo & White 2010).

- **modification of large-scale modes in an N -body simulation**

By constructing a simulation, which has only small-scale power, one can add different large-scale power realizations to it relying on the Zel'dovich approximation. Care must be taken to correct for the mode coupling originated through the large-scale modes, which cause a local modification of the growth. Therefore by looking up the positions of dark matter particles at different redshift snapshots in the N -body simulation, one can mimic that mode coupling. This technique allows to make a reduce number of N -body simulations combined with large numbers of Zel'dovich large-scale calculations to compute covariance matrices (Schneider *et al.* 2011).

Figure 1: Flowchart of the KIGEN-code based on a Bayesian Networks Machine Learning approach. The primordial density fluctuations (initial conditions) are obtained from iteratively sampling Gaussian fields, which lead to cosmic structures compatible with the distribution of galaxies given a particular structure formation model. Starting from some initial guess (steps 0-3) structure formation is simulated forward (steps 4-5), observational effects like redshift-space distortions caused by peculiar motions (steps 6-7) and selection function effects according to the magnitude limited survey (step 8) are taken into account and the resulting mock observations (step 9) are matched with the observations in a likelihood comparison process (steps 0-1). The results are used to improve the initial conditions in the next iteration.



- **modeling the correlation function of halo clustering**

The Zel'dovich approximation has been demonstrated to give a precise description of the gravitational mode coupling introduced in the baryon acoustic oscillations (Tassev & Zaldarriaga 2013). Hence it can be used to model the correlation function after cosmic evolution (McCullagh & Szalay 2012). Including galaxy bias and redshift space distortions originated by the peculiar motions of galaxies requires additional modeling (White 2015).

- **reconstruction of baryon acoustic oscillations**

The baryon acoustic oscillations (BAO) can be used as a standard ruler to measure the scale of the Universe at different epochs, and thereby study dark energy. However, they are distorted by gravitational evolution. To enhance the BAO signal one can undo gravity by moving the galaxies back in time using the Zel'dovich approximation (Eisenstein *et al.* 2007).

- **reconstruction of primordial fluctuations**

Lagrangian perturbation theory can be used to recover the initial conditions of the Universe on scales smaller than the BAO scale, as we will discuss below.

All the above-mentioned techniques enable an efficient analysis of the large-scale structure. We will discuss below how these techniques can be combined with Bayesian techniques to recover the full phase space information of the cosmic web.

1. Big Data of the Cosmic Web

The analysis of the large-scale structure from a galaxy distribution requires the characterization of the underlying dark matter field, which governs the dynamics. Although the primordial fluctuations are closely Gaussian distributed, gravity couples different modes and the formation of the cosmic web introduces an anisotropy in the three-point correlation function. One therefore needs, from a statistical point of view, to jointly constrain all the higher order moments of the dark matter distribution, the galaxy bias, and the

peculiar velocity field (Kitaura & Enßlin 2008). This can be a very complex task from the mathematical and computational point of view (Schaap & van de Weygaert 2000, Kitaura *et al.* 2010, Jasche & Kitaura 2010, Jasche *et al.* 2010, Platen *et al.* 2011, Kitaura *et al.* 2012c), especially including higher order correlation functions (e.g., Kitaura 2012).

The initial conditions of the Universe encode the full phase-space (density and peculiar velocity fields) information of any later cosmic time with a given cosmological structure formation model. It is thus tempting to reconstruct the initial conditions to characterize the large-scale structure.

Previous pioneering attempts to recover them have in most of the cases either ignored the relative movement of structures due to gravitation (see, e.g., Weinberg 1992, Kravtsov *et al.* 2002, Klypin *et al.* 03), or relied on linear theory (Nusser 1992, Kolatt *et al.* 1996, Mathis *et al.* 2002, Eisenstein *et al.* 2007, Padmanabhan *et al.* 2012, Doumler *et al.* 2013). Some nonlinear attempts can be found in the literature (see, e.g., Gramann93, Croft & Gaztanaga 1997, Narayanan 1998, Monaco & Efstathiou 1999, Kitaura & Angulo 2012). Other approaches have aimed at solving the boundary problem of finding the initial positions of a set of matter tracers governed by the Eulerian equation of motion and gravity with the least action principle (see Peebles 1989, Nusser *et al.* 2000, Branchini *et al.* 2002). A similar approach consists on relating the observed positions of galaxies in a geometrical way to a homogeneous distribution by minimizing a cost function (Frisch *et al.* 2002, Brenier *et al.* 2003, Lavaux 2010). All these approaches have one fundamental aspect in common: they aim at finding a single *optimal* solution to the initial conditions boundary problem.

Nevertheless, once shell-crossing starts, two matter tracers can have extremely close positions, but very different peculiar motions. It, thus, becomes impossible to know where the tracers came from in a unique way. Moreover, matter collapses to compact objects, which did not exist in the past, but had an extended Lagrangian region. Therefore, statistical forward approaches have been introduced (see the KIGEN-code: Kitaura 2013, Kitaura *et al.* 2012 and other approaches: Jasche & Wandelt 2013, Wang *et al.* 2013). While KIGEN is a particle-based approach, the other approaches are grid based. Patrick Bos presented in this IAU meeting the first method of the latter kind, which also includes a self-consistent treatment of RSD.

These methods exploit the fact that the statistical description of the density fields at the initial conditions is simple, as it must be closely Gaussian distributed. In particular the KIGEN-code permits one for the first time to deal with any kind of structure formation model in a probabilistic way. It is based on a Bayesian networks machine-learning algorithm, which iteratively samples Gaussian fields, whose phases are constrained by the distribution of observed tracers given a structure formation and a cosmological model. It also includes redshift-space distortions (coherent and virialised peculiar motions) in the likelihood comparison to the observations. In this sense, it is also the first self-consistent phase-space reconstruction method.

One can write down the Hamiltonian equations of motion for any analytical structure formation model, and sample from a posterior defined with such a model within a Bayesian approach. However, every time one changes the structure formation model one needs to rewrite the Hamiltonian sampler. Also it is very difficult to include in a self-consistent way RSD. The approach used in KIGEN splits the sampling problem into two main Gibbs sampling steps. The first step assumes that the positions of dark matter particles at the initial conditions tracing the primordial fluctuations are known. This is a statistical problem known in cosmology as the constrained realization (Hoffman & Ribak 1991, van de Weygaert & Bertschinger 1996), in which the Gaussian field is sampled compatible with a number of constraints. The second step is based on assuming that the

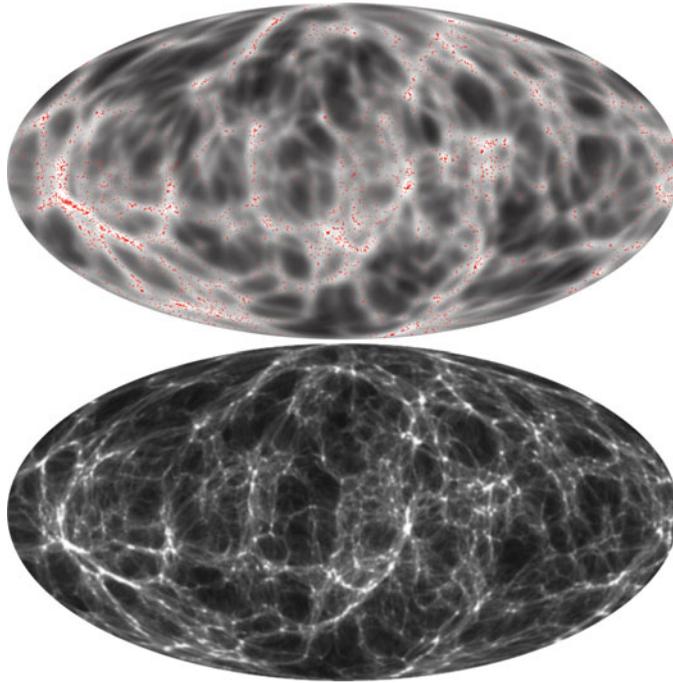


Figure 2: The upper panel shows the sky projection of all galaxies in the 2MRS catalog (red dots) at distances of 170 to 280 million light-years and their exquisite correlation with the mean over 25 reconstructed samples of the nonlinear ALPT cosmic web (grey scale) using the KIGEN code. The lower panel shows the same projection, but for the dark matter field of one constrained N -body simulation.

primordial Gaussian field is known, and given a set of constraints at low redshift, one needs to find the constraints at high redshift. This problem is also a statistical problem which needs to include a structure formation model translating the position of matter tracers at initial times to collapsed objects at late times, which can be in turn compared with the position of observed objects. A likelihood comparison process selects the collapsed objects, which are compatible under certain criteria (some minimum distance with a scatter). Since these collapsed objects have been obtained from a constrained simulation starting at initial times, we have full knowledge of its phase-space at all times. This enables us to look up the linking list of objects and the corresponding initial positions. Once we have them we can go to the first step. One can include RSD in the position of the objects and all kind of systematic effects. We note, that this approach is completely flexible to adopt any kind of structure formation model. As a generalization one can substitute positions of matter tracers with displacement fields (Kitaura in prep.). The method does not change. Here the idea is based on the reverse concept of setting initial conditions for an N -body simulation with the Zel'dovich approximation. In the same way, a Gaussian field determines the initial linear displacement field, the reverse is also true. One has only to consider that the likelihood comparison must be done with a nonlinear displacement, while the constrained realization must be done with the linear one.

The flowchart of the KIGEN Bayesian Networks Machine Learning approach is presented in Fig. 1. We note that the likelihood comparison can be used to improve the

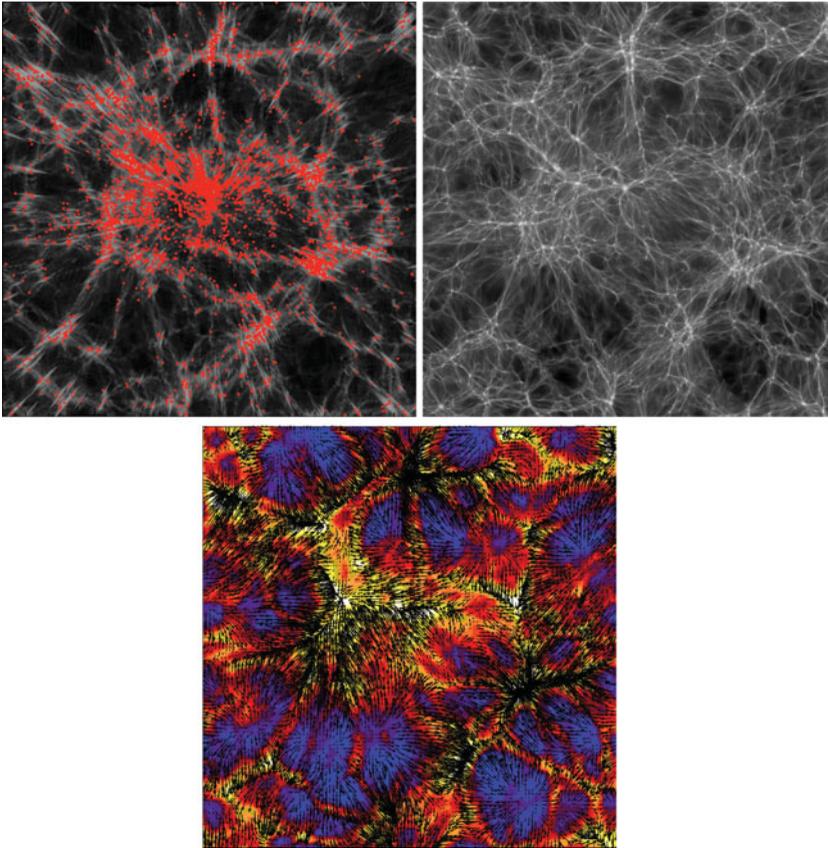


Figure 3: Left panel: slice through the super-galactic plane of the rendering of dark matter particles moved to redshift space from one particular constrained simulation presented in Heß *et al.* (2013), where the plots are instead shown based on the haloes to avoid an excess of virial motions, and the 2MRS galaxies (including $<5\%$ mocks in the galactic plane) are represented with red dots. Right panel: same slice, but after applying the phase-space mapping with the method presented in Abel *et al.* (2012). (Credit for the phase-space computation based on the reconstruction performed with KIGEN: Steffen Heß, Devon Powell, Ralf Kaehler & Tom Abel 2012). The lower panel shows the peculiar velocity field obtained with KIGEN using ALPT.

initial conditions since we have the full information available about the trajectories of the matter tracers from some starting high redshift until the redshift of the observations.

The KIGEN-code has been tested with a semi-analytic halo-model based galaxy mock catalog to demonstrate that the recovered initial conditions are closely unbiased with respect to the actual ones from the corresponding N-body simulation (see Kitaura 2013). It has also been applied to the Two-Micron All-Sky Redshift Survey (2MRS: Huchra *et al.* 2012) to perform a cosmography analysis and determine the proper motion of the Local Group finding a close agreement with the direction of the Cosmic Microwave Background (CMB) dipole and explaining about 80 % of its speed (see Figs. 2, 3 and Kitaura *et al.* 2012), and to search for the missing baryons in the warm hot inter-galactic medium (Suarez-Velasquez *et al.* 2013). A thorough analysis of the high performance of the KIGEN-code and its robustness with constrained N-body simulations has also

been done (see Heß *et al.* 2013). We have also investigated the cosmic web in the local Universe (Nuza *et al.* 2014). Here it could be shown that the relation between galaxy morphology and environment becomes clearer when properly correcting for RSD. The left panel in Fig. 3 shows that the likelihood comparison is done in redshift space. The knowledge of the primordial fluctuations opens a new possibility to analyze the cosmic web by using the full phase-space (Shandarin *et al.* 2012, Abel *et al.* 2012, Falck *et al.* 2012). We present a first application of such a method on observations in the right panel of Fig. 3, showing the reconstructed real-space cosmic web.

2. Conclusions

The huge amount of data from galaxy surveys will permit us to map the Universe with unprecedented accuracy. We have reached an era in which we need to develop complex data mining techniques to extract the hidden information in the data. We have shown that great advances have been carried in the study and characterization of the large-scale structure. Here we find that the statistics of the cosmic primordial fluctuations are well described by Gaussian distribution functions. Within a Bayesian framework, we can thus use simple priors and encode the physical structure formation models in the likelihood, when comparing to observations. Many developments need still to be done, to sample over the cosmological parameters, over the growth rate, and over the bias model, to include light-cone effects, and to include in the analysis additional cosmological probes, such as CMB, lensing, Lyman-alpha forest, etc. This is an exciting time to work in cosmology.

Acknowledgements

I thank the organizers of the IAU meeting for letting me give my first key note speaker talk at the celebration of Zel'dovich's anniversary in June 2014. I want to specially thank Steffen Heß for a wonderful collaboration. The upper plots in Fig. 3 have been made by him. I also thank Tom Abel and his group, in particular Devon Powell, for providing us the phase-space mapping code, which permitted us to make the first calculation of such kind, based on real observations. A more advanced calculation served for an article in National Geographic <http://ngm.nationalgeographic.com/2015/01/hidden-cosmos/fly-through-video>. All the material presented in this work had been presented in the original talk I gave in Tallin.

References

- T. Abel, O. Hahn, & R. Kaehler, *MNRAS*, **427**, 61, (2012)
- R. E. Angulo & S. D. M. White, *MNRAS*, **405**, 143, (2010)
- P. S. Behroozi, C. Conroy, & R. H. Wechsler, *ApJ*, **717**, 379, (2010)
- A. A. Berlind, & D. H. Weinberg, *ApJ*, **575**, 587, (2002)
- J. R. Bond & S. T. Myers, *ApJS*, **103**, 1, (1996)
- E. Branchini, A. Eldar, & A. Nusser, *MNRAS*, **335**, 53, (2002)
- A. J. Benson, *NA*, **17**, 175, (2012)
- Y. Brenier, U. Frisch, M. Henon, G. Loeper, S. Matarrese, R. Mohayaee, & A. Sobolevskii, *MNRAS*, **346**, 501, (2003)
- C.-H. Chuang, F.-S. Kitaura, F. Prada, C. Zhao, & G. Yepes, *MNRAS*, **446**, 2621, (2015)
- S. Cole, C. G. Lacey, C. M. Baugh, & C. S. Frenk, *MNRAS*, **319**, 168, (2000)
- S. Cole, A. Aragon-Salamanca, C. S. Frenk, J. F. Navarro, & S. E. Zepf, *MNRAS*, **271**, 781, (1994)
- C. Conroy, R. H. Wechsler, & A. V. Kravtsov, *ApJ*, **647**, 201, (2006)

- M. Crocce, S. Pueblas, & R. Scoccimarro, *MNRAS*, **373**, 369, (2006)
- R. A. C. Croft & E. Gaztanaga, *MNRAS*, **285**, 793, (1997)
- D. J. Croton, V. Springel, S. D. M. White, G. De Lucia, C. S. Frenk, L. Gao, A. Jenkins, G. Kauffmann, *et al.*, *MNRAS*, **365**, 11, (2006)
- K. S. Dawson, D. J. Schlegel, C. P. Ahn, S. F. Anderson, É. Aubourg, S. Bailey, R. H. Barkhouser, J. E. Bautista, *et al.*, *AJ*, **145**, 10, (2013)
- G. De Lucia & J. Blaizot, *MNRAS*, **375**, 2, (2007)
- & T. Doumler, Y. Hoffman, H. Courtois, & S. Gottlöber, 2013, *MNRAS*, **430**, 888
- D. J. Eisenstein, H.-J. Seo, E. Sirko, & D. N. Spergel, *ApJ*, **664**, 675, (2007)
- B. L. Falck, M. C. Neyrinck, & A. S. Szalay, *ApJ*, **323**, 1, (2012)
- U. Frisch, S. Matarrese, R. Mohayaee, & A. Sobolevski, *Nature*, **417**, 260, (2002)
- Q. Guo, S. White, C. Li, & M. Boylan-Kolchin, *MNRAS*, **404**, 1111, (2010)
- M. Gramann, *ApJ*, **405**, 449, (1993)
- S. Heß, F.-S. Kitaura, & S. Gottlöber, *MNRAS*, **435**, 3, 2065, (2013)
- Y. Hoffman & E. Ribak, *ApJL*, **380**, L5, (1991)
- J. P. Huchra, L. M. Macri, K. L. Masters, T. H. Jarrett, *et al.*, *Rev.Astrn.Astrophys.*, **199**, 26, (2012)
- J. Jasche, & F.-S. Kitaura, *MNRAS*, **407**, 29, (2010)
- J. Jasche, F.-S. Kitaura, C. Li, & T. A. Enßlin, *MNRAS*, **409**, 355, (2010)
- J. Jasche & B. D. Wandelt, *MNRAS*, **425**, 1042, (2013)
- A. Jenkins, *MNRAS*, **403**, 1859, (2009)
- G. Kauffmann & S. D. M. White, *MNRAS*, **261**, 921, (1993)
- J. Kim, C. Park, & Y.-Y. Choi, *ApJ*, **683**, 123, (2008)
- F.-S. Kitaura, *MNRAS*, **420**, 2737, (2012)
- F.-S. Kitaura, *MNRAS*, **429**, 1, L84, (2013)
- F.-S. Kitaura & R. E. Angulo, *MNRAS*, **425**, 4, 2443, (2012)
- F.-S. Kitaura, R. E. Angulo, Y. Hoffman & S. Gottlöber, *MNRAS*, **425**, 4, 2422, (2012)
- F.-S. Kitaura & T. A. Enßlin, *MNRAS*, **389**, 497, (2008)
- F.-S. Kitaura, J. Jasche, & R. B. Metcalf, *MNRAS*, **403**, 589, (2010)
- F.-S. Kitaura, P. Erdogdu, S. E. Nuza, A. Khalatyan, R. E. Angulo, Y. Hoffman & S. Gottlöber, *MNRAS*, **427**, 1, L35, (2012)
- F.-S. Kitaura, S. Gallerani, & A. Ferrara, *MNRAS*, **420**, 61, (2012)
- F.-S. Kitaura & S. Heß, *MNRAS*, **435**, 1, L78, (2013)
- F.-S. Kitaura, G. Yepes, & F. Prada, *MNRAS*, **439**, 21, (2014)
- A. Klypin, Y. Hoffman, A. V. Kravtsov, & S. Gottlöber, *ApJ*, **596**, 19, (2003)
- T. Kolatt, A. Dekel, G. Ganon, & J. A. Willick, *ApJ*, **458**, 419, (1996)
- A. V. Kravtsov, A. A. Berlind, R. H. Wechsler, A. A. Klypin, S. Gottlöber, B. Allgood, & J. R. Primack, *ApJ*, **609**, 35, (2004)
- A. V. Kravtsov, A. Klypin, & Y. Hoffman, *ApJ*, **571**, 563, (2002)
- G. Lavaux, *MNRAS*, **406**, 1007, (2010)
- A. Leauthaud, J. Tinker, P. S. Behroozi, M. T. Busha, & R. H. Wechsler, *ApJ*, **738**, 45, (2011)
- H. Mathis, G. Lemson, V. Springel, G. Kauffmann, S. D. M. White, A. Eldar, & A. Dekel, *MNRAS*, **333**, 739, (2002)
- N. McCullagh, & A. S. Szalay, *MNRAS*, **752**, 21, (2012)
- P. Monaco & G. Efstathiou, *MNRAS*, **308**, 763, (1999)
- P. Monaco, T. Theuns, G. Taffoni, F. Governato, T. Quinn, & J. Stadel, *ApJ*, **564**, 8, (2002)
- V. K. Narayanan & D. H. Weinberg, *ApJ*, **508**, 440, (1998)
- A. Nusser, & E. Branchini, *MNRAS*, **313**, 587, (2000)
- A. Nusser, & A. Dekel, *ApJ*, **391**, 443, (1992)
- S. E. Nuza, F.-S. Kitaura, S. Heß, N. I. Libeskind, & V. Müller, *MNRAS*, **445**, 988, (2014)
- N. Padmanabhan, X. Xu, D. J. Eisenstein, R. Scalzo, A. J. Cuesta, K. T. Mehta, & E. Kazin, *MNRAS*, **427**, 2132, (2012)
- J. Peebles, *ApJ*, **344**, L53, (1989)

- E. Platen, R. van de Weygaert, B. J. T. Jones, G. Vegter, & M. A. A. Calvo, *MNRAS*, **416**, 2494, (2011)
- A. J. Ross & R. J. Brunner, *MNRAS*, **399**, 878, (2009)
- W. E. Schaap & R. van de Weygaert, *AAP*, **363**, L29, (2000)
- M. D. Schneider, S. Cole, C. S. Frenk, & I. Szapudi, *ApJ*, **737**, 11, (2011)
- R. Scoccimarro, & R. K. Sheth, *MNRAS*, **329**, 629, (2002)
- R. Scoccimarro, L. Hui, M. Manera, & K. C. Chan, *PRD*, **85**, 083002, (2012)
- S. Shandarin, S. Habib, & K. Heitmann, *PRD*, **85**, 083005, (2012)
- R. A. Skibba & R. K. Sheth, *MNRAS*, **392**, 1080, (2009)
- R. S. Somerville & J. R. Primack, *MNRAS*, **310**, 1087, (1999)
- V. Springel, S. D. M. White, A. Jenkins, C. S. Frenk, N. Yoshida, L. Gao, J. Navarro, R. Thacker, *et al.*, *Nature*, **435**, 629, (2005)
- I. Suarez-Velasquez, F.-S. Kitaura, F. Atrio-Barandela & J. Mucket, *ApJ*, **769**, 1, 7, (2013)
- A. Tasitsiomi, A. V. Kravtsov, S. Gottlöber, & A. A. Klypin, *ApJ*, **607**, 125, (2004)
- S. Tassev & M. Zaldarriaga, *JCAP*, **4**, 13, (2013)
- S. Tassev, M. Zaldarriaga, & D. J. Eisenstein, *JCAP*, **6**, 36, (2013)
- S. Trujillo-Gomez, A. Klypin, J. Primack, & A. J. Romanowsky, *ApJ*, **742**, 16, (2011)
- A. Vale & J. P. Ostriker, *MNRAS*, **353**, 189, (2004)
- M. Vogelsberger, S. Genel, V. Springel, P. Torrey, D. Sijacki, D. Xu, G. Snyder, D. Nelson, & L. Hernquist, *MNRAS*, **444**, 1518, (2014)
- H. Wang, H. J. Mo, X. Yang, & F. C. van den Bosch, *ApJ*, **772**, 63, (2013)
- D. H. Weinberg, *MNRAS*, **254**, 315, (1992)
- A. R. Wetzel & M. White, *MNRAS*, **403**, 1072, (2010)
- R. van de Weygaert, & E. Bertschinger, *MNRAS*, **281**, 84, (1996)
- M. White, *MNRAS*, **450**, 3822, (2015)
- M. White, M. Blanton, A. Bolton, D. Schlegel, J. Tinker, A. Berlind, L. da Costa, E. Kazin, *et al.*, *ApJ*, **728**, 126, (2011)
- S. D. M. White, & C. S. Frenk, *ApJ*, **379**, 52, (1991)
- I. Zehavi, Z. Zheng, D. H. Weinberg, J. A. Frieman, A. A. Berlind, M. R. Blanton, R. Scoccimarro, R. K. Sheth *et al.*, *ApJ*, **630**, 1, (2005)
- Y. B. Zel'dovich, *AAP*, **5**, 84, (1970)
- A. R. Zentner, A. A. Berlind, J. S. Bullock, A. V. Kravtsov, & R. H. Wechsler, *ApJ*, **624**, 505, (2005)
- Z. Zheng, A. L. Coil & I. Zehavi, *ApJ*, **667**, 760, (2007)
- Z. Zheng, I. Zehavi, D. J. Eisenstein, D. H. Weinberg, & Y. P. Jing, *ApJ*, **707**, 554, (2009)