

Bayesian analysis of the dynamic cosmic web in the SDSS galaxy survey

Contents

| | | |
|------------|--|------------|
| 9.1 | Introduction | 132 |
| 9.2 | Methods | 134 |
| 9.2.1 | Bayesian large-scale structure inference with BORG | 134 |
| 9.2.2 | Non-linear filtering of samples with COLA | 135 |
| 9.2.3 | Classification of the cosmic web | 135 |
| 9.3 | The late-time large-scale structure | 136 |
| 9.3.1 | Tidal environment | 136 |
| 9.3.2 | Probabilistic web-type cartography | 137 |
| 9.3.3 | Volume and mass filling fractions | 140 |
| 9.4 | The primordial large-scale structure | 141 |
| 9.4.1 | Tidal environment | 142 |
| 9.4.2 | Probabilistic web-type cartography | 142 |
| 9.4.3 | Volume and mass filling fractions | 143 |
| 9.5 | Evolution of the cosmic web | 143 |
| 9.5.1 | Evolution of the probabilistic maps | 147 |
| 9.5.2 | Volume filling fraction | 147 |
| 9.5.3 | Mass filling fraction | 147 |
| 9.6 | Summary and Conclusion | 148 |

“I just wonder how things were put together.”
— Claude Shannon

Abstract

Recent application of the Bayesian algorithm BORG to the Sloan Digital Sky Survey main sample galaxies resulted in the physical inference of the formation history of the observed large-scale structure from its origin to the present epoch. In this work, we use these inferences as inputs for a detailed probabilistic cosmic web-type analysis. To do so, we generate a large set of data-constrained realizations of the large-scale structure using a fast, fully non-linear gravitational model. We then perform a dynamic classification of the cosmic web into four distinct components (voids, sheets, filaments, and clusters) on the basis of the tidal field. Our inference framework automatically and self-consistently propagates typical observational uncertainties to web-type classification. As a result, this study produces accurate cosmographic classification of large-scale structure elements in the SDSS volume. By also providing the history of these structure maps, the approach allows an analysis of the origin and growth of the early traces of the cosmic web present in the initial density field and of the evolution of global quantities such as the volume and mass filling fractions of different structures. For the problem of web-type classification, the results described in this chapter constitute the first connection between theory and observations at non-linear scales including a physical model of structure formation and the demonstrated capability of uncertainty quantification. A connection between cosmology and information theory using real data also naturally emerges from our probabilistic approach. Our results constitute quantitative chrono-cosmography of the complex web-like patterns underlying the observed galaxy distribution.

This chapter is adapted from its corresponding publication, [Leclercq, Jasche & Wandelt \(2015c\)](#).

9.1 Introduction

The large-scale distribution of matter in the Universe is known to form intricate, complex patterns traced by galaxies. The existence of this large-scale structure, also known as the *cosmic web* (Bond, Kofman & Pogosyan, 1996), has been suggested by early observational projects aiming at mapping the Universe (Gregory & Thompson, 1978; Kirshner *et al.*, 1981; de Lapparent, Geller & Huchra, 1986; Geller & Huchra, 1989; Shectman *et al.*, 1996), and has been extensively analyzed since then by massive surveys such as the 2dFGRS (Colless *et al.*, 2003), the SDSS (e.g. Gott *et al.*, 2005) or the 2MASS redshift survey (Huchra *et al.*, 2012). The cosmic web is usually segmented into different elements: voids, sheets, filaments, and clusters. At late times, low-density regions (voids) occupy most of the volume of the Universe. They are surrounded by walls (or sheets) from which departs a network of denser filaments. At the intersection of filaments lie the densest clumps of matter (clusters). Dynamically, matter tends to flow out of the voids to their compensation walls, transits through filaments and finally accretes in the densest halos.

Describing the cosmic web morphology is an involved task due to the intrinsic complexity of individual structures, but also to their connectivity and the hierarchical nature of their global organization. First approaches (e.g. Barrow, Bhavsar & Sonoda, 1985; Gott, Dickinson & Melott, 1986; Babul & Starkman, 1992; Mecke, Buchert & Wagner, 1994; Sahni, Sathyaprakash & Shandarin, 1998) often characterized the LSS with a set of global and statistical diagnostics, without providing a way to locally identify cosmic web elements. In the last decade, a variety of methods has been developed for segmenting the LSS into its components and applied to numerical simulations and observations. Among them, some focus on investigating one component at a time, in particular filaments (e.g. the Candy model – Stoica *et al.*, 2005; Stoica, Martínez & Saar, 2007, 2010, the skeleton analysis – Novikov, Colombi & Doré, 2006; Sousbie *et al.*, 2008, and DisPerSE – Sousbie, 2011; Sousbie, Pichon & Kawahara, 2011) or voids (e.g. Plionis & Basilakos, 2002; Colberg *et al.*, 2005; Shandarin *et al.*, 2006; Platen, van de Weygaert & Jones, 2007; Neyrinck, 2008; Sutter *et al.*, 2015b; Elyiv *et al.*, 2015, see also Colberg *et al.*, 2008 for a void finder comparison project). Unfortunately, this approach does not allow an analysis of the connections between cosmic web components, identified in the same framework. Another important class of web classifiers dissects clusters, filaments, walls, and voids at the same time. In particular, several recent studies deserve special attention due to their methodological richness. The “T-web” and “V-web” (Hahn *et al.*, 2007a; Forero-Romero *et al.*, 2009; Hoffman *et al.*, 2012) characterize the cosmic web based on the tidal and velocity shear fields. DIVA (Lavaux & Wandelt, 2010) rather uses the shear of the Lagrangian displacement field. ORIGAMI (Falck, Neyrinck & Szalay, 2012) identifies single and multi-stream regions in the full six-dimensional phase-space information (Abel, Hahn & Kaehler, 2012; Neyrinck, 2012; Shandarin, Habib & Heitmann, 2012). The Multiscale Morphology Filter (Aragón-Calvo *et al.*, 2007) and later refinements NEXUS/NEXUS+ (Cautun, van de Weygaert & Jones, 2013) follow a multiscale approach which probes the hierarchical nature of the cosmic web.

In the standard theoretical picture, the cosmic web arises from the anisotropic nature of gravitational collapse, which drives the formation of structure in the Universe from primordial fluctuations (Peebles, 1980). The capital importance of the large-scale tidal field in the formation and evolution of the cosmic web was first pointed out in the seminal work of Zel’dovich (1970). In the Zel’dovich approximation, the late-time morphology of structures is linked to the eigenvalues of the tidal tensor in the initial conditions. Gravitational collapse amplifies any anisotropy present in the primordial density field to give rise to highly asymmetrical structures. This picture explains the segmented nature of the LSS, but not its connectivity. The cosmic web theory of Bond, Kofman & Pogosyan (1996) asserted the deep connection between the tidal field around rare density peaks in the initial fluctuations and the final web pattern, in particular the filamentary cluster-cluster bridges. More generally, the shaping of the cosmic web through gravitational clustering is essentially a deterministic process described by Einstein’s equations and the main source of stochasticity in the problem enters in the generation of initial conditions, which are known, from inflationary theory, to resemble a Gaussian random field to very high accuracy (Guth & Pi, 1982; Hawking, 1982; Bardeen, Steinhardt & Turner, 1983). For these reasons, considerable effort has been devoted to a theoretical understanding of the LSS in terms of perturbation theory in the Eulerian and Lagrangian frames (for a review, see Bernardeau *et al.*, 2002). While this approach offers important analytical insights, it only permits to describe structure formation in the linear and mildly non-linear regimes and it is usually limited to the first few correlation functions of the density field. The complete description of the connection between primordial fluctuations and the late-time LSS, including a full phase-space treatment and the entire hierarchy of correlators, has to rely on a numerical treatment through N -body simulations. The characterization of cosmic web environments in the non-linear regime and

the description of their time evolution has only been treated recently, following the application of web classifiers to state-of-the-art simulations. In particular, [Hahn *et al.* \(2007a\)](#); [Aragón-Calvo, van de Weygaert & Jones \(2010\)](#) presented a local description of structure types in high-resolution cosmological simulations. [Hahn *et al.* \(2007b\)](#); [Bond, Strauss & Cen \(2010\)](#); [Cautun *et al.* \(2014\)](#) analyzed the time evolution of the cosmic web in terms of the mass and volume content of web-type components, their density distribution, and a set of new analysis tools especially designed for particular elements.

To the best of our knowledge, neither the classification of cosmic environments at non-linear scales in physical realizations of the LSS nor the investigation of their genesis and growth, using real data and with demonstrated capability of uncertainty quantification, have been treated in the existing literature. In this work, we propose the first probabilistic web-type analysis conducted with observational data in the deeply non-linear regime of LSS formation. We build accurate maps of dynamic cosmic web components with a resolution of around $3 \text{ Mpc}/h$, constrained by observations. In addition, our approach leads to the first quantitative inference of the formation history of these environments and allows the construction of maps of the embryonic traces in the initial perturbations of the late-time morphological features of the cosmic web.

Cosmographic descriptions of the LSS in terms of three-dimensional maps, and in particular a dynamic structure type cartography carry potential for a rich variety of applications. Such maps characterize the anisotropic nature of gravitational structure formation, the clustering behavior of galaxies as a function of their tidal environment and permit to describe the traces of the cosmic web already imprinted in the initial conditions. So far, most investigations focused on understanding the physical properties of dark halos and galaxies in relation to the LSS. [Hahn *et al.* \(2007a,b, 2009\)](#); [Hahn, Angulo & Abel \(2015\)](#); [Aragón-Calvo, van de Weygaert & Jones \(2010\)](#) found a systematic dependence of halo properties such as morphological type, color, luminosity and spin parameter on their cosmic environment (local density, velocity and tidal field). In addition, a correlation between halo shapes and spins and the orientations of nearby filaments and sheets, predicted in simulations ([Altay, Colberg & Croft, 2006](#); [Hahn *et al.*, 2007a,b, 2009](#); [Paz, Stasyszyn & Padilla, 2008](#); [Zhang *et al.*, 2009](#); [Codis *et al.*, 2012](#); [Libeskind *et al.*, 2013](#); [Welker *et al.*, 2014](#); [Aragon-Calvo & Yang, 2014](#); [Laigle *et al.*, 2015](#)), has been confirmed by observational galaxy data ([Paz, Stasyszyn & Padilla, 2008](#); [Jones, van de Weygaert & Aragón-Calvo, 2010](#); [Tempel, Stoica & Saar, 2013](#); [Zhang *et al.*, 2013](#)). Cartographic descriptions of the cosmic web also permit to study the environmental dependence of galaxy properties (see e.g. [Lee & Lee, 2008](#); [Lee & Li, 2008](#); [Park, Kim & Park, 2010](#); [Yan, Fan & White, 2012](#); [Kovač *et al.*, 2014](#)) and to make the connection between the sophisticated predictions for galaxy properties in hydrodynamic simulations (e.g. [Vogelsberger *et al.*, 2014](#); [Dubois *et al.*, 2014](#); [Codis *et al.*, 2015](#)) and observations. Another wide range of applications of structure type reconstructions is to probe the effect of the inhomogeneous large-scale structure on photon properties and geodesics. For example, it is possible to interpret the weak gravitational lensing effects of voids ([Melchior *et al.*, 2014](#); [Clampitt & Jain, 2015](#)). Dynamic information can also be used to produce prediction templates for secondary effects expected in the cosmic microwave background such as the kinetic Sunyaev-Zel'dovich effect ([Li *et al.*, 2014](#)), the integrated Sachs-Wolfe and Rees-Sciama effects (e.g. [Cai *et al.*, 2010](#); [Ilić, Langer & Douspis, 2013](#); [Planck Collaboration, 2014a](#)). Lastly, as the cosmic web morphology arises from gravitational instability, it can be used to test general relativity ([Falck, Koyama & Zhao, 2015](#)).

Building such refined cosmographic descriptions of the Universe requires high-dimensional, non-linear inferences. In chapter 5 ([Jasche, Leclercq & Wandelt, 2015](#)), we presented a chrono-cosmography project, aiming at reconstructing simultaneously the density distribution, the velocity field and the formation history of the LSS from galaxies. To do so, we used an advanced Bayesian inference algorithm to assimilate the Sloan Digital Sky Survey DR7 data into the forecasts of a physical model of structure formation (second order Lagrangian perturbation theory). Besides inferring the four-dimensional history of the matter distribution, these results permit us an analysis of the genesis and growth of the complex web-like patterns that have been observed in our Universe. Therefore, this work constitutes a new chrono-cosmography project, aiming at the analysis of the evolving cosmic web.

Our investigations rely on the inference of the initial conditions in the SDSS volume (see chapter 5). Starting from these, we generate a large set of constrained realizations of the Universe using the COLA method ([Tassev, Zaldarriaga & Eisenstein, 2013](#), see also section 7.3.1). This physical model allows us to perform the first description of the cosmic web in the non-linear regime, using real data, and to follow the time evolution of its constituting elements. Throughout this chapter, we adopt the [Hahn *et al.* \(2007a\)](#) dynamic “T-web” classifier, which segments the LSS into voids, sheets, filaments, and clusters. This choice is motivated by the close relation between the equations that dictate the dynamics of the growth of structures in the Zel'dovich formalism and the Lagrangian description of the LSS which naturally emerges with BORG. As this procedure

relies on the estimation of the eigenvalues of the tidal tensor in Fourier space, it constitutes a non-linear and non-local estimator of structure types, requiring adequate means to propagate observational uncertainties to finally inferred products (web-type maps and all derived quantities), in order not to misinterpret results. The BORG algorithm naturally addresses this problem by providing a set of density realizations constrained by the data. The variation between these samples constitute a thorough quantification of uncertainty coming from all observational effects (in particular the incompleteness of the data because of the survey mask and the radial selection functions, as well as luminosity-dependent galaxy biases, see chapter 5 for details), not only with a point estimation but with a detailed treatment of the likelihood. Hence, for all problems addressed in this work, we get a fully probabilistic answer in terms of a prior and a posterior distribution. Building upon the robustness of our uncertainty quantification procedure, we are able to make the first observationally-supported link between cosmology and information theory (see Neyrinck, 2015, for theoretical considerations related to this question) by looking at the entropy and Kullback-Leibler divergence of probability distribution functions.

This chapter is organized as follows. In section 9.2, we describe our methodology: Bayesian large-scale structure inference with the BORG algorithm, non-linear filtering of samples with COLA and web-type classification using the T-web procedure. In sections 9.3 and 9.4, we describe the cosmic web at present and primordial times, respectively. In section 9.5, we follow the time evolution of web-types as structures form in the Universe. Finally, we summarize our results and offer concluding comments in section 9.6.

9.2 Methods

In this section, we describe our methodology step by step:

1. inference of the initial conditions with BORG (section 9.2.1),
2. generation of data-constrained realizations of the SDSS volume via non-linear filtering of BORG samples with COLA (section 9.2.2),
3. classification of the cosmic web in voids, sheets, filaments, and clusters, using the T-web algorithm (section 9.2.3).

9.2.1 Bayesian large-scale structure inference with BORG

This work builds upon results previously obtained by application of the BORG (Bayesian Origin Reconstruction from Galaxies, Jasche & Wandelt, 2013a) algorithm to the Sloan Digital Sky Survey data release 7 (Jasche, Leclercq & Wandelt, 2015). BORG is a full-scale Bayesian inference code which permits to simultaneously analyze morphology and formation history of the cosmic web (see chapters 4 and 5 for a complete description).

As discussed in Jasche & Wandelt (2013a), accurate and detailed cosmographic inferences from observations require modeling the mildly non-linear and non-linear regime of the presently observed matter distribution. The exact statistical behavior of the LSS in terms of a full probability distribution function for non-linearly evolved density fields is not known. For this reason, the first full-scale reconstructions relied on phenomenological approximations, such as multivariate Gaussian or log-normal distributions, incorporating a cosmological power spectrum to accurately represent correct two-point statistics of density fields (see e.g. Lahav *et al.*, 1994; Zaroubi, 2002; Erdođdu *et al.*, 2004; Kitaura & Enßlin, 2008; Kitaura *et al.*, 2009; Kitaura, Jasche & Metcalf, 2010; Jasche & Kitaura, 2010; Jasche *et al.*, 2010b,a). However, these prescriptions only model the one and two-point statistics of the matter distribution. Additional statistical complexity of the evolved density field arises from the fact that gravitational structure formation introduces mode coupling and phase correlations. This manifests itself not only in a sheer amplitude difference of density and velocity fields at different redshifts, but also in a modification of their statistical behavior by the generation of higher-order correlation functions. An accurate modeling of these high-order correlators is of crucial importance for a precise description of the connectivity and hierarchical nature of the cosmic web, which is the aim of this chapter.

While the statistical nature of the late-time density field is poorly understood, the initial conditions from which it formed are known to obey Gaussian statistics to very great accuracy (Planck Collaboration, 2015). Therefore, it is reasonable to account for the increasing statistical complexity of the evolving matter distribution by a dynamical model of structure formation linking initial and final conditions. This naturally turns the problem of LSS analysis to the task of inferring the initial conditions from present cosmological observations (Jasche & Wandelt, 2013a; Kitaura, 2013; Wang *et al.*, 2013). This approach yields a very high-dimensional

and non-linear inference problem. Typically, the parameter space to explore comprises on the order of 10^6 to 10^7 elements, corresponding to the voxels of the map to be inferred. For reasons linked to computational cost, the BORG algorithm employs 2LPT as an approximation for the actual gravitational dynamics linking initial three-dimensional Gaussian density fields to present, non-Gaussian density fields. As known from perturbation theory (see e.g. [Bernardeau *et al.*, 2002](#)), in the linear and mildly non-linear regime, 2LPT correctly describes the one-, two- and three-point statistics of the matter distribution and also approximates very well higher-order correlators. It accounts in particular for tidal effects in its regime of validity. Consequently, the BORG algorithm correctly transports the observational information corresponding to complex web-like features from the final density field to the corresponding initial conditions. Note that such an explicit Bayesian forward-modeling approach is always more powerful than constraining (part of) the sequence of correlation functions, as it accounts for the entire dark matter dynamics (in particular for the infinite hierarchy of correlators), in its regime of validity. This is of particular importance, since the hierarchy of correlation functions has been shown to be an insufficient description of density fields in the non-linear regime ([Carron, 2012](#); [Carron & Neyrinck, 2012](#)).

As discussed in chapter 5 ([Jasche, Leclercq & Wandelt, 2015](#)), our analysis comprehensively accounts for observational effects such as selection functions, survey geometry, luminosity-dependent galaxy biases and noise. Corresponding uncertainty quantification is provided by sampling from the high-dimensional posterior distribution via an efficient implementation of the Hamiltonian Markov Chain Monte Carlo method (see chapter 4 and [Jasche & Wandelt, 2013a](#), for details). In particular, luminosity-dependent galaxy biases are explicitly part of the BORG likelihood and the bias amplitudes are inferred self-consistently during the run. Though not explicitly modeled, redshift-space distortions are automatically mitigated: due to the prior preference for homogeneity and isotropy, such anisotropic features are treated as noise in the data.

In the following, we make use of the 12,000 samples of the posterior distribution for primordial density fields, obtained in chapter 5. These reconstructions, constrained by SDSS observations, act as initial conditions for the generation of constrained large-scale structure realizations. It is important to note that we directly make use of BORG outputs without any further post-processing, which demonstrates the remarkable quality of our inference results.

9.2.2 Non-linear filtering of samples with COLA

In section 2.3 ([Leclercq *et al.*, 2013](#), section 2.A), we performed a study of differences in the representation of structure types in density fields predicted by LPT and N -body simulations. To do so, we used the same web-type classification procedure as in this work (see sections 9.2.3 and C.2). In spite of the visual similarity of LPT and N -body density fields at large and intermediate scales (above a few Mpc/h), we found crucial differences in the representation of structures. Specifically, LPT predicts fuzzier halos than full gravity, and incorrectly assigns the surroundings of voids as part of them. This manifests itself in an overprediction of the volume occupied by clusters and voids at the detriment of sheets and filaments. The substructure of voids is also known to be incorrectly represented in 2LPT ([Sahni & Shandarin, 1996](#); [Neyrinck, 2013](#); [Leclercq *et al.*, 2013](#)).

For these reasons, in this chapter we cannot directly make use of the final BORG density samples, which are a prediction of the 2LPT model. Instead, we rely on the inferred initial conditions, which contain the data constraints (as described in chapter 5) and on a non-linear filtering step (see chapter 7) similar to the one described in chapter 8 ([Leclercq *et al.*, 2015](#)). Due to the large number of samples to be processed for this work, we do not use a fully non-linear simulation code as in chapter 8, but the COLA method ([Tassev, Zaldarriaga & Eisenstein, 2013](#), see also section 7.3.1).

The generation of the set of non-linear BORG-COLA samples, used in this chapter, is described in section 7.3.2.

9.2.3 Classification of the cosmic web

The BORG filtered reconstructions permit a variety of scientific analyses of the large scale structure in the observed Universe. In this work, we focus specifically on the possibility to characterize the cosmic web by distinct structure types. Generally, any of the methods cited in the introduction (section 9.1) can be employed for analysis of our density samples, however for the purpose of this chapter, we follow the ‘‘T-web’’ classification procedure as proposed by [Hahn *et al.* \(2007a\)](#), described in section C.2.

The basic idea of this dynamical classification approach is that the eigenvalues $\mu_1 \leq \mu_2 \leq \mu_3$ of the tidal tensor $\mathcal{T}_{ij} \equiv \text{H}(\tilde{\Phi})_{ij}$ (Hessian of the rescaled gravitational potential) characterize the geometrical properties of each point in space. With these definitions, the three eigenvalues of the tidal tensor form a decomposition of the density contrast field, in the sense that the trace of \mathcal{T} is $\mu_1 + \mu_2 + \mu_3 = \delta$. Each spatial point can then be classified as a specific web type by considering the signs of μ_1, μ_2, μ_3 , according to the rules given in table C.1.

Several extensions of this classification procedure exist, that permit different classification up to sub-megaparsec scales (see section C.2.4). In this work, we will probe scales down to $\sim 3 \text{ Mpc}/h$ (the voxel size in our reconstructions). Therefore, we will be content with the original classification procedure as proposed by Hahn *et al.* (2007a).

It is important to note that the tidal tensor and the rescaled gravitational potential are both physical quantities, and hence their calculation requires the availability of a full physical density field in contrast to a smoothed mean reconstruction of the density field. As described in chapter 5, density samples obtained by the BORG algorithm provide such required full physical density fields. The tidal tensor can therefore easily be calculated in each density sample from the Fourier space representations of equations (C.6) and (C.7) (see section C.2.5 and Hahn *et al.*, 2007a; Forero-Romero *et al.*, 2009, for details on the technical implementation).

The web classifier provides four voxel-wise scalar fields that characterize the large scale structure. In a specific realization, the answer is unique, meaning that these fields obey the following conditions at each voxel position \vec{x}_k :

$$\mathbb{T}_i(\vec{x}_k) \in \{0, 1\} \text{ for } i \in \llbracket 0, 3 \rrbracket \quad \text{and} \quad \sum_{i=0}^3 \mathbb{T}_i(\vec{x}_k) = 1 \quad (9.1)$$

where $\mathbb{T}_0 = \text{void}$, $\mathbb{T}_1 = \text{sheet}$, $\mathbb{T}_2 = \text{filament}$, $\mathbb{T}_3 = \text{cluster}$. In this work, we follow the Bayesian approach of Jasche, Leclercq & Wandelt (2015) and quantify the degree of belief in structure type classification. Specifically, our web classification is given in terms of four voxel-wise scalar fields that obey the following conditions at each voxel position \vec{x}_k :

$$\mathcal{T}_i(\vec{x}_k) \in [0, 1] \text{ for } i \in \llbracket 0, 3 \rrbracket \quad \text{and} \quad \sum_{i=0}^3 \mathcal{T}_i(\vec{x}_k) = 1. \quad (9.2)$$

Here, $\mathcal{T}_i(\vec{x}_k) \equiv \langle \mathbb{T}_i(\vec{x}_k) \rangle_{\mathcal{P}(\mathbb{T}_i(\vec{x}_k)|d)} = \mathcal{P}(\mathbb{T}_i(\vec{x}_k)|d)$ are the posterior probabilities indicating the possibility to encounter specific structure types at a given position in the observed volume, conditional on the data. These are estimated by applying the web classification to all density samples and counting the relative frequencies at each individual spatial coordinate within the set of samples (see section 5 in Jasche *et al.*, 2010b). With this definition, the cosmic web-type posterior mean is given by

$$\langle \mathcal{P}(\mathbb{T}_i(\vec{x}_k)|d) \rangle = \frac{1}{N} \sum_{n=1}^N \sum_{j=0}^3 \delta_{\mathbb{T}_i(\vec{x}_k) \mathbb{T}_j^n(\vec{x}_k)}, \quad (9.3)$$

where n labels one of the N samples, $\mathbb{T}_j^n(\vec{x}_k)$ is the result of the web classifier on the n -th sample (i.e. a unit four-vector at each voxel position \vec{x}_k containing zeros except for one component, which indicates the structure type), and $\delta_{\mathbb{T}_i \mathbb{T}_j}^{ab}$ is a Kronecker symbol.

9.3 The late-time large-scale structure

In this section, we discuss the results of our analysis of the final density field, at $a = 1$. For reasons of computational time with COLA filtering (see section 9.2.2), we kept around 10% of the original set of samples obtained in chapter 5. In order to mitigate as much as possible the effects of correlation among samples, we maximally separated the samples kept for the present analysis, keeping one out of ten consecutive samples of the original Markov Chain. Hence, for all results discussed in this section, we used a total of 1,097 samples inferred by BORG and filtered with COLA.

9.3.1 Tidal environment

As a natural byproduct, the application of the T-web classifier to density samples yields samples of the pdfs for the three eigenvalues of the tidal field tensor. These pdfs account for the assumed physical model of structure formation and the data constraints, and quantify uncertainty coming in particular from selection

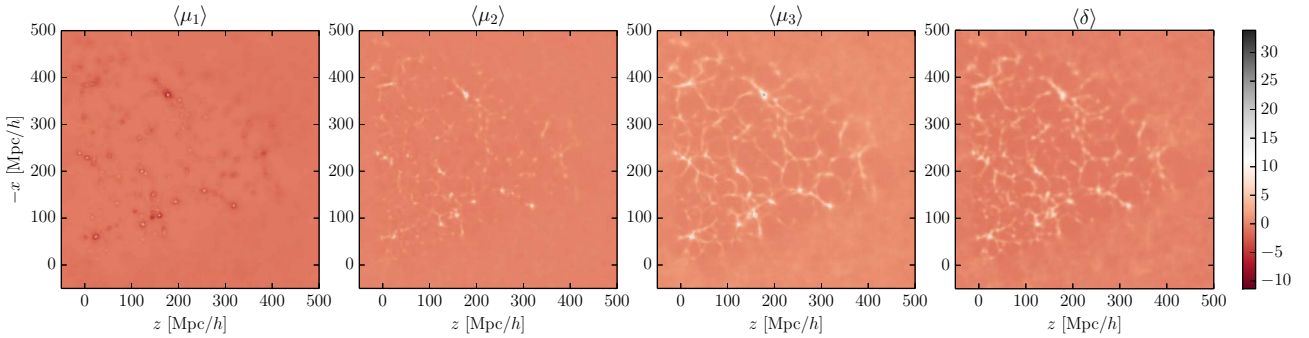


Figure 9.1: Slices through the three-dimensional ensemble posterior mean for the eigenvalues $\mu_1 \leq \mu_2 \leq \mu_3$ of the tidal field tensor in the final conditions, estimated from 1,097 samples. The rightmost panel shows the corresponding slice through the posterior mean for the final density contrast $\delta = \mu_1 + \mu_2 + \mu_3$, obtained in section 5.3.1.

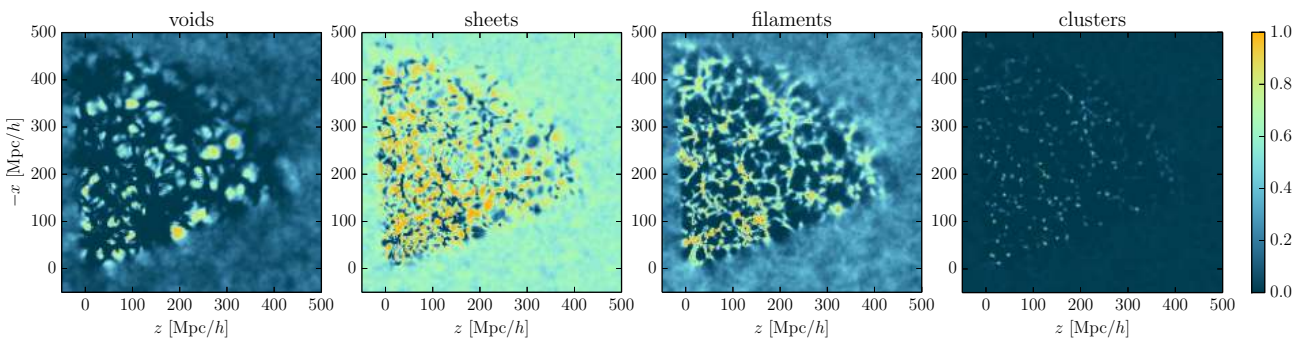


Figure 9.2: Slices through the posterior mean for different structure types (from left to right: void, sheet, filament, and cluster) in the late-time large-scale structure in the Sloan volume ($a = 1$). These four three-dimensional voxel-wise pdfs sum up to one on a voxel basis.

effects, surveys geometries and galaxy biases. In a similar fashion as described in section 5.3, the ensemble of samples permits us to provide any desired statistical summary such as mean and variance.

In figure 9.1, we show slices through the ensemble mean fields μ_1 , μ_2 and μ_3 . For visual comparison, the rightmost panel of figure 9.1 shows the corresponding slice through the posterior mean of the final density contrast, $\delta = \mu_1 + \mu_2 + \mu_3$, obtained in section 5.3.1. Different morphologies can be observed in the data-constrained parts of these slices: μ_1 , μ_2 and μ_3 respectively trace well the clusters, filaments and sheets, as we now argue. The μ_1 field is rather homogeneous, apart for small spots where all eigenvalues are largely positive, i.e. undergoing dramatic gravitational collapse along three axes. These correspond to the dynamic clusters. Note that there exists a form of “tidal compensation”: these clusters are surrounded by regions where μ_1 is smaller than its cosmic mean. More patterns can be observed in the μ_2 field: it also exhibits filaments (appearing as dots when piercing the slice). Finally, the μ_3 field is highly-structured, as it also traces sheets (which appear filamentary when sliced). Dynamic voids can also be easily distinguished in this field, wherever μ_3 is negative.

9.3.2 Probabilistic web-type cartography

Building upon previous results and using the procedure described in section 9.2.3, we obtain probabilistic maps of structures. More precisely, we obtain four probability distributions at each spatial position, $\mathcal{P}(T_i(x_k)|d)$, indicating the possibility to encounter a specific structure type (cluster, filament, sheet, void) at that position. As noted in section 9.2.3, these pdfs take their values in the range $[0, 1]$ and sum up to one on a voxel-basis. Figure 9.2 shows slices through their means (see equation (9.3)). The plot shows the anticipated behavior, with a high degree of structure and values close to certainty (i.e. zero or one) in regions covered by data, while the unobserved regions approach a uniform value corresponding to the prior. At this point, it is worth noting that the T-web classifier has a prior preference for some structure types. Using unconstrained large-scale structure

| Structure type | $\mu_{\mathcal{P}(\mathbb{T}_i)}$ | $\sigma_{\mathcal{P}(\mathbb{T}_i)}$ |
|---|-----------------------------------|--------------------------------------|
| Late-time large-scale structure ($a = 1$) | | |
| Void | 0.14261 | 6.1681×10^{-4} |
| Sheet | 0.59561 | 6.3275×10^{-4} |
| Filament | 0.24980 | 5.5637×10^{-4} |
| Cluster | 0.01198 | 5.8793×10^{-5} |

Table 9.1: Prior probabilities assigned by the T-web classifier to the different structures types, in the late-time large-scale structure ($a = 1$).

realizations produced with the same setup,¹ we measured that these prior probabilities, $\mathcal{P}(\mathbb{T}_i)$, can be well described by Gaussians whose mean and standard deviation are given in table 9.1.

In addition to their ensemble mean, the set of samples permits to propagate uncertainty quantification to web-type classification. In particular, it allows us to locally assess the strength of data constraints. In information theory, a convenient way to characterize the uncertainty content of a random source \mathcal{S} is the Shannon entropy (Shannon, 1948), defined by

$$H[\mathcal{S}] \equiv - \sum_i p_i \log_2(p_i), \quad (9.4)$$

where the p_i are the probabilities of possible events. This definition yields expected properties and accounts for the intuition that the more likely an event is, the less information it provides when it occurs (i.e. the more it contributes to the source entropy). We follow this prescription and write the voxel-wise entropy of the web-type posterior, $\mathcal{P}(\mathbb{T}(\vec{x}_k)|d)$, as

$$H[\mathcal{P}(\mathbb{T}(\vec{x}_k)|d)] \equiv - \sum_{i=0}^3 \mathcal{P}(\mathbb{T}_i(\vec{x}_k)|d) \log_2(\mathcal{P}(\mathbb{T}_i(\vec{x}_k)|d)). \quad (9.5)$$

It is a number in the range $[0, 2]$ and its natural unit is the shannon (Sh). $H = 0$ Sh in the case of perfect certainty, i.e. when the data constraints entirely determine the underlying structure type: $\mathcal{P}(\mathbb{T}_{i_0}(\vec{x}_k)|d)$ is 1 for one i_0 and 0 for $i \neq i_0$. H reaches its maximum value of 2 Sh when all $\mathcal{P}(\mathbb{T}_i(\vec{x}_k)|d)$ are equal to $1/4$. This is the case of maximal randomness: all the events being equally likely, no information is gained when one occurs.

A slice through the voxel-wise entropy of the web-type posterior is shown in the left panel of figure 9.3. Generally, the entropy map reflects the information content of the posterior pdf, which comes from augmenting the information content of the prior pdf with the data constraints, in the Bayesian way.

The entropy takes low values and shows a high degree of structure in the regions where data constraints exist, and even reaches zero in some spots where the data are perfectly informative. Comparing with figures 9.1 and 9.2, one can note that this structure is highly non-trivial and does not follow any of the previously described patterns. This is due to the facts that in a Poisson process, the signal (here the density, inferred in section 5.3.1) is correlated with the uncertainty and that structure types classification further is a non-linear function of the density field. In the unobserved regions, the entropy fluctuates around a constant value of about 1.4 Sh, which characterizes the information content of the prior. This value is consistent with the expectation, which can be computed using equation (9.5) (unconditional on the data) and the numbers given in table 9.1.

The information-theoretic quantity that measures the information gain (in shannons) due to the data is the relative entropy or Kullback-Leibler divergence (Kullback & Leibler, 1951) of the posterior from the prior,

$$\begin{aligned} D_{\text{KL}}[\mathcal{P}(\mathbb{T}(\vec{x}_k)|d) \parallel \mathcal{P}(\mathbb{T})] &\equiv \sum_{i=0}^3 \mathcal{P}(\mathbb{T}_i(\vec{x}_k)|d) \log_2 \left(\frac{\mathcal{P}(\mathbb{T}_i(\vec{x}_k)|d)}{\mathcal{P}(\mathbb{T}_i)} \right) \\ &= -H[\mathcal{P}(\mathbb{T}(\vec{x}_k)|d)] - \sum_{i=0}^3 \mathcal{P}(\mathbb{T}_i(\vec{x}_k)|d) \log_2(\mathcal{P}(\mathbb{T}_i)). \end{aligned} \quad (9.6)$$

¹ By this, we specifically mean realizations obtained from initial randomly-generated Gaussian density fields with an Eisenstein & Hu (1998, 1999) power spectrum using the fiducial cosmological parameters of the BORG analysis ($\Omega_m = 0.272$, $\Omega_\mu = 0.728$, $\Omega_b = 0.045$, $h = 0.702$, $\sigma_8 = 0.807$, $n_s = 0.961$, see equation (5.1)). The density field is defined on a $750 \text{ Mpc}/h$ cubic grid of 256^3 -voxels and populated by 512^3 dark matter particles, which are evolved to $z = 69$ with 2LPT and from $z = 69$ to $z = 0$ with COLA, using 30 timesteps logarithmically-spaced in the scale factor. The particles are binned on a 256^3 -voxel grid with the CiC scheme to get the final density field.

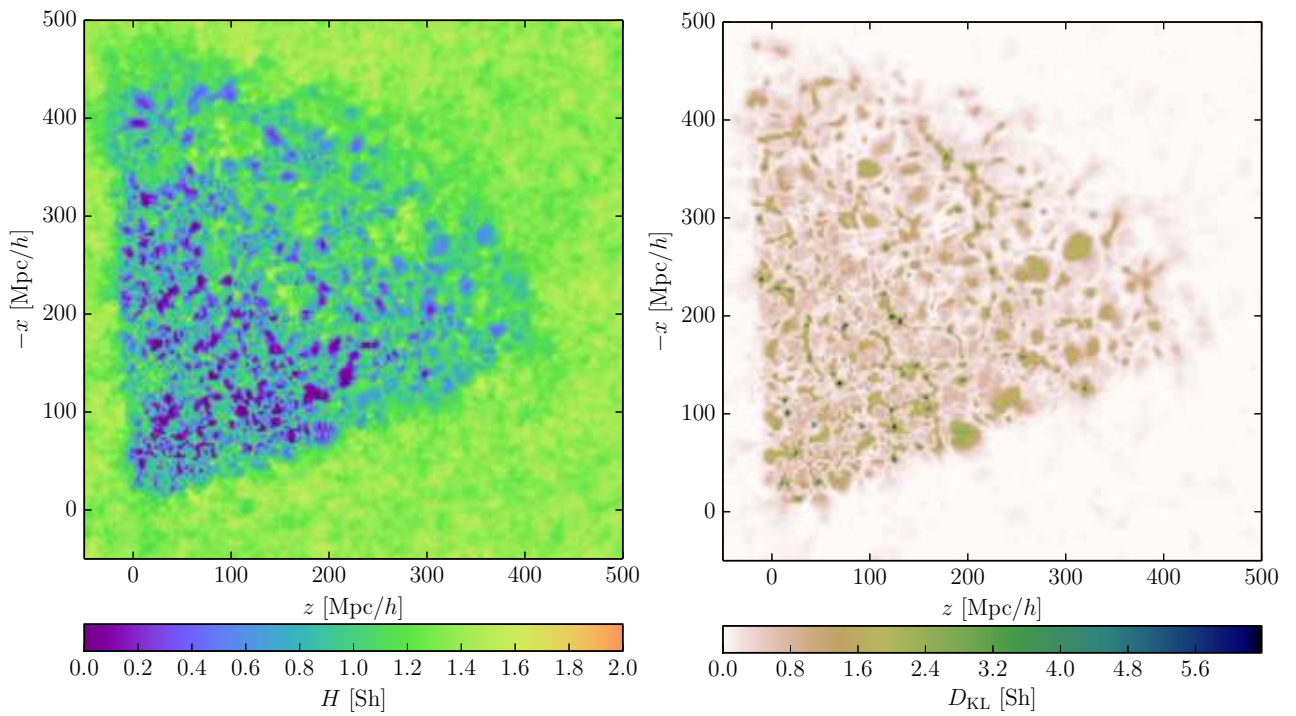


Figure 9.3: Slices through the entropy of the structure types posterior (left panel) and the Kullback-Leibler divergence of the posterior from the prior (right panel), in the final conditions. The entropy H , defined by equation (9.5), quantifies the information content of the posterior pdf represented in figure 9.2, which results from fusing the information content of the prior and the data constraints. The Kullback-Leibler divergence D_{KL} , defined by equation (9.6), represents the information gained in moving from the prior to the posterior. It quantifies the information that has been learned on structure types by looking at SDSS galaxies.

It is a non-symmetric measure of the difference between the two probability distributions.

A slice through the voxel-wise Kullback-Leibler divergence of the web-type posterior from the prior is shown in the right panel of figure 9.3. As expected, the information gain is zero out of the survey boundaries. In the observed regions, SDSS galaxies are informative on underlying structure types at the level of at least ~ 1 Sh. This number can go to ~ 3 Sh in the interior of deep voids and up to ~ 6 Sh in the densest clusters. This map permits to visualize the regions where additional data would be needed to improve structure type classification, e.g. in some high-redshift regions where uncertainty remains due to selection effects.

9.3.3 Volume and mass filling fractions

A characterization of large scale environments commonly found in literature involves evaluating global quantities such as the volume and mass content of these structures. In a particular realization, the volume filling fraction (VFF) for structure type T_i is the number of voxels of type T_i divided by the total number of voxels in the considered volume,

$$\text{VFF}(T_i) \equiv \frac{\sum_{\vec{x}_k} \sum_{j=0}^3 \delta_{\mathbf{K}}^{T_i(\vec{x}_k)T_j^n(\vec{x}_k)}}{N_{\text{vox}}}. \quad (9.7)$$

The mass filling fraction (MFF) can be obtained in a similar manner by weighting the same sum by the local density $\rho(\vec{x}_k) = \bar{\rho}(1 + \delta(\vec{x}_k))$,

$$\text{MFF}(T_i) \equiv \frac{\sum_{\vec{x}_k} \sum_{j=0}^3 (1 + \delta(\vec{x}_k)) \delta_{\mathbf{K}}^{T_i(\vec{x}_k)T_j^n(\vec{x}_k)}}{\sum_{\vec{x}_k} (1 + \delta(\vec{x}_k))}. \quad (9.8)$$

To ensure that results are not prior-dominated, we measured the VFFs and MFFs in the data-constrained parts of our realizations. More precisely, we limited ourselves to the voxels where the survey response operator (representing simultaneously the survey geometry and the selection effects, see sections 4.2 and 5.1) is strictly positive. This amounts to $N_{\text{vox}} = 3,148,504$ out of $256^3 = 16,777,216$ voxels, around 18.7% of the full box (see also section 8.2.3.2 and figure 8.2). In equations (9.7) and (9.8), \vec{x}_k labels one of these voxels.

By measuring the VFF and MFF of different structure types in each constrained realization of our ensemble, we obtained the posterior pdfs, $\mathcal{P}(\text{VFF}(T_i)|d)$ and $\mathcal{P}(\text{MFF}(T_i)|d)$, conditional on the data. Similarly, we computed the prior pdfs, $\mathcal{P}(\text{VFF}(T_i))$ and $\mathcal{P}(\text{MFF}(T_i))$, using unconstrained realizations produced with the same setup. We found that all these pdfs can be well described by Gaussians, the mean and variance of which are given in tables 9.2 and 9.3.

Previous studies on this topic (e.g. Doroshkevich, 1970b; Shen *et al.*, 2006; Hahn *et al.*, 2007a; Forero-Romero *et al.*, 2009; Jasche *et al.*, 2010b; Aragón-Calvo, van de Weygaert & Jones, 2010; Shandarin, Habib & Heitmann, 2012; Cautun *et al.*, 2014) have found a wide range of values for the VFF and MFF of structures (see e.g. table 3 in Cautun *et al.*, 2014). For example, existing studies found that clusters occupy at most a few percent of the volume of the Universe but contribute significantly to the mass content, with a MFF ranging from $\sim 10\%$ (Hahn *et al.*, 2007a; Cautun *et al.*, 2014) to $\sim 40\%$ (Shandarin, Habib & Heitmann, 2012). The void volume fraction can vary from $\sim 10\%$ (Hahn *et al.*, 2007a) to $\sim 80\%$ (Aragón-Calvo, van de Weygaert & Jones, 2010; Shandarin, Habib & Heitmann, 2012; Cautun *et al.*, 2014); in the Forero-Romero *et al.* (2009) formalism (see section C.2.4), it is a very sensitive function of the threshold μ_{th} (figure 9 in Jasche *et al.*, 2010b). These large disparities in the literature arise because different algorithms use various information and criteria for classifying the cosmic web. For this reason, we believe that it is only relevant to make relative statements for the same setup, i.e. to compare our results to the corresponding prior quantities, as done in tables 9.2 and 9.3. In this purpose, the large number of samples used allowed a precise characterization of the pdfs so that all digits quoted in the tables are significant. Note that all our analyses are repeatable for different setups, which allows in principle a comparison with any previous work.

As expected for a Bayesian update of the degree of belief, the posterior quantities generally have smaller variance and a mean value displaced from the prior mean. For the MFF, the posterior means are always within two standard deviations of the corresponding prior means. The analysis shows that in the SDSS, a larger mass fraction is occupied by clusters, sheets, and voids, at the detriment of filaments, in comparison to the prior expectation. The data also favor a smaller filling of the Sloan volume by filaments and sheets and larger filling by voids and clusters. For the cluster VFF, the posterior mean, $\mu_{\text{VFF}(T_3)|d} = 0.01499$ is at about 15 standard deviations ($\sigma_{\text{VFF}(T_3)} = 1.9194 \times 10^{-4}$) of the prior mean, $\mu_{\text{VFF}(T_3)} = 0.01198$. Given other results on the VFF and MFF, we believe that the data truly favor a higher volume content in clusters as compared to the

| Structure type | μ_{VFF} | σ_{VFF} | μ_{VFF} | σ_{VFF} |
|---|--------------------|-------------------------|--------------------|-------------------------|
| Late-time large-scale structure ($a = 1$) | | | | |
| | Posterior | | Prior | |
| Void | 0.14897 | 1.8256×10^{-3} | 0.14254 | 6.2930×10^{-3} |
| Sheet | 0.58914 | 1.3021×10^{-3} | 0.59562 | 2.2375×10^{-3} |
| Filament | 0.24689 | 1.1295×10^{-3} | 0.24986 | 4.4440×10^{-3} |
| Cluster | 0.01499 | 8.7274×10^{-5} | 0.01198 | 1.9194×10^{-4} |

Table 9.2: Mean and standard deviation of the prior and posterior pdfs for the volume filling fraction of different structure types in the late-time large-scale structure ($a = 1$).

| Structure type | μ_{MFF} | σ_{MFF} | μ_{MFF} | σ_{MFF} |
|---|--------------------|-------------------------|--------------------|-------------------------|
| Late-time large-scale structure ($a = 1$) | | | | |
| | Posterior | | Prior | |
| Void | 0.04050 | 8.3531×10^{-4} | 0.03876 | 2.3352×10^{-3} |
| Sheet | 0.35605 | 1.2723×10^{-3} | 0.35286 | 3.6854×10^{-3} |
| Filament | 0.47356 | 1.5661×10^{-3} | 0.48170 | 4.2215×10^{-3} |
| Cluster | 0.12990 | 6.4966×10^{-4} | 0.12666 | 1.8284×10^{-3} |

Table 9.3: Mean and standard deviation of the prior and posterior pdfs for the mass filling fraction of different structure types in the late-time large-scale structure ($a = 1$).

structure formation model used as prior. However, this surprising result should be treated with care; part of the discrepancy is likely due to the original BORG analysis, which optimizes the initial conditions for evolution with 2LPT (instead of the non-linear evolution with COLA used for the present work). LPT predicts fuzzier halos than N -body dynamics, which results in the incorrect prediction of a high cluster VFF (see section 2.3; Leclercq *et al.*, 2013).

9.4 The primordial large-scale structure

In this section, we discuss the results of our analysis of the initial density field, at $a = 10^{-3}$. Since the analysis of the primordial large-scale structure does not involve an additional filtering step, we have been able to keep a larger number of samples of the posterior pdf for initial conditions, obtained in chapter 5. Hence, for all results described in this section, we used a total of 4,473 samples.

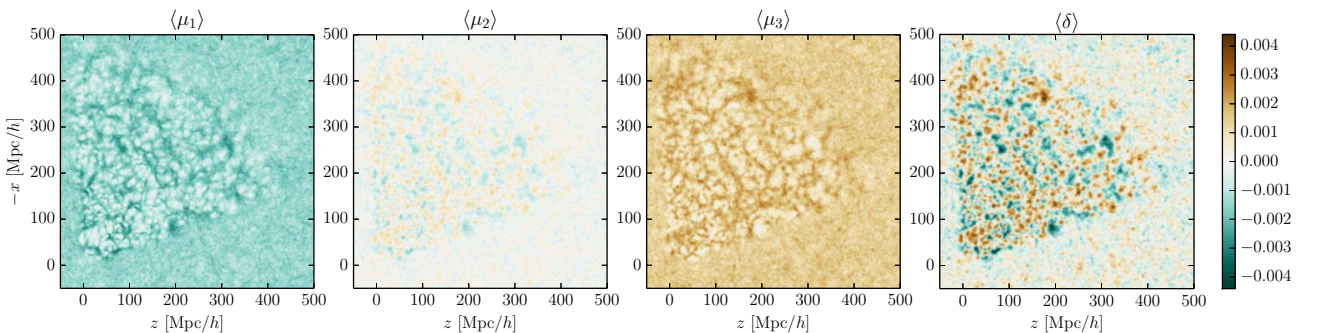


Figure 9.4: Slices through the three-dimensional ensemble posterior mean for the eigenvalues $\mu_1 \leq \mu_2 \leq \mu_3$ of the tidal field tensor in the initial conditions, estimated from 4,473 samples. The rightmost panel shows the corresponding slice through the posterior mean for the initial density contrast $\delta = \mu_1 + \mu_2 + \mu_3$, obtained in section 5.3.1.

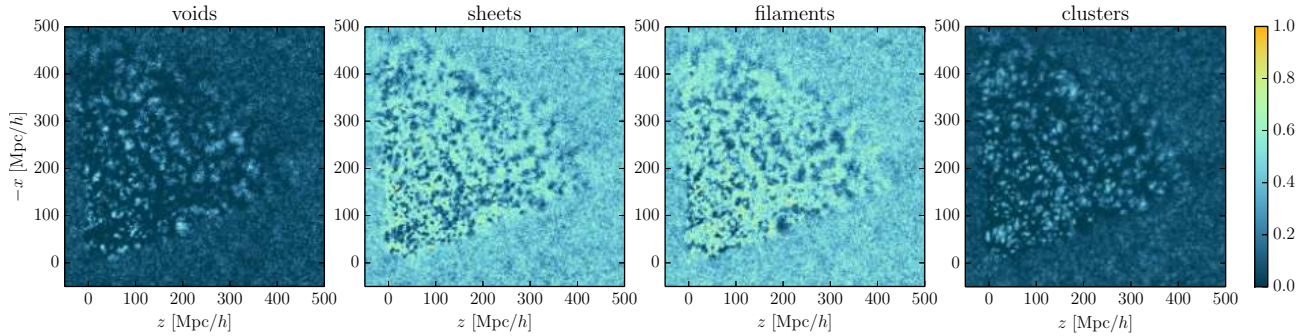


Figure 9.5: Slices through the posterior mean for different structure types (from left to right: void, sheet, filament, and cluster) in the primordial large-scale structure in the Sloan volume ($a = 10^{-3}$). These four three-dimensional voxel-wise pdfs sum up to one on a voxel basis.

9.4.1 Tidal environment

In a similar fashion as in section 9.3.1, the application of the T-web classifier to initial density samples yields the posterior pdf for the three eigenvalues, μ_1 , μ_2 and μ_3 , of the initial tidal field tensor. Figure 9.4 shows slices through their means. For visual comparison, the rightmost panel shows the corresponding slice through the posterior mean of the initial density contrast, $\delta = \mu_1 + \mu_2 + \mu_3$, obtained in section 5.3.1.

In a Gaussian random field, μ_1 is generally negative, μ_3 is generally positive and μ_2 close to zero (see the unobserved parts of the slices in figure 9.4). In addition, μ_2 closely resembles the total density contrast δ up to a global scaling. In the constrained regions, the eigenvalues of the initial tidal tensor follow this behavior. The structure observed in their maps is visually consistent with the decomposition of Gaussian density fluctuations as shown by the right panel.

9.4.2 Probabilistic web-type cartography

Looking at the sign of the eigenvalues of the initial tidal tensor and following the procedure described in section 9.2.3, we obtain a probabilistic cartography of the primordial large-scale structure. As before, we obtain four voxel-wise pdfs $\mathcal{P}(T_i(\vec{x}_k)|d)$, taking their values in the range $[0, 1]$ and summing up to one. Figure 9.5 shows slices through their means, defined by equation (9.3). As in the final conditions, the maps exhibit structure in the data-constrained regions and approach uniform values in the unobserved parts, corresponding to the respective priors. Using unconstrained realizations of Gaussian random fields produced with the same setup,² we measured these prior probabilities. Their means and standard deviations are given in table 9.4.

At this point, it is worth mentioning that there exists an additional symmetry for Gaussian random fields. Since the definition of the tidal tensor is linear in the density contrast (see equations (C.6) and (C.7)) and since positive and negative density contrasts are equally likely, a positive and negative value for a given μ_i have the same probabilities. Because of this sign symmetry, the pdfs for voids and clusters (0 or 3 positive/negative eigenvalues) and the pdfs for sheets and filaments (1 or 2 positive/negative eigenvalues) are equal. This can be checked both in table 9.4 and in the unconstrained regions of the maps in figure 9.5. In the constrained regions, a qualitative complementarity between pdfs for voids and clusters and for sheets and filaments can be observed. This can be explained by the following. As $\sum_i \mathcal{P}(T_i(\vec{x}_k)|d) = 1$ and assuming that $\mathcal{P}(T_i(\vec{x}_k)|d) \approx \mathcal{P}(T_{3-i}(\vec{x}_k)|d)$ for unlikely events, consistently with the previous remark, we get $\mathcal{P}(T_0(\vec{x}_k)|d) \approx 1 - \mathcal{P}(T_3(\vec{x}_k)|d)$ wherever $\mathcal{P}(T_1(\vec{x}_k)|d) \approx \mathcal{P}(T_2(\vec{x}_k)|d)$ is sufficiently small; and $\mathcal{P}(T_1(\vec{x}_k)|d) \approx 1 - \mathcal{P}(T_2(\vec{x}_k)|d)$ wherever $\mathcal{P}(T_0(\vec{x}_k)|d) \approx \mathcal{P}(T_3(\vec{x}_k)|d)$ is sufficiently small. These results are therefore consistent with expectations based on Gaussianity for the primordial large-scale structure in the Sloan volume.

In a similar fashion as in section 9.3.2, the ensemble of samples permits us to propagate uncertainties to structure type classification and to characterize the strength of data constraints. In the left panel of figure 9.6, we show a slice through the voxel-wise entropy of the web-type posterior pdf in the initial conditions, defined by equation (9.5). This function quantifies the information content of the posterior, which comes from both the prior and the data constraints. As in the final conditions, the entropy takes lower values inside the survey

² We used the initial conditions of our set of unconstrained simulations (see footnote 1).

| Structure type | $\mu_{\mathcal{P}(T_i)}$ | $\sigma_{\mathcal{P}(T_i)}$ |
|--|--------------------------|-----------------------------|
| Primordial large-scale structure ($a = 10^{-3}$) | | |
| Void | 0.07979 | 5.4875×10^{-5} |
| Sheet | 0.42022 | 1.0240×10^{-4} |
| Filament | 0.42022 | 1.0412×10^{-4} |
| Cluster | 0.07978 | 5.6337×10^{-5} |

Table 9.4: Prior probabilities assigned by the T-web classifier to the different structures types, in the primordial large-scale structure ($a = 10^{-3}$).

| Structure type | μ_{VFF} | σ_{VFF} | μ_{VFF} | σ_{VFF} |
|--|--------------------|-------------------------|--------------------|-------------------------|
| Primordial large-scale structure ($a = 10^{-3}$) | | | | |
| | Posterior | | Prior | |
| Void | 0.07994 | 4.0221×10^{-4} | 0.07977 | 1.0200×10^{-3} |
| Sheet | 0.41994 | 6.1770×10^{-4} | 0.42019 | 1.7885×10^{-3} |
| Filament | 0.42048 | 6.3589×10^{-4} | 0.42024 | 1.7820×10^{-3} |
| Cluster | 0.07964 | 3.8043×10^{-4} | 0.07980 | 1.0260×10^{-3} |

Table 9.5: Mean and standard deviation of the prior and posterior pdfs for the volume filling fraction of different structure types in the primordial large-scale structure ($a = 10^{-3}$).

region. In the unobserved parts, the entropy fluctuates around 1.6 Sh, value which characterizes the information content of the prior. Using equation (9.5) (unconditional on the data) and the numbers given in table 9.4, one can check that this number is consistent with the expectation. In the right panel of figure 9.6, we show a map of the Kullback-Leibler divergence of the posterior from the prior, which represents the information gain due to the data.

9.4.3 Volume and mass filling fractions

We computed the volume and mass filling fractions (defined by equations (9.7) and (9.8)) of different structure types in the primordial large-scale structure in the Sloan volume. As for the final conditions, we kept only the regions where the survey response operator is strictly positive. Consequently, we obtained the posterior pdfs $\mathcal{P}(\text{VFF}(T_i)|d)$ and $\mathcal{P}(\text{MFF}(T_i)|d)$. Using a set of unconstrained Gaussian random fields, we also measured $\mathcal{P}(\text{VFF}(T_i))$ and $\mathcal{P}(\text{MFF}(T_i))$ and found that all these pdfs are well described by Gaussians, the means and standard deviations of which are given in table 9.5 and 9.6.

All posterior quantities obtained are within two standard deviations of the corresponding prior means, and show smaller variance, as expected. Hence, all results obtained are consistent with Gaussian initial conditions.

9.5 Evolution of the cosmic web

In addition to the inference of initial and final density fields, BORG allows to simultaneously analyze the formation history and morphology of the observed large-scale structure, a subject that we refer to as chrono-

| Structure type | μ_{MFF} | σ_{MFF} | μ_{MFF} | σ_{MFF} |
|--|--------------------|-------------------------|--------------------|-------------------------|
| Primordial large-scale structure ($a = 10^{-3}$) | | | | |
| | Posterior | | Prior | |
| Void | 0.07958 | 4.0122×10^{-4} | 0.07941 | 1.0163×10^{-3} |
| Sheet | 0.41933 | 6.1907×10^{-4} | 0.41957 | 1.7912×10^{-3} |
| Filament | 0.42110 | 6.3543×10^{-4} | 0.42087 | 1.7785×10^{-3} |
| Cluster | 0.07999 | 3.8206×10^{-4} | 0.08015 | 1.0293×10^{-3} |

Table 9.6: Mean and standard deviation of the prior and posterior pdfs for the mass filling fraction of different structure types in the primordial large-scale structure ($a = 10^{-3}$).

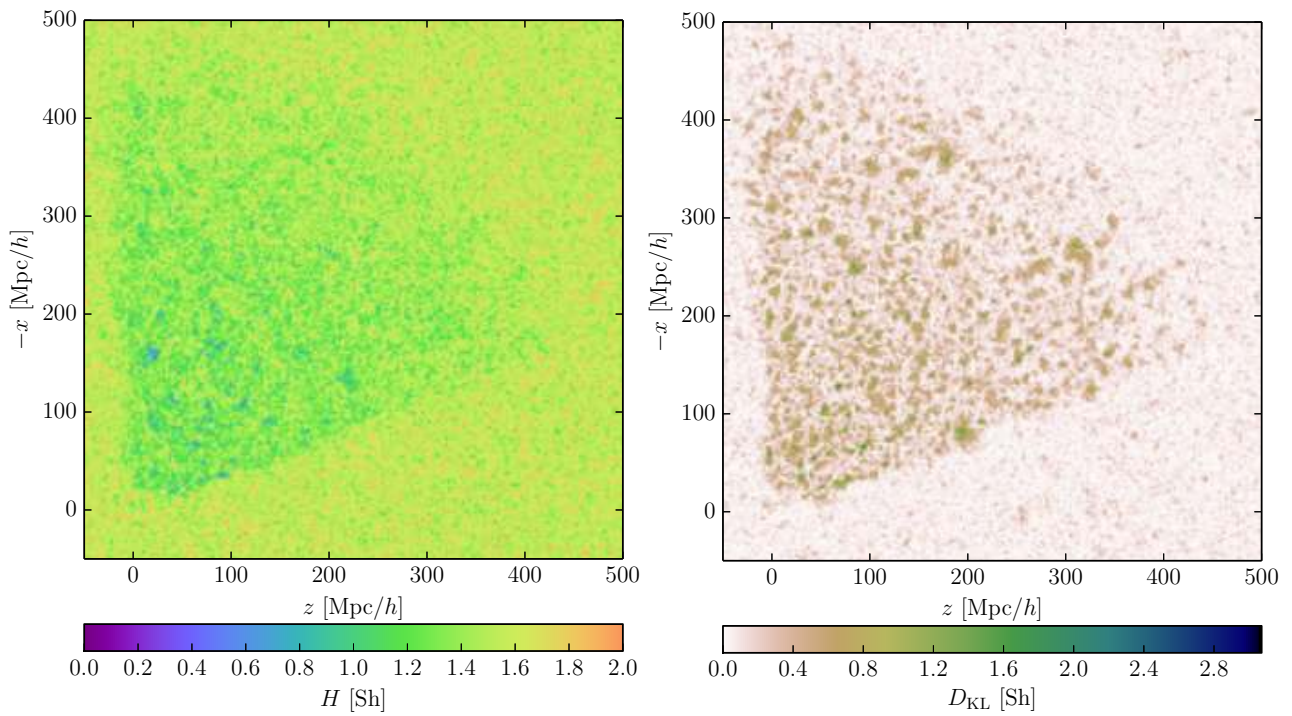


Figure 9.6: Slices through the entropy of the structure types posterior (left panel) and the Kullback-Leibler divergence of the posterior from the prior (right panel), in the initial conditions. The entropy H , defined by equation (9.5), quantifies the information content of the posterior pdf represented in figure 9.5, which results from fusing the information content of the prior and the data constraints. The Kullback-Leibler divergence D_{KL} , defined by equation (9.6), represents the information gained in moving from the prior to the posterior. It quantifies the information that has been learned on structure types by looking at SDSS galaxies.

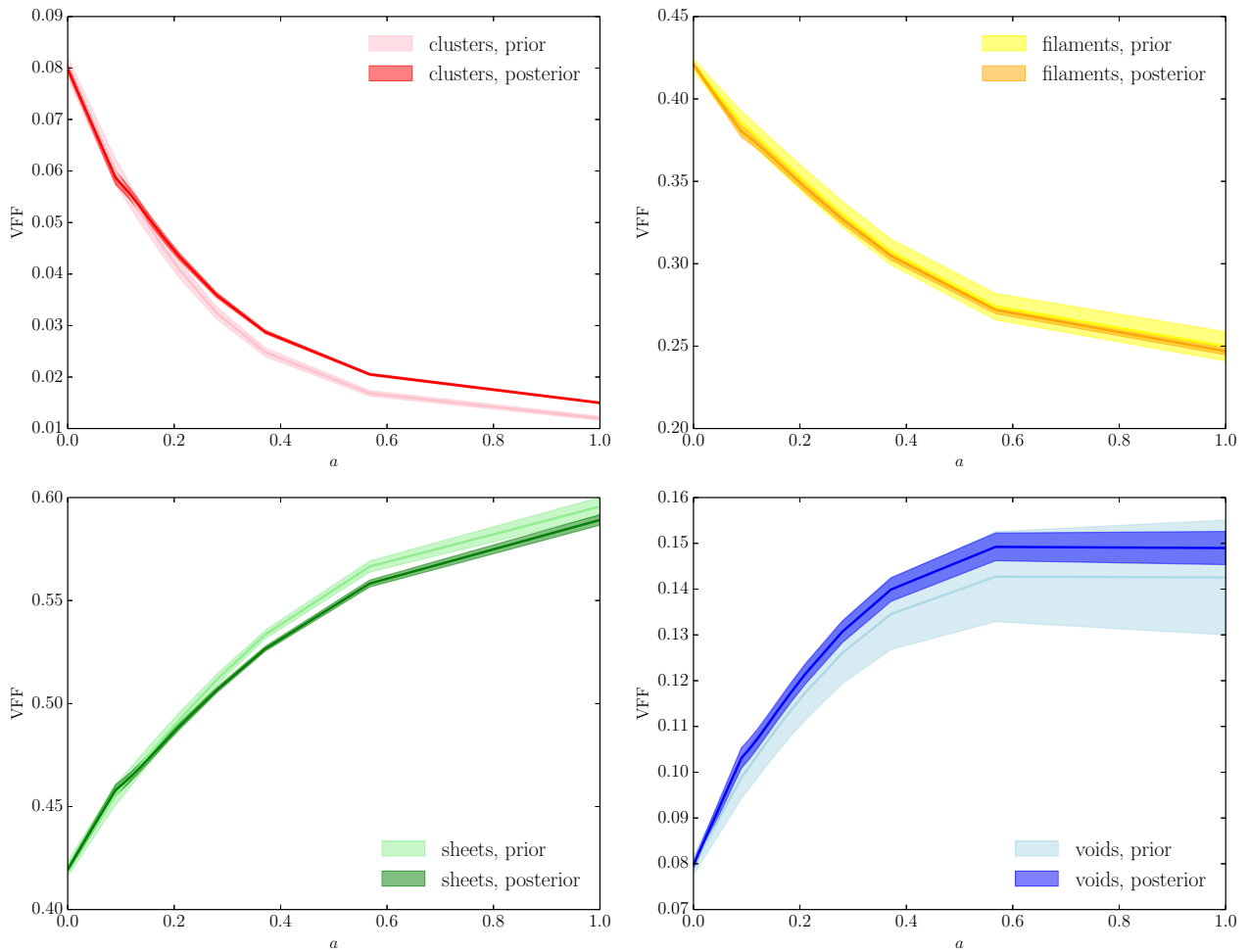


Figure 9.7: Time evolution of the volume filling fractions of different structure types (from left to right and top to bottom: clusters, filaments, sheets, voids). The solid lines show the pdf means and the shaded regions are the $2\text{-}\sigma$ credible intervals. Light colors are used for the priors and dark colors for the posteriors.

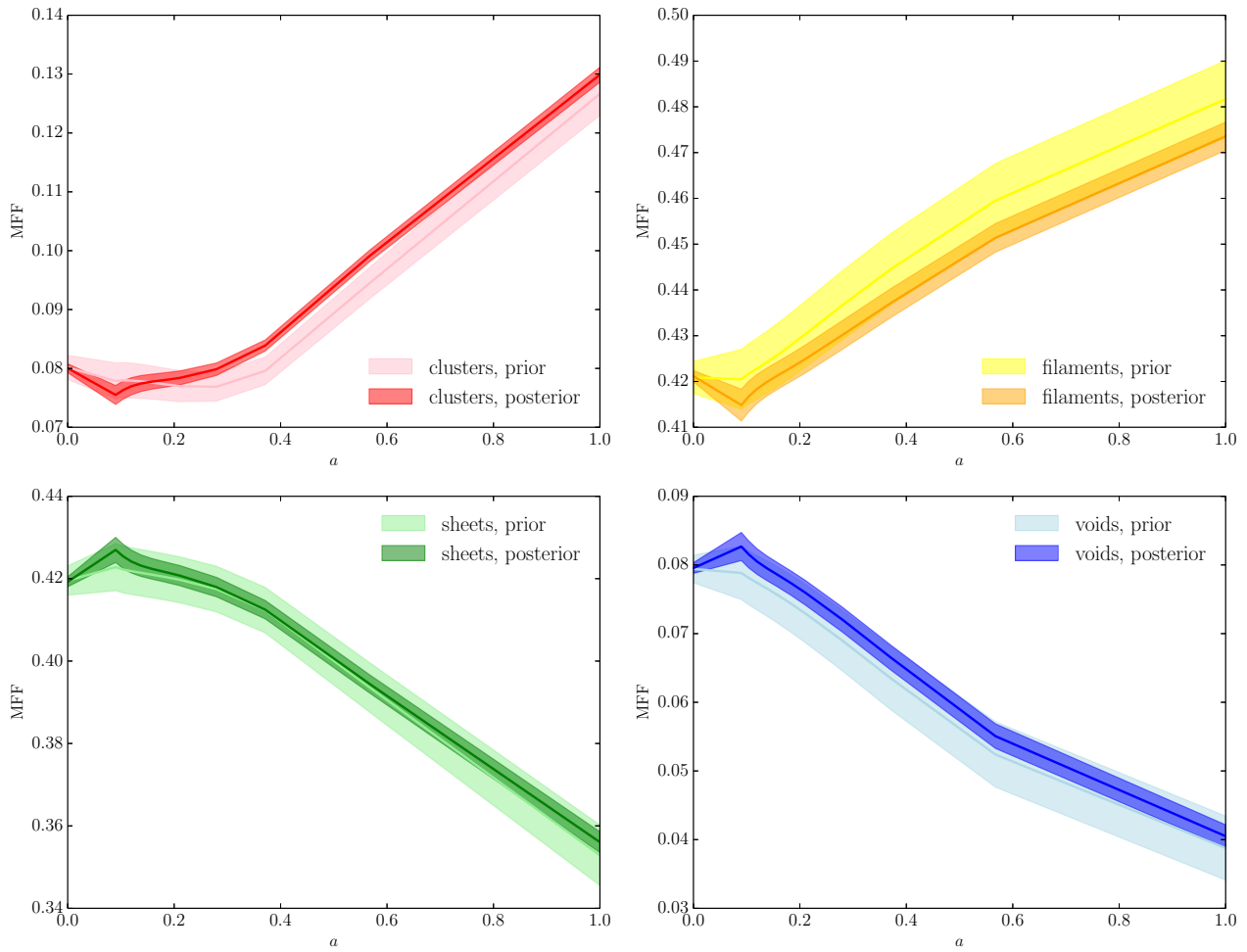


Figure 9.8: Same as figure 9.7 but for the mass filling fractions.

cosmography. In this section, we discuss the evolution of the cosmic web from its origin ($a = 10^{-3}$, analyzed in section 9.4) to the present epoch ($a = 1$, analyzed in section 9.3). To do so, we use 11 snapshots saved during the COLA filtering of our results (see section 9.2.2). These are linearly separated in redshift from $z = 10$ to $z = 0$. We perform this analysis in the 1,097 samples filtered with COLA considered in section 9.3. For each of these samples and for each redshift, we follow the procedure described in sections 9.2.2 and 9.2.3 to compute the density field and to classify the structure types.

9.5.1 Evolution of the probabilistic maps

We followed the time evolution of the probabilistic web-type maps from the primordial (figure 9.5) to the late-time large-scale structure (figure 9.2). In unconstrained regions, these maps show the evolution of the prior preference for specific structure types (see tables 9.1 and 9.4), in particular the breaking of the initial symmetry between voids and clusters and between sheets and filaments, discussed in section 9.4.2.

In data-constrained regions, the time evolution of web-type maps permits to visually check the expansion history of individual regions where the posterior probability of one specific structure is high. In particular, it is easy to see that, as expected from their dynamical definition, voids expand and clusters shrink in comoving coordinates, from $a = 10^{-3}$ to $a = 1$ (the reader is invited to compare the leftmost and rightmost panels of figures 9.2 and 9.5). Similarly, regions corresponding with high probability to sheets and filaments expand along two and one axis, respectively, and shrink along the others. This phenomenon is more difficult to see in slices, however, as the slicing plane intersects randomly the eigendirections of the tidal tensor.

The time evolution of maps of the web-type posterior entropy (absolute and relative to the prior) also exhibit some interesting features. There, it is possible to simultaneously check the increase of the information content of the prior (from $H \approx 1.6$ Sh to $H \approx 1.4$ Sh) and the displacement of observational information operated by the physical model. As the large-scale structure forms in the Sloan volume, data constraints are propagated and the complex structure of the final entropy map (figure 9.3), discussed in section 9.3.2, takes shape.

9.5.2 Volume filling fraction

Our ensemble of snapshots allows us to check the time evolution of global characterizations of the large-scale structure such as the volume and mass filling fractions of different structures. As in sections 9.3.3 and 9.4.3, we computed these quantities using only the volume where the survey response operator is non-zero. In figure 9.7, we plot these VFF as a function of the scale factor. There, the solid lines correspond to the pdf means and the shaded regions to the $2\text{-}\sigma$ credible intervals, with light colors for the priors and dark colors for the posteriors.

The time variation of the VFF in figure 9.7 is consistent with the expected dynamical behavior of structures. As voids and sheets expand along three and two axes, respectively, their volume fraction increases. Here, the posterior probabilities are mild updates of this prediction. Conversely, as clusters and filaments shrink along three and two axes, respectively, their volume fraction decreases. An explanation for the substantial displacement of the posterior from the prior, observed for clusters, can be found in section 9.3.3.

As already noted, the VFF is a very sensitive function of the precise definition of structures, grid size, density assignment scheme, smoothing scale, etc. For this reason, even for prior probabilities, our results can be in qualitative disagreement with previous authors (e.g. figure 23 in Cautun *et al.*, 2014), due to their very different definitions of structures. Therefore, we only found relevant to compare our posterior results with the prior predictions based on unconstrained realizations. The same remark applies to the MFF in the following section.

9.5.3 Mass filling fraction

In figure 9.8, we show the time evolution of the mass filling fractions using the same plotting conventions. Results are consistent with an interpretation based on large scale flows of matter. According to this picture, voids always loose mass while clusters always become more massive. The behavior of sheets and filaments can in principle be more complex, since these regions have both inflows and outflows of matter depending on the detail of their expansion profiles. In our setup, we found that the number of axes along which there is expansion dominates in the determination of the balance of inflow versus outflow, for global quantities such as the MFF. Therefore, filaments always gain mass and sheets always loose mass. Summing up our prior predictions, as they expand along at least two axes, matter flows out of voids and sheets and streams towards filaments and clusters.

The posterior probabilities slightly update this picture. Observations support smaller outflowing of matter from voids. For structures globally gaining matter, the priors are displaced towards less massive filaments and more massive clusters. All posterior predictions fall within the $\sim 2\text{-}\sigma$ credible interval from corresponding prior means.

9.6 Summary and Conclusion

Along with chapter 8 (Leclercq *et al.*, 2015), this work exploits the high quality of inference results produced by the application of the Bayesian code BORG (chapter 4; Jasche & Wandelt, 2013a) to the Sloan Digital Sky Survey main galaxy sample (chapter 5; Jasche, Leclercq & Wandelt, 2015). We presented a Bayesian cosmic web analysis of the nearby Universe probed by the northern cap of the SDSS and its surrounding. In doing so, we produced the first probabilistic, four-dimensional maps of dynamic structure types in real observations.

As described in section 9.2.1, our method relies on the physical inference of the initial density field in the LSS (Jasche & Wandelt, 2013a; Jasche, Leclercq & Wandelt, 2015). Starting from these, we generated a large set of data-constrained realizations using the fast COLA method (section 9.2.2). The use of 2LPT as a physical model in the inference process and of the fully non-linear gravitational dynamics, provided by COLA, as a filter allowed us to describe structures at the required statistical accuracy, by very well representing the full hierarchy of correlation functions. Even though initial conditions were inferred with the approximate 2LPT model, we checked that the clustering statistics of constrained non-linear model evaluations agree with theoretical expectations up to scales considered in this work. As described in section 9.2.3, we used the dynamic web-type classification algorithm proposed by Hahn *et al.* (2007a) to dissect the cosmic web into voids, sheets, filaments, and clusters.

In sections 9.3 and 9.4, we presented the resulting maps of structures in the final and initial conditions, respectively, and studied the distribution of global quantities such as volume fraction and mass filling fractions. In section 9.5, we further analyzed the time evolution of our results, in a rigorous chrono-cosmographic framework.

For all results presented in this chapter, we demonstrated a thorough capability of uncertainty quantification. Specifically, for all inferred maps and derived quantities, we got a probabilistic answer in terms of a prior and a posterior distribution. The variation between samples of the posterior distribution quantifies the remaining uncertainties of various origins (in particular noise, selection effects, survey geometry and galaxy bias, see chapters 4 and 5 for a detailed discussion). Building upon our accurate probabilistic treatment, we looked at the entropy of the structure type posterior and at the relative entropy between posterior and prior. In doing so, we quantified the information gain due to SDSS galaxy data with respect to the underlying dynamic cosmic web and analyzed how this information is propagated during cosmic history. This study constitutes the first link between cosmology and information theory using real data.

In summary, our methodology yields an accurate cosmographic description of web types in the non-linear regime of structure formation, permits to analyze their time evolution and allows a precise uncertainty quantification in a full-scale Bayesian framework. These inference results can be used for a rich variety of applications, ranging from studying galaxies inside their environment to cross-correlating with other cosmological probes. They count among the first steps towards accurate chrono-cosmography, the subject of simultaneously analyzing the morphology and formation history of the inhomogeneous Universe.

Note added: As we were finalizing the paper corresponding to this chapter (Leclercq, Jasche & Wandelt, 2015c) for submission, the works by Zhao *et al.* (2015) and Shi, Wang & Mo (2015) appeared where the relationship between halos and the cosmic web environment defined by the tidal tensor is being studied.