Mon. Not. R. Astron. Soc. **380,** 551–570 (2007) doi:10.1111/j.1365-2966.2007.12125.x

A cosmic watershed: the WVF void detection technique

Erwin Platen,* Rien van de Weygaert and Bernard J. T. Jones
Kapteyn Astronomical Institute, University of Groningen, PO Box 800, 9700 AV Groningen, the Netherlands

Accepted 2007 June 15. Received 2007 June 11; in original form 2007 March 27

ABSTRACT

On megaparsec scales the Universe is permeated by an intricate filigree of clusters, filaments, sheets and voids, the cosmic web. For the understanding of its dynamical and hierarchical history it is crucial to identify objectively its complex morphological components. One of the most characteristic aspects is that of the dominant underdense voids, the product of a hierarchical process driven by the collapse of minor voids in addition to the merging of large ones.

In this study we present an objective void finder technique which involves a minimum of assumptions about the scale, structure and shape of voids. Our void finding method, the watershed void finder (WVF), is based upon the watershed transform, a well-known technique for the segmentation of images. Importantly, the technique has the potential to trace the existing manifestations of a void hierarchy. The basic watershed transform is augmented by a variety of correction procedures to remove spurious structure resulting from sampling noise.

This study contains a detailed description of the WVF. We demonstrate how it is able to trace and identify, relatively parameter free, voids and their surrounding (filamentary and planar) boundaries. We test the technique on a set of kinematic Voronoi models, heuristic spatial models for a cellular distribution of matter. Comparison of the WVF segmentations of low-noise and high-noise Voronoi models with the quantitatively known spatial characteristics of the intrinsic Voronoi tessellation shows that the size and shape of the voids are successfully retrieved. WVF manages to even reproduce the full void size distribution function.

Key words: methods: data analysis – methods: numerical – cosmology: theory – large-scale structure of Universe.

1 INTRODUCTION

Voids form a prominent aspect of the distribution of galaxies and matter on megaparsec scales. They are enormous regions with sizes in the range of $20 - 50\,h^{-1}$ Mpc that are practically devoid of any galaxy and usually roundish in shape. Forming an essential ingredient of the cosmic web (Bond, Kofman & Pogosyan 1996), they are surrounded by elongated filaments, sheet-like walls and dense compact clusters. Together they define the salient web-like pattern of galaxies and matter which pervades the observable Universe.

Voids have been known as a feature of galaxy surveys since the first surveys were compiled (Chincarini & Rood 1975; Gregory & Thompson 1978; Einasto, Joeveer & Saar 1980). Following the discovery by Kirshner et al. (1981, 1987) of the most dramatic specimen, the Boötes void, a hint of their central position within a web-like arrangement came with the first CfA (Center for Astrophysics) redshift slice (de Lapparent, Geller & Huchra 1986). This view has recently been expanded dramatically as maps of the spatial distribution of hundreds of thousands of galaxies in the 2dF Galaxy Redshift Survey (2dFGRS; Colless et al. 2003) and Sloan Digital Sky Survey (SDSS; Abazajian et al. 2003) have become available.

Voids are a manifestation of the cosmic structure formation process as it reaches a non-linear stage of evolution. Structure forms by gravitational instability from a primordial Gaussian field of small amplitude density perturbations, where voids emerge out of the depressions (e.g. Icke 1984; van de Weygaert & van Kampen 1993). They mark the transition scale at which perturbations have decoupled from the Hubble flow and organized themselves into recognizable structural features. Early theoretical models of void formation (Hoffman & Shaham 1982; Icke 1984; Bertschinger 1985; Blumenthal et al. 1992) were followed and generalized by the first numerical simulations of void centred universes (Martel & Wassermann 1990; Regős & Geller 1991; Dubinski et al. 1993; van de Weygaert & van Kampen 1993).

In recent years the huge increase in computational resources has enabled *N*-body simulations to resolve in detail the intricate substructure of voids within the context of hierarchical cosmological structure formation scenarios (Arbabi-Bidgoli & Müller 2002; Mathis & White 2002; Gottlöber et al. 2003; Goldberg & Vogeley

*E-mail: platen@astro.rug.nl

2004; Colberg et al. 2005; Padilla, Ceccarelli & Lambas 2005; Hoeft et al. 2007). They confirm the theoretical expectation of voids having a rich substructure as a result of their hierarchical buildup. Theoretically this evolution has been successfully embedded in the extended Press–Schechter description (Press & Schechter 1974; Bond et al. 1991; Sheth 1998). Sheth & van de Weygaert (2004) showed how this can be described by a two-barrier excursion set formalism (also see Furlanetto & Piran 2006). The two barriers refer to the two processes dictating the evolution of voids: their merging into ever larger voids as well as the collapse and disappearance of small ones embedded in overdense regions (see van de Weygaert, Sheth & Platen 2004).

Besides representing a key constituent of the cosmic matter distribution voids are interesting and important for a variety of reasons. First, they are a prominent feature of the megaparsec Universe. A proper and full understanding of the formation and dynamics of the cosmic web is not possible without understanding the structure and evolution of voids (Sheth & van de Weygaert 2004). Secondly, they are a probe of cosmological parameters. The outflow from the voids depends on the matter density parameter $\Omega_{\rm m}$, the Hubble parameter $H(t)$ and possibly on the cosmological constant Λ (see e.g. Martel & Wassermann 1990; Dekel & Rees 1994; Bernardeau & van de Weygaert 1996; Fliche & Triay 2006). These parameters also dictate their redshift space distortions (Ryden & Melott 1996; Schmidt, Ryden & Melott 2001) while their intrinsic structure and shape is sensitive to various aspects of the power spectrum of density fluctuations (Lee & Park 2006). A third point of interest concerns the galaxies in voids. Voids provide a unique and still largely pristine environment for studying the evolution of galaxies (Hoffman, Silk & Wyse 1992; Little & Weinberg 1994; Peebles 2001). The recent interest in environmental influences on galaxy formation has stimulated substantial activity in this direction (Benson et al. 1996; Szomoru et al. 1998; Grogin & Geller 1999; Friedmann & Piran 2001; Hoyle & Vogeley 2002; Mathis & White 2002; Gottlöber et al. 2003; Rojas et al. 2005; Ceccarelli et al. 2006; Furlanetto & Piran 2006; Patiri et al. 2006b; Hoeft et al. 2007).

Despite the considerable interest in voids a fairly basic yet highly significant issue remains: identifying voids and tracing their outline within the complex spatial geometry of the cosmic web. There is not an unequivocal definition of what a void is and as a result there is considerable disagreement on the precise outline of such a region (see e.g. Shandarin et al. 2006). Because of the vague and diverse definitions, and the diverse interests in voids, there is a plethora of void identification procedures (Kauffmann & Fairall 1991; El-Ad & Piran 1997; Aikio & Mähönen 1998; Arbabi-Bidgoli & Müller 2002; Hoyle & Vogeley 2002; Plionis & Basilakos 2002; Colberg et al. 2005; Patiri et al. 2006a; Shandarin et al. 2006; Hahn et al. 2007; Neyrinck, in preparation).

The 'sphere-based' void finder algorithm of El-Ad & Piran (1997) has been at the basis of most void finding methods. However, this successful approach will not be able to analyse complex spatial configurations in which voids may have arbitrary shapes and contain a range and variety of substructures. A somewhat related and tessellation-based void finding technique that still is under development is ZOBOV (ZOnes Bordering On Voidiness, Neyrinck, in preparation). It is the void finder equivalent to the VOBOZ (Voronoi bound zones) halo finder method (Neyrinck, Gnedin & Hamilton 2005).

Here we introduce and test a new and objective void finding formalism that has been specifically designed to dissect the multiscale character of the void network and the web-like features marking its boundaries. Our watershed void finder (WVF) is based on the watershed algorithm (Beucher & Lantuejoul 1979; Beucher & Meyer 1993). It stems from the field of mathematical morphology (MM) and image analysis.

The WVF is defined with respect to the Delaunay tessellation field estimator (DTFE) density field of a discrete point distribution (Schaap & van de Weygaert 2000). This assures an optimal sensitivity to the morphology of spatial structures and yields an unbiased probe of substructure in the mass distribution (see e.g. Okabe et al. 2000; Schaap & van de Weygaert 2000). Because the WVF void finder does not impose a priori constraints on the size, morphology and shape of a void it provides a basis for analysing the intricacies of an evolving void hierarchy. Indeed, this has been a major incentive towards its development.

This study is the first in a series. Here we will define and describe the WVF and investigate its performance with respect to a test model of spatial web-like distributions, Voronoi kinematic models. Having assured the success of WVF to trace and measure the spatial characteristics of these models the follow-up study will address the application of WVF on a number of GIF *N*-body simulations of structure formation (Kauffmann et al. 1999). Amongst others, WVF will be directed towards characterizing the hierarchical structure of the megaparsec void population (Sheth & van de Weygaert 2004). For a comparison of the WVF with other void finder methods we refer to the extensive study of Colberg et al. (in preparation).

In the following sections we will first describe how the fundamental concepts of MM have been translated into a tool for the analysis of cosmological density fields inferred from a discrete *N*-body simulation or galaxy redshift survey point distribution (Sections 2 and 3). To test our method we have applied it to a set of heuristic and flexible models of a cellular spatial distribution of points, Voronoi clustering models. These are described in Section 4. In Section 5 we present the quantitative analysis of our test results and a comparison with the known intrinsic properties of the test models. In Section 6 we evaluate our findings and discuss the prospects for the analysis of cosmological *N*-body simulations.

2 THE WATERSHED VOID FINDER

The new void finding algorithm which we introduce here is based on the watershed transform (WST) of Beucher & Lantuejoul (1979) and Beucher & Meyer (1993). A more extensive and technical description of the basic concepts of MM and the basic watershed algorithm in terms of homotopy transformations on lattices (Kresch 1998) is provided in Appendices A and B.

2.1 The watershed transform

The WST is used for segmenting images into distinct regions and objects. The WST is a concept defined within the context of MM, and was first introduced by Beucher & Lantuejoul (1979). The basic idea behind the WST finds its origin in geophysics. The WST delineates the boundaries of the separate domains, i.e. the basins, into which yields of, for example, rainfall will collect.

The word watershed refers to the analogy of a landscape being flooded by a rising level of water. Suppose we have a surface in the shape of a landscape (first image of Fig. 1). The surface is pierced at the location of each of the minima. As the water level rises a growing fraction of the landscape will be flooded by the water in the expanding basins. Ultimately basins will meet at the ridges corresponding to saddle points in the density field. This intermediate step is plotted in the second image of Fig. 1. The ridges define the boundaries of the basins, enforced by means of a sufficiently high

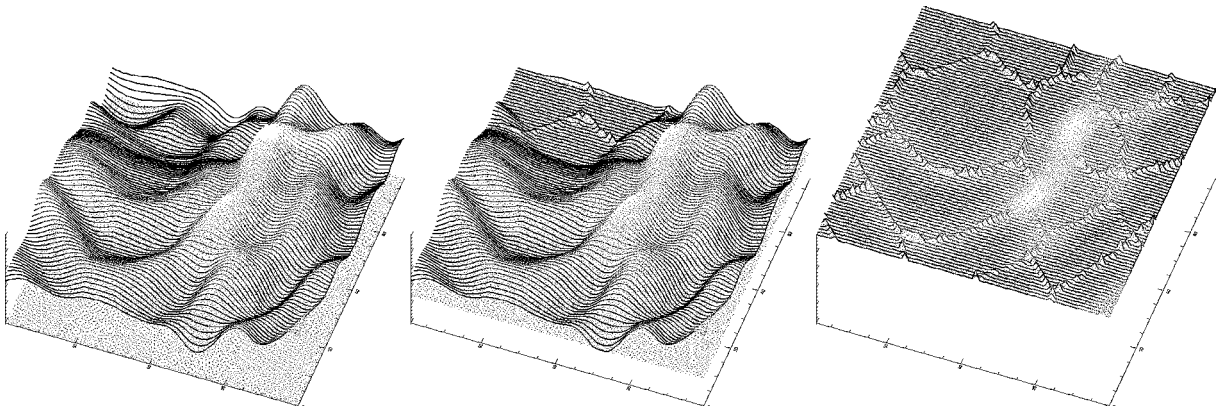

Figure 1. Three frames illustrating the principle of the WST. The left-hand frame shows the surface to be segmented. Starting from the local minima the surrounding basins of the surface start to flood as the water level continues to rise (dotted plane initially below the surface). Where two basins meet up near a ridge of the density surface, a 'dam' is erected (central frame). Ultimately, the entire surface is flooded, leaving a network of dams defines a segmented volume and delineates the corresponding cosmic web (right-hand frame).

dam. The final result (see third image in Fig. 1) of the completely immersed landscape is a division of the landscape into individual cells, separated by the ridge dams. In the remainder of this study we will use the word 'segment' to describe the watershed's cells.

2.2 Watershed segments: qualities

The watershed algorithm holds several advantages with respect to other void finders.

(i) Within an ideal smooth density field (i.e. without noise) it will identify voids in a parameter free way. No pre-defined values have to be introduced. In less ideal, and realistic, circumstances a few parameters have to be set for filtering out discreteness noise. Their values are guided by the properties of the data.

(ii) The watershed works directly on the topology of the field and does not reply on a pre-defined geometry/shape. By implication the identified voids may have any shape.

(iii) The watershed naturally places the divide lines on the crests of a field. The void boundary will be detected even when its boundary is distorted.

(iv) The transform naturally produces closed contours. As long as minima are well chosen the WST will not be sensitive to local protrusions between two adjacent voids.

Obviously we can only extract structural information to the extent that the point distribution reflects the underlying structure. Undersampling and shot noise always conspire to obfuscate the results, but we believe the present methodology provides an excellent way of handling this.

2.3 Voids and watersheds

The WVF is an implementation of the WST within a cosmological context. The watershed method is perfectly suited to study the holes and boundaries in the distribution of galaxies, and holds the specific promise of being able to recognize the void hierarchy that has been the incentive for our study.

The analogy of the WST with the cosmological context is straightforward: voids are to be identified with the basins, while the filaments and walls of the cosmic web are the ridges separating the voids from each other.

2.4 The watershed void finder: outline

An outline of the steps of the watershed procedure within its cosmological context is as follows.

(i) *DTFE.* Given a point distribution (*N*-body, redshift survey), the DTFE (Schaap & van de Weygaert 2000) is used to define a continuous density field throughout the sample volume. This guarantees a density field which retains the morphological character of the underlying point distribution, i.e. the hierarchical nature, the web-like morphology dominated by filaments and walls, and the presence voids is warranted.

(ii) *Grid sampling.* For practical processing purposes the DTFE field is sampled on a grid. The optimal grid size has to assure the resolution of all morphological structures while minimizing the number of needed grid cells. This criterion suggests a grid with grid cells whose size is in the order of the interparticle separation.

(iii) *Rank-ordered filtering.* The DTFE density field is adaptively smoothed by means of natural neighbour (NN) maxmin and median filtering. This involves the computation of the median, minimum or maximum of densities within the contiguous Voronoi cell, the region defined by a point and its NNs.

(iv) *Contour levels.* The image is transformed into a discrete set of density levels. The levels are defined by a uniform partitioning of the cumulative density distribution.

(v) *Pixel noise.* With an opening and closing (operation to be defined in Appendix A) of 2-pixel radius we further reduce pixel-by-pixel fluctuations.

(vi) *Field minima.* The minima in the smoothed density field are identified as the pixels (grid cells) which are exclusively surrounded by neighbouring grid cells with a higher density value.

(vii) *Flooding.* The flooding procedure starts at the location of the minima. At successively increasing flood levels the surrounding region with a density lower than the corresponding density threshold is added to the basin of a particular minimum. The flooding is illustrated in Fig. 1.

(viii) *Segmentation.* Once a pixel is reached by two distinct basins it is identified as belonging to their segmentation boundary. By continuing this procedure up to the maximum density level the whole region has been segmented into distinct void patches.

(ix) *Hierarchy correction.* A correction is necessary to deal with effects related to the intrinsic hierarchical nature of the void distribution. The correction involves the removal of segmentation

boundaries whose density is lower than some density threshold. The natural threshold value would be the typical void underdensity $\Delta = -0.8$ (see Section 3.4.1). Alternatively, dependent on the application, one may chose to take a user-defined value.

2.5 WVF by example: voids in a Λ cold dark matter (ΛCDM) simulation

A direct impression of the watershed void finding method is most readily obtained via the illustration of a representative example. In Fig. 2 the watershed procedure has been applied to the cosmological GIF2 simulation (Kauffmann et al. 1999).

The *N*-body particle distribution (left-hand frame, Fig. 2) is translated into a density field using the DTFE method. The application of the DTFE method is described in Section 3.1, the details of the DTFE procedure are specified in Appendix D.

The DTFE density field is sampled and interpolated on a 256^3 grid, the result of which is shown in the top right-hand frame of Fig. 2. The grey-scales are fixed by uniformly sampling the cumulative density distribution, ensuring that all grey-scale values have the same amount of volume.

The DTFE density field is smoothed by means of the adaptive NN median filtering described in Section 3.2. This procedure determines the filtered density values at the location of the particles. Subsequently, these are interpolated on to a grid. This field is translated into a grey-scale image following the same procedure as that for the raw DTFE image (bottom left-hand frame).

The minima in the smoothed density field are identified and marked as the flooding centres for the WST. The resulting WVF segmentation is shown in the bottom right-hand frame of Fig. 2.

The correspondence between the cosmic web, its voids and the watershed segmentation is striking. There is an almost perfect one-to-one correspondence between the segmentation and the void regions in the underlying density field. The WVF method does not depend on any pre-defined shape. As a result, the recovered voids do follow their natural shape. A qualitative assessment of the whole simulation cube reveals that voids are very elongated and have a preferential orientation within the cosmic web, perhaps dictated by the megaparsec tidal force field (see e.g. Lee & Park 2006).

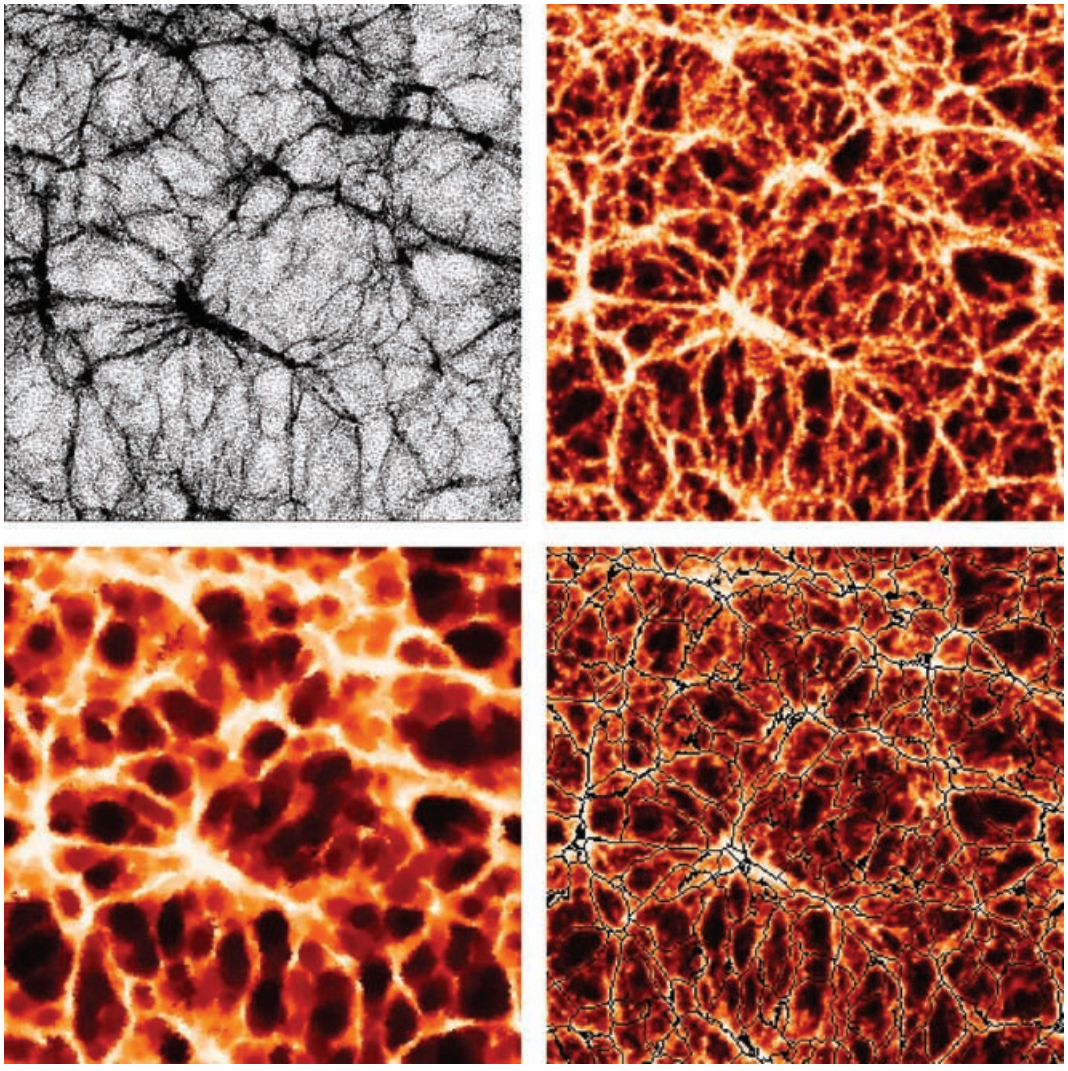

Figure 2. A visualization of several intermediate steps of the watershed void finding method. The top left-hand frame shows the particles of a slice in the ΛCDM GIF simulation. The corresponding DTFE density field is shown in the top right-hand frame. The next, bottom left-hand, frame shows the resulting fifth order median-filtered image. Bottom right-hand frame: the resulting WVF segmentation, computed on the basis of the median filtered image. The image shows the superposition of WVF ridges (black) on the original density field.

Clearly, the WVF is able to extract substructure at any level present in the density distribution. While this is an advantage with respect to tracing the presence of substructure within voids it does turn into a disadvantage when seeking to trace the outline of large-scale voids or when dealing with noise in the data set. While the noise-induced artificial segments are suppressed by means of the full machinery of markers (Section 3.3), void patch merging (Section 3.4) and NN rank filtering (Section 3.2), it are the latter two which may deal with intrinsic void hierarchy.

The follow-up study (Platen et al., in preparation) will involve a detailed quantitative analyse of volume and shapes of the voids in the GIF2 mass distribution for a sequence of time-steps.

3 METHOD: DETAILED DESCRIPTION

In order to appreciate the various steps of the WVF outlined in the previous section we need to describe a few of the essential steps in more detail.

To process a point sample into a spatial density field we use DTFE. To detect voids of a particular scale it is necessary to remove statistically insignificant voids generated by the shot noise of the discrete point sample as well as physically significant subvoids. In order to retain only the statistically significant voids we introduce and apply NN rank-order filtering. Hierarchy merging is used for the removal of subvoids which one would wish to exclude from a specific void study.

3.1 The DTFE density field

The input samples for our analysis are mostly samples of galaxy positions obtained by galaxy redshift surveys or the positions of a large number of particles produced by *N*-body simulations of cosmic structure formation. In order to define a proper continuous field from a discrete distribution of points – computer particles or galaxies – we translate the spatial point sample into a continuous density field by means of the DTFE (Schaap & van de Weygaert 2000).

3.1.1 DTFE

The DTFE technique (Schaap & van de Weygaert 2000) recovers fully volume-covering and volume-weighted continuous fields from a discrete set of sampled field values. The method has been developed by Schaap & van de Weygaert (2000) and forms an elaboration of the velocity interpolation scheme introduced by Bernardeau & van de Weygaert (1996). It is based upon the use of the Voronoi and Delaunay tessellations of a given spatial point distribution to form the basis of a natural, fully self-adaptive filter in which the Delaunay tessellations are used as multidimensional interpolation intervals. A typical example of a DTFE processed field is the one shown in the top row of Fig. 2: the particles of a GIF *N*-body simulation (Kauffmann et al. 1999) are translated into the continuous density field in the right-hand frame.

The primary ingredient of the DTFE method is the Delaunay tessellation of the particle distribution. The Delaunay tessellation of a point set is the uniquely defined and volume-covering tessellation of mutually disjunct Delaunay tetrahedra [triangles in two-dimensional (2D)]. Each is defined by the set of four points whose circumscribing sphere does not contain any of the other points in the generating set (Delaunay 1934). The Delaunay tessellation and the Voronoi tessellation of the point set are each others dual. The Voronoi tessellation is the division of space into mutually disjunct polyhedra, each polyhedron consisting of the part of space closer to the defining point than any of the other points (Voronoi 1908; Okabe et al. 2000).

DTFE exploits three properties of Voronoi and Delaunay tessellations (Schaap 2007; Schaap & van de Weygaert 2007). The tessellations are very sensitive to the local point density. DTFE uses this to define a local estimate of the density on the basis of the inverse of the volume of the tessellation cells. Equally important is their sensitivity to the local geometry of the point distribution. This allows them to trace anisotropic features such as encountered in the cosmic web. Finally, DTFE exploits the adaptive and minimum triangulation properties of Delaunay tessellations in using them as adaptive spatial interpolation intervals for irregular point distributions. In this way it is the first order version of the NN method (Watson 1992; Braun & Sambridge 1995; Sukumar 1998).

Within the cosmological context a major – and crucial – characteristic of a processed DTFE density field is that it is capable of delineating three fundamental characteristics of the spatial structure of the megaparsec cosmic matter distribution. It outlines the full hierarchy of substructures present in the sampling point distribution, relating to the standard view of structure in the Universe having arisen through the gradual hierarchical buildup of matter concentrations. DTFE also reproduces any anisotropic patterns in the density distribution without diluting their intrinsic geometrical properties. This is particularly important when analysing the prominent filamentary and planar features marking the cosmic web. A third important aspect of DTFE is that it outlines the presence and shape of void-like regions. Because of the interpolation definition of the DTFE field reconstruction voids are rendered as regions of slowly varying and moderately low-density values.

A more detailed outline of the DTFE reconstruction procedure can be found in Appendix D.

3.1.2 DTFE grid

DTFE involves the estimate of a continuous field throughout the complete sample volume. To process the DTFE field through the WVF machinery we sample the field on a grid. It is important to choose a grid which is optimally suited for the void finding purpose of the WVF method. On the one hand, the grid values should represent all physically significant structural features (voids) in the sample volume. On the other hand, the grid needs to be as coarse as possible in order to suppress the detection of spurious and insignificant features. The latter is also beneficial from a viewpoint of computational efficiency. This is achieved by adopting a grid size in the order of the mean interparticle distance.

The DTFE grid sampling is accomplished through Monte Carlo sampling within each grid cell. Within each grid cell the DTFE density value is measured at 10 randomly distributed sample points. The grid value is taken to be their average.

3.2 Natural neighbour rank-ordered filtering

A major and novel ingredient of our WVF method intended to eliminate shot noise in the DTFE density field reconstructions is that of a natural non-linear filtering extension: the NN rank-ordered filtering.

We invoke two kinds of non-linear adaptive smoothing techniques, median filtering and max/min filtering, the latter originating in MM. Both filters are rank order filters, and both have well known behaviour. They have a few important properties relevant for our purposes. Median filtering is very effective in removing shot noise while preserving the locations of edges. The max/min filters are designed to remove morphological features arising from shot noise (see Appendix A).

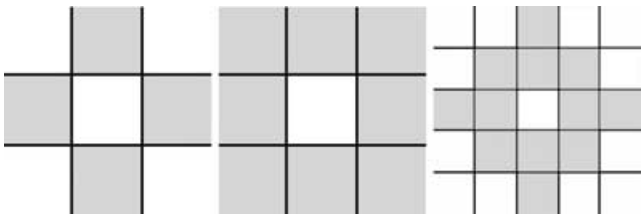

Figure 3. Examples of 2D grid connectivities. By default the central square is white. Cells connected to the centre are represented by grey squares. Left-hand frame: a four connectivity. Centre frame: an eight connectivity. Right-hand frame: a structure element representing a ball of 2 pixels.

The filters are defined over neighbourhoods. These are often named connectivity or, alternatively, structure elements. Image analysis usually deals with regular 2D image grids. The most common situation for such grids is straightforward four connectivities or eight connectivities (see Fig. 3). When a more arbitrary shape is used one usually refers to it as a structure element.

In the situation of our interest we deal with irregularly spaced data, rendering it impossible to use any of the above neighbourhoods. It is the Delaunay triangulation which defines a natural neighbourhood for these situations. For any point it consists of its NNs, i.e. all points to which it is connected via an edge of the Delaunay triangulation (see Fig. 4). This may be extended to any higher order natural neighbourhood: e.g. a second-order neighbourhood would include the NNs of the (first order) NNs.

The advantages of following this approach are the same as those for the DTFE procedure: the NN filtering – shortly named NN-median filtering or NN-min/max filtering – forms a natural extension to our DTFE-based formalism. It shares in the major advantage of being an entirely natural and self-adaptive procedure. The smoothing kernel is compact in regions of high point concentrations, while it is extended in regions of low density.

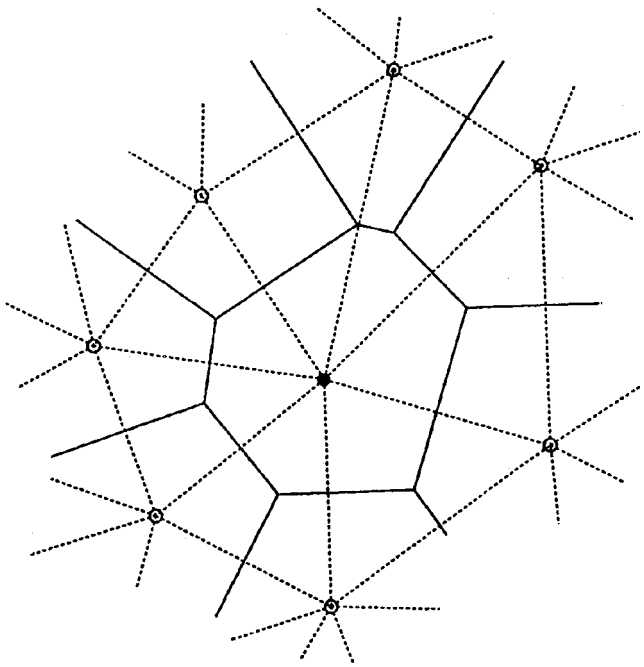

Figure 4. NNs of a point. The black dot represents the central point, the open circles its NNs. The solid edges mark the Voronoi cell surrounding the central point, along with the connecting Voronoi edges. The dashed lines delineate the corresponding Delaunay triangles. The central Voronoi cell is surrounded by its related Delaunay triangles, defining the NNs. The image is an illustration of the dual relationship between Voronoi and Delaunay tessellations.

3.2.1 Implementation of NN rank-order filtering

Implementing the min/max and median NN filters within the DTFE method is straightforward. The procedure starts with the DTFE density value at each of the (original) sample points. These may be the particles in an *N*-body simulation or the galaxies in a redshift survey. For each point in the sample the next step consists of the determination of the median, maximum or minimum value over the set of density values made up by that of the point itself and those of its NNs. The new 'filtered' density values are assigned to the points as the first-order filter value. This process is continued for a number of iterative steps, each step yielding a higher order filtering step.

The number of iterative steps of the NN smoothing is dependent on the size of the structure to be resolved and the sampling density within its realm. Testing has shown that a reasonable order of magnitude estimate is the mean number of sample points along the diameter of the structure. As an illustration of this criterion one may want to consult the low-noise and high-noise Voronoi models in Fig. 6 (below). While the void cells of the low-noise models contain on average six points per cell diameter, the void cells of the high-noise model contain around 16. Fifth-order filtering sufficed for the low-noise model, 20th order for the high-noise model (Figs 7 and 8, below).

In the final step, following the specified order of the filtering process, the filtered density values – determined at the particle positions – are interpolated on to a regular grid for practical processing purposes (see Section 3.1.2).

An example of a fifth-order median filtering process is shown in the bottom left-hand frame of Fig. 2. The comparison with the original DTFE field (top right-hand frame, Fig. 2) reveals the adaptive nature of the filtering process, suppressing noise in the low-density areas while retaining the overall topology of the density field. Figs 7(b) and 8(b) show it in the presence of controlled noise.

3.3 Markers and false segment removal

Following the NN-median smoothing of the DTFE density field, and some minor pixel noise removals, the WVF proceeds by identifying the significant minima of the density field. These are the markers for the WST. In the case of a cosmological density field the markers are the central deep minima in the (smoothed) density field.

Almost without exception the markers do not involve all minima in a raw unfiltered density field. The minima originating from shot noise need to be eliminated. In the unfiltered field each regional minimum would correspond to a catchment basin, producing oversegmented results: significant watershed basins would tend to get subdivided into an overabundance of smaller insignificant patches. While most of these segments are not relevant a beneficial property of the WST is that truly relevant edges constitute a subset of the oversegmented segmentation. This notion will be further exploited in Section 4.

Once the markers have been selected we compute the WST on the basis of an ordered queues algorithm. This process is described in detail in Beucher & Meyer (1993), and outlined in Appendix B. The process has a few important advantages. It is rather efficient because each point is processed only once while it naturally involves watershed by markers.

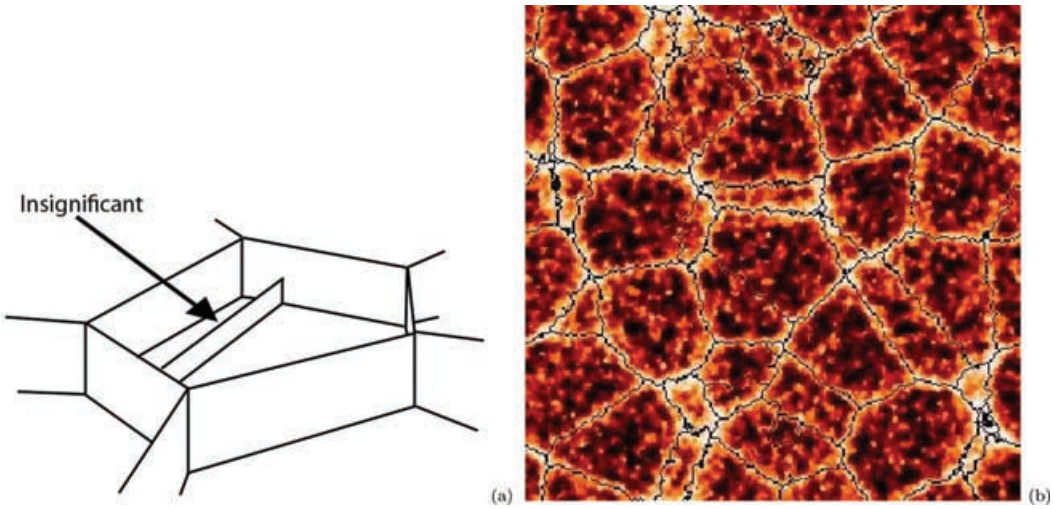

Figure 5. The concept of hierarchical watershed. Not all divide lines produced by the watershed may be relevant. They are removed if they do not fulfil a particular criterion (e.g. if they have a contrast lower than some threshold). Only the significant watershed segments survive. The segmentation after five iterative density smoothings and removal of boundaries below a contrast of 0.8.

Table 1. Parameters of the Voronoi kinematic model realizations: number of cells, cell filling factor, percentages of galaxies within each of the morphological components (clusters, filaments, walls, field) and the Gaussian width of clusters, filaments and walls.

Model	M	Cell filling factor	Field	Wall	$R_{\rm w}$ (h^{-1} Mpc)	Filament	$R_{\rm f}$ (h^{-1} Mpc)	Cluster	$R_{\rm c}$ (h^{-1} Mpc)
High noise	180	0.500	50.0	38.3	1.0	10.6	1.0	1.1	0.5
Low noise	180	0.025	2.5	16.4	1.0	40.6	1.0	40.5	0.5

3.4 Hierarchy merging

The WVF procedure combines two strategies to remove the artefacts generated by Poisson noise resulting from a density field discretely sampled by particles or galaxies:

(i) the pre-processing of the image such that the insignificant minima are removed;
(ii) merging of subdivided cells into larger ones.

The first strategy involves the previously described reconstruction of the density field by DTFE, followed by a combination of edge preserving median filtering and smoothing with the morphological erosion and dilation operators (Appendix A). In general, as will be argued and demonstrated in this study, it provides a good strategy for recovering only significant voids. The second strategy involves the merging of neighbouring patches via a user-specified scheme.

Amongst a variety of possibilities we have pursued a well-known method for merging patches, the watershed hierarchy. In its original form it assigns to each boundary a value dependent on the difference in density values between the minima of the neighbouring patches on either side of the ridge. We implemented a variant of this scheme where the discriminating value is that of the density value integrated over the boundary. A critical contrast threshold determines the outcome of the procedure. For an integral density value lower than the contrast threshold the two patches are merged. If the value is higher the edge is recognized as a genuine segment boundary.

The watershed hierarchy procedure is illustrated in Fig. 5(a). An example of its operation is provided by Fig. 5(b), one of the Voronoi clustering models extensively analysed in the remainder of this study. It depicts the segmentation resulting from watershed processing of a five times iteratively NN-median smoothed density field, followed by the hierarchical removal of boundaries. The improvement compared to the segmentation of a merely five times median smoothed density field is remarkable (cf. left-hand and right-hand frames, Fig. 8).

3.4.1 Merger threshold

In addition to the removal of features on morphological grounds, we also have to possibility to remove features on the basis of the involved density values.

In the case of voids we expect that they mature as they reach a density deficit of $\Delta \approx -0.8$ (see e.g. Sheth & van de Weygaert 2004). Any structures with a lower density may be residual features, the diminishing low-density boundaries of the subvoids which have merged (see e.g. Dubinski et al. 1993). Various void finding techniques do in fact exploit this notion and restrict their search to regions with $\Delta < -0.8$ (see e.g. Colberg et al. 2005). Note that in practice it may also involve noise, of considerable significance in these diluted regions.

A density threshold may indeed be applied within the WVF. This threshold is applied following the WST. Any ridges and features with a density contrast lower than a specified threshold are removed. The threshold $\Delta = -0.8$ is a natural value of choice. The goal is twofold: to suppress noise or spurious features within voids and to select out subvoids.

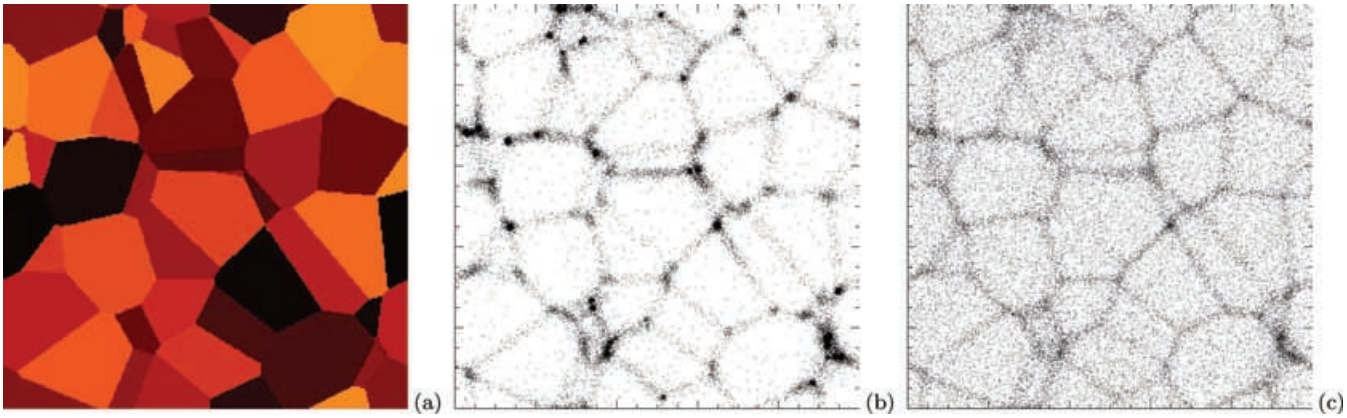

Figure 6. Frame (a) shows a slice through the original (geometrically defined) Voronoi tessellation. For two different Voronoi clustering models defined within this same tessellation, frames (b) and (c) depict the particles within the same slice. Frame (b) shows the low noise case with a high density contrast between the voids and walls. Frame (c) is a high noise model with a relatively low contrast between voids and walls.

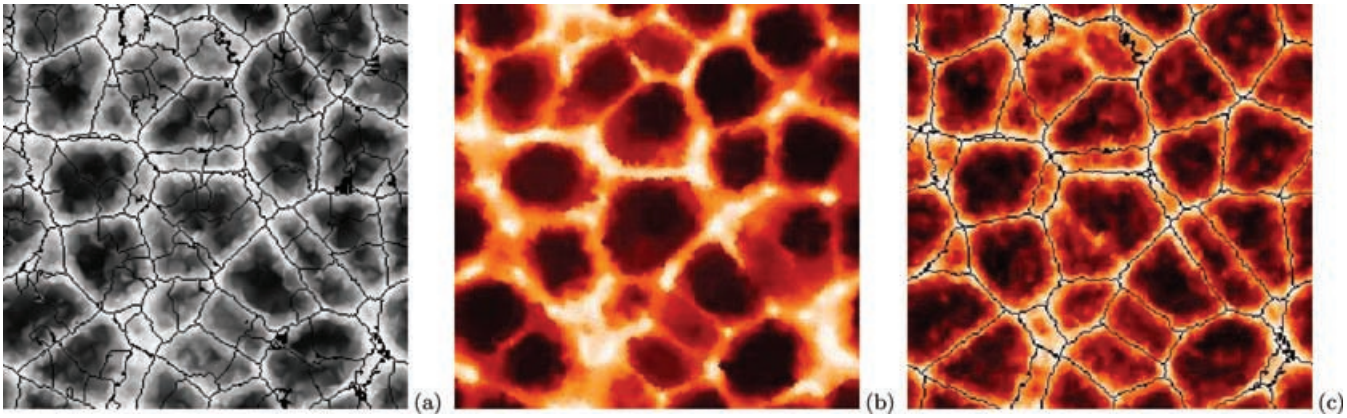

Figure 7. The density field of the particle distribution in the low noise model (a). Superimposed are the WVF segmentation boundaries. The central frame (b) shows the resulting fifth order median-filtered density field (b). This filtered field is the input for the watershed procedure whose segmentation is delineated in frame (c), superimposed on top of the original density field.

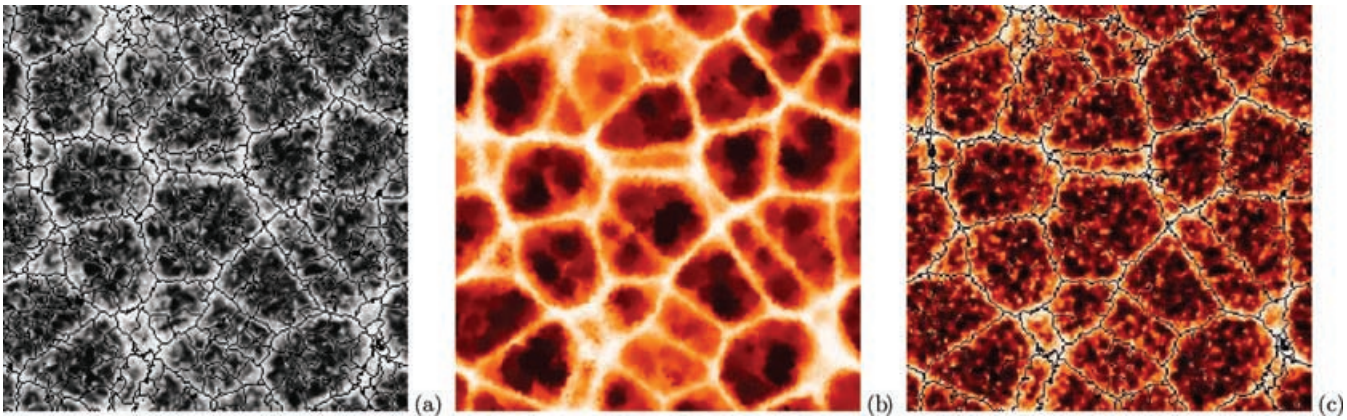

Figure 8. The density field of the particle distribution in the high noise model (a). Superimposed are the WVF segmentation boundaries. The central frame (b) shows the resulting 20th order median-filtered density field. The WVF segmentation of the fifth order median filtered density field, followed by removal of boundaries below a contrast of 0.8, is depicted in frame (c), superimposed on top of the original density field.

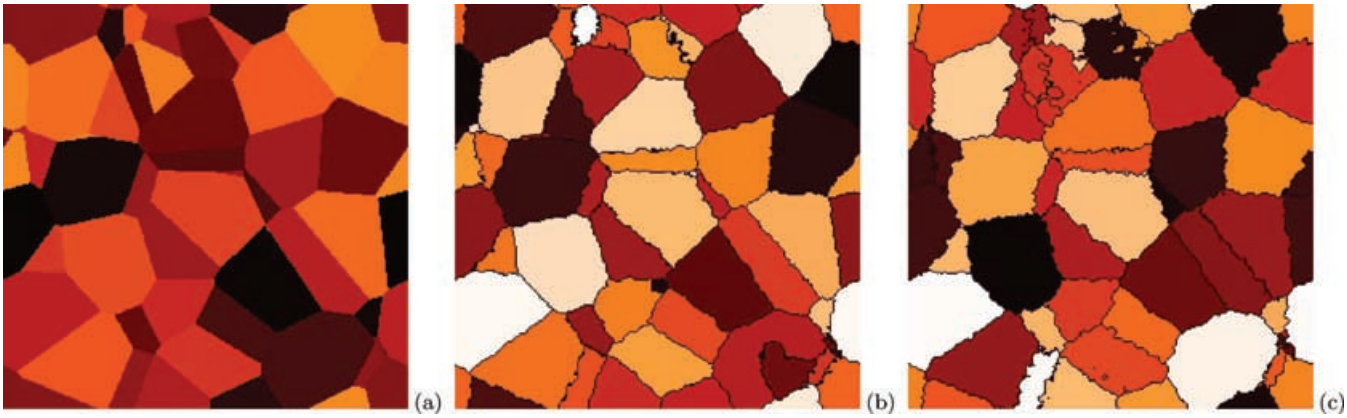

Figure 9. Frame (a): the original (geometric) Voronoi tessellation. Frames (b) and (c): the best recovered WVF segmentation of the low-noise (b) and high-noise (c) models.

4 WVF TEST: VORONOI CLUSTERING MODEL

To test and calibrate the WVF we have applied the WVF to a kinematic Voronoi model (van de Weygaert & Icke 1989; van de Weygaert 1991, 2002, and in preparation). In the case of the Voronoi models we have exact quantitative information on the location, geometry and identity of the Voronoi cells, whose interior functions as the voids in the matter distribution, against which we compare the outcome of the WVF analysis. These models combine the spatial intricacies of the cosmic web with the virtues of a model that has a priori known properties. They are particularly suited for studying systematic properties of spatial galaxy distributions confined to one or more structural elements of non-trivial geometric spatial patterns. The Voronoi models offer flexible templates for cellular patterns, and they are easy to tune towards a particular spatial cellular morphology.

Kinematic Voronoi models belong to the class of Voronoi clustering models. These are heuristic models for cellular spatial patterns which use the Voronoi tessellation as the skeleton of the cosmic matter distribution. The tessellation defines the structural frame around which matter will gradually assemble during the formation and growth of cosmic structure (Voronoi 1908; Okabe et al. 2000). The interior of Voronoi cells correspond to voids and the Voronoi planes with sheets of galaxies. The edges delineating the rim of each wall are identified with the filaments in the galaxy distribution. What is usually denoted as a flattened 'supercluster' will consist of an assembly of various connecting walls in the Voronoi foam, as elongated 'superclusters' of 'filaments' will usually include a few coupled edges. The most outstanding structural elements are the vertices, corresponding to the very dense compact nodes within the cosmic web, rich clusters of galaxies. We distinguish two different yet complementary approaches, Voronoi element models and kinematic Voronoi models. The kinematic Voronoi models are based upon the notion that voids play a key organizational role in the development of structure and make the Universe resemble a soapsud of expanding bubbles Icke (1984). It forms an idealized and asymptotic description of the outcome of the cosmic structure formation process within gravitational instability scenarios with voids forming around a dip in the primordial density field. This is translated into a scheme for the displacement of initially randomly distributed galaxies within the Voronoi skeleton (see Section C1 for a detailed specification). Within a void, the mean distance between galaxies increases uniformly in the course of time. When a galaxy tries to enter an adjacent cell, the velocity component perpendicular to the cell wall disappears. Thereafter, the galaxy continues to move within the wall, until it tries to enter the next cell; it then loses its velocity component towards that cell, so that the galaxy continues along a filament. Finally, it comes to rest in a node, as soon as it tries to enter a fourth neighbouring void. A detailed description of the model construction may be found in Section C1.

To test and calibrate the WVF technique we have applied the WVF to a high contrast/low noise Voronoi galaxy distribution and a low contrast/high noise one. Both concern two stages of the same kinematic Voronoi model, the high noise one to an early time-step with a high abundance of field galaxies and the low noise one to an advanced stage in which most galaxies have moved on towards filament or cluster locations. While the models differ substantially in terms of cell filling factor, the underlying geometric pattern remains the same: the position of the nodes, edges and walls occupy the same location. Most importantly for our purposes: the Voronoi cells, identified with the interior of the voids, are the same ones, be it that the high noise cells are marked by a substantial population of randomly distributed points.

The model has been set up in a (periodic) box with $141\,h^{-1}$ Mpc size, and is based on a Voronoi tessellation defined by 180 Voronoi cells. In total 128^3 particles were displaced following the kinematic Voronoi evolution. Table 1 specifies the distinctive parameters defining the model realizations, and Fig. 6 shows the particle distribution for the two model distributions in a central slice through the model box.

4.1 Voronoi model: watershed segmentation

The density/intensity field is determined by DTFE, yielding a 256^3 grid of density values. Fig. 7 contains an example of the outcome of the resulting DTFE density interpolation, with the contour levels determined according to the description in Section 2. The density map clearly reflects the filaments and nodes that were seen in the particle distribution. The void interiors are dominated by noise, visible as islands within a large zero density ocean.

A direct application of the WST results in a starkly oversegmented tessellation (Figs 7 and 8). Amongst the overabundance of mostly artificial, noise-related segments we may also discern real significant watersheds. Their boundary ridges (divide lines) are defined by

filaments, walls and clusters surrounding the voids. Many of these genuine voids are divided into small patches. They are the result of oversegmentation induced by the noisy Poisson point distribution within the cells. The local minima within this background noise will act as individual watershed flood centres marking corresponding, superfluous, watershed segments.

While for a general cosmological distribution it may be challenging to separate genuine physical subvoids from artificial noise-generated ones, the Voronoi kinematic models have the unique advantage of having no intrinsic substructure. Any detected substructure has to be artificial, rendering it straightforward to assess the action of the various steps intent on removing the noise contributions.

4.1.1 Smoothing and segment merging

The first step in the removal of insignificant minima consists of the application of the iterative NN median filtering process. This procedure, described in Section 3.2, removes some of the shot noise in the low-density regions. At the same time it is edge preserving. The result of five NN-median filtering iterations on the high noise version of the Voronoi kinematic clustering model is shown in Fig. 7. With the exception of a few artificial edges the resulting watershed segmentation almost perfectly matches the intrinsic Voronoi tessellation.

Fig. 8 shows the result for the high noise version of the same Voronoi kinematic clustering model. In this case pure NN-median filtering is not sufficient. A much more acceptable result is achieved following the application of the watershed hierarchy segment merging operation and the removal of ridges with a density contrast lower than the 0.8 contrast threshold.

For both the low-noise and high-noise realizations we find that the intrinsic and prominent edges of the Voronoi pattern remain in place. Nonetheless, a few shot-noise-induced artificial divisions survive the filtering and noise removal operations (Fig. 9). They mark prominent coherent but fully artificial features in the noise. Given their rare occurrence we accept these oversegmentations as inescapable yet insignificant contaminations.

5 VORONOI CLUSTERING MODEL: QUANTITATIVE RESULTS WATERSHED

The watershed segmentation retrieved by the WVF is compared with the intrinsic (geometric) Voronoi tessellation. The first test assesses the number of false and correct WVF detections. A second test concerns the volume distribution of the Voronoi cells and the corresponding watershed void segments.

5.1 Data sets

For our performance study we have three basic models: the intrinsic (geometric) Voronoi tessellation, and the low-noise and high-noise Voronoi clustering models (Table 1). The Voronoi clustering models are processed by WVF. In order to assess the various steps in the WVF procedure the models are subjected to different versions of the WVF.

The second column of Table 2 lists the differently WVF processed data sets. These are as follows.

(i) *Original.* The pure DTFE density field, without any smoothing or boundary removal, subjected to the WST.

Table 2. Quantitative comparison of the original and retrieved voids.

Model	Parameters	Voids	Splits	Mergers	Correct	Correctness
Intrinsic		180	–	–	–	–
	Original	847	–	–	–	–
	Max/min	259	82	3	118	66
Low noise	Med2	180	6	6	159	88
	Med5	162	9	30	119	66
	Med20	136	20	80	33	18
	Original	4293	–	–	–	–
	Max/min	3540	–	–	0	–
	Med5	723	529	0	8	4
High noise	Med20	275	95	3	100	55
	Hierarch	251	75	44	90	50
	Med5hr	172	6	12	144	80
	Med20hr	175	1	6	160	89

(ii) *Minmax.* Only the NN-min/max filtering is applied to the DTFE density field before watershed segmentation.

(iii) *Medn*. *n* iterations of median NN filtering is applied to the DTFE density field. In all situations this includes max/min filtering afterwards.

(iv) *Hierarch.* Following the WST, on the pure non-filtered DTFE density, a density threshold is applied. The applied hierarchy threshold level is $\rho/\rho_u = 0.8$: all segment boundaries with a density lower than $\delta < -0.2$ are removed as physically insignificant.

(v) *Mednhr*. Mixed process involving an *n* times iterated median filtered DTFE density field, followed by the WST, after which the segment boundaries below the hierarchy threshold $\delta < -0.2$ are removed.

Note that the physically natural threshold of $\Delta = -0.8$ is not really applicable to the heuristic Voronoi models. On the basis of the model specifications the threshold level has been set to $\Delta = -0.2$.

5.2 Detection rate

Each of the resulting segmentations is subjected to a range of detection assessments. These are listed in the third to seventh column of Table 2. The columns of the table contain, respectively, the number of WVF void detections, the amount of false splits, the amount of false mergers, the number of correctly identified voids and the correctness measure. While the top block contains information on the intrinsic (geometric) Voronoi tessellation, the subsequent two blocks contain the detection evaluations for the low-noise and high-noise models.

The false detections are split into two cases. The first case we name false splits: a break up of a genuine cell into two or more watershed voids. The second class is that of the false mergers: the spurious merging of two Voronoi cells into one watershed void. The splits, mergers and correct voids are computed by comparing the overlap between the volume of the Voronoi cell and that of the retrieved watershed void. A split is identified if the overlap percentage with respect to the Voronoi volume is lower than a threshold of 85 per cent of the overlapping volume. Along the same line, a merger concerns an overlap deficiency with respect to the watershed void volume. When both measures agree for at least 85 per cent a void is considered to be correct. The correctness of a certain segmentation is the percentage of correctly identified voids with respect the 180 intrinsic Voronoi cells.

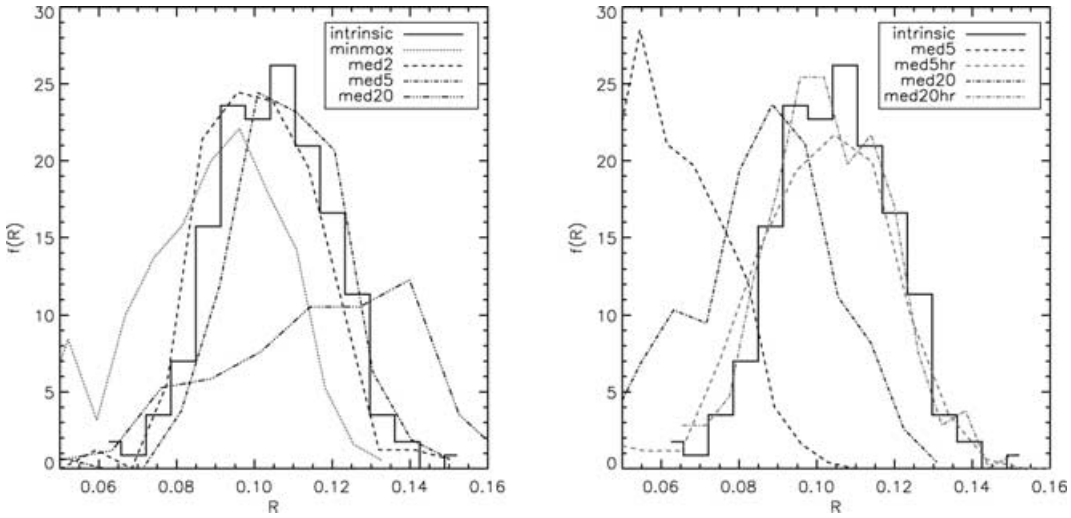

Figure 10. Left-hand frame: the volume distributions for void segments in low-noise models. The histogram shows the intrinsic distribution of the Voronoi cell volumes. Superimposed are the inferred volume distribution functions for the WVF segmentations of various Voronoi clustering models. The line style of each of the models is indicated in the insert. Right-hand frame: similar plot for a set of noisy Voronoi clustering models.

5.2.1 Low-noise model

Judging by the number of voids in the low-noise model, it is clear that smoothing or any other selection criterion remain necessary to reduce the number of minima from 850 to a number close to the intrinsic value 180. The second row shows the results for the case when just the maxmin filter is applied. This step already reduces the number of insignificant minima by already 60 per cent. It is an indication for the local character of the shot noise component. The next three rows list the results for various iterations of the median filtering. With just two iterations almost 90 per cent of the voids are retrieved. Most of the splits are removed at two iterations. This result does not improve with more median filtering, even up to 20 iterations this just increases the number of mergers as more walls are smoothed away. The number of splits also increases as minima begin to merge.

5.2.2 High-noise model

In general the same conclusion can be drawn for the high-noise model. Rank-ordered NN-median and NN-min/max filters manage to reduce the number of insignificant minima by a factor of 80 per cent (cf. the number of voids in the second and third row). These models attain a correctness of approximately 50 per cent. Mere rank-ordered filtering is evidently insufficient.

We also ran a threshold model which did not include median filtering. Instead only insignificant boundaries were removed. It achieved a recovery of 50 per cent. Combining both methods (med5hr and med20hr) recovers 80 to 90 per cent of the voids. The success rate may be understood by the complementarity of both methods: while the median filtering recovers the coherent structures, the thresholding will remove those coherent walls that are far underdense.

The translation to a cosmological density field is straightforward. The rank-ordered filtering ensures that insignificant minima are removed and that the watershed will pick up only coherent boundaries. Thresholding is able to order these walls by significance and to remove the very underdense and insignificant walls.

5.3 Volume comparison

In Fig. 10 we compare the distribution of the void volumes. The histogram shows the distribution of equivalent radii for the segment cells,

$$R \equiv \sqrt[3]{\frac{3}{4\pi}\,V}. \tag{1}$$

The solid line histogram shows the (geometric) volume distribution for the intrinsic Voronoi tessellations. On top of this we show the distributions for the various (parametrized) watershed segmentation models listed in Table 2. Not surprisingly the best segmentations have nearly equivalent volume distributions. For the low-noise models this is med2 (left-hand frame), for the high-noise models med20hr (right-hand frame). This conclusion is in line with the detection rates listed in Table 2.

The visual comparison of the intrinsic geometric Voronoi tessellations and the two best segmentations – med2 for the low-noise model and med20hr for the high-noise version – confirms that also the visual impression between these watershed renderings and the original Voronoi model is very much alike.

We have also assessed the cell-by-cell correspondence between the watershed segmentations and the Voronoi model. Identifying each watershed segment with its original Voronoi cell we have plotted the volume of all watershed cells against the corresponding Voronoi cell volumes. The scatter plots in Fig. 11 form a convincing confirmation of the almost perfect one-to-one relation between the volumes derived by the WVF procedure and the original volumes. The only deviations concern a few outliers. These are the hierarchy merger segments for which the watershed volumes are too large, resulting in a displacement to the right.

5.4 Surface comparison

While the volumes occupied by the watershed segments in Fig. 9 do overlap almost perfectly with that of the original Voronoi cells, their surfaces have a more noisy and erratic appearance. This is mostly a consequence of the shot noise in the (DTFE) density field, induced

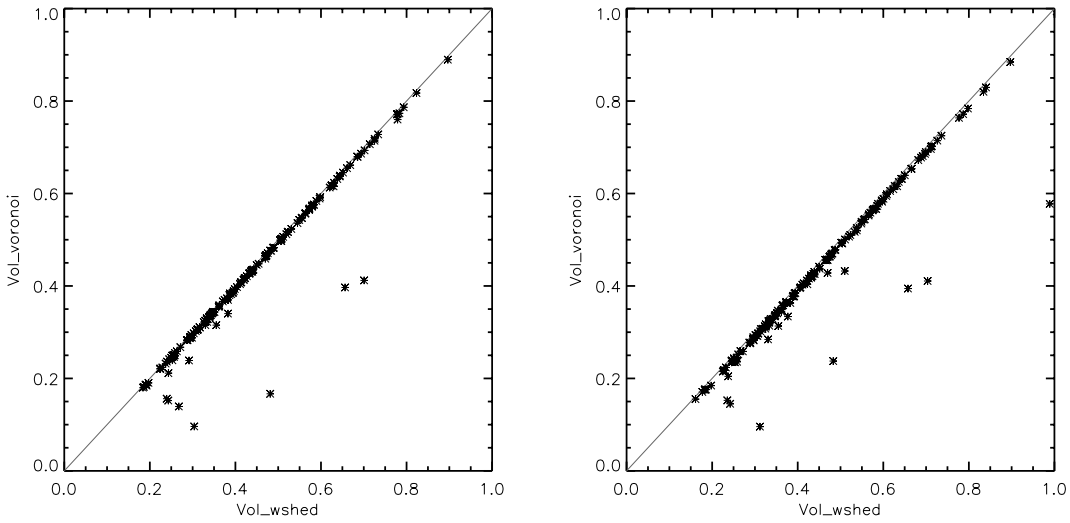

Figure 11. Scatter diagram plotting the WVF void segment volumes against the intrinsic geometric Voronoi cell volume. The solid line is the linear 1–1 relation. Left-hand frame: low-noise Voronoi clustering model. Right-hand frame: noisy Voronoi clustering model.

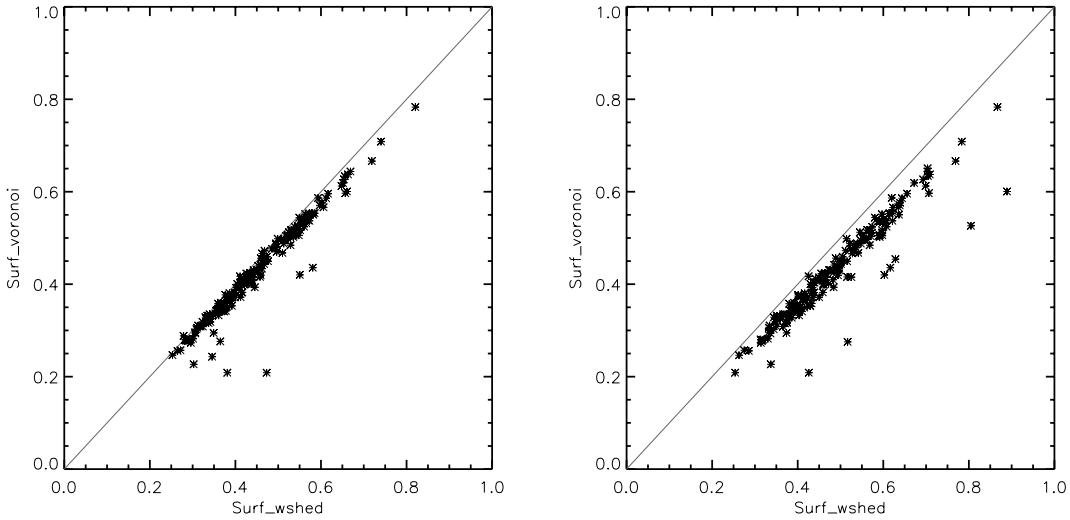

Figure 12. Scatter diagram plotting the WVF void segment surface area against the intrinsic geometric Voronoi cell volume. The solid line is the linear 1–1 relation. Left-hand frame: low-noise Voronoi clustering model. Right-hand frame: noisy Voronoi clustering model.

by the noise in the underlying point process. The crests in the density field are highly sensitive to any noise.

In addition to assess the impact of the noise on the surfaces of the watershed segments we compared the watershed segment surface areas with the Voronoi cell surface areas. The results are shown in Fig. 12. We tested the low noise med2 and the high noise med20hr. In both cases we find a linear relationship between the watershed surface and the genuine Voronoi surface area. Both cases involve considerably more scatter than that for the volumes of the cells. In addition to an increased level of scatter we also find a small be it significant offset from the 1–1 relation. The slope of the low-noise model is only slightly less than unity, the high-noise model slope deviates considerably more. These offsets do reflect the systematically larger surface areas of the watershed segments, a manifestation of their irregular surfaces. It is evident that the level of irregularity is more substantial for the high-noise than for the low-noise reconstructions (cf. Fig. 9).

The scatter plots do also reveal several cells with huge deviations in surface area. Unlike expected there is no systematic trend for smaller cells to show larger deviations. Some of the small deviating cells can be recognized in Fig. 9 as highly irregular patches. The large deviant cells correspond to watershed segments which as a result of noisy boundaries got wrongly merged.

While the irregularity of the surface areas forms a good illustration of the noise characteristics of the watershed patches, for the purpose of void identification it does not pose a serious problem. Smoother contours may always be obtained by applying the flooding process on a properly smoothed field. Some suggestions for how this may be achieved follow in the discussion.

6 DISCUSSION AND PROSPECTS

The WVF void finder technique is based on the WST known from the field of image processing. Voids are identified with the basins of

the cosmological mass distribution, the filaments and walls of the cosmic web with the ridges separating the voids from each other. Stemming from the field of MM, watershed cells are considered as patches locally minimizing the 'topographic distance'.

The WVF operates on a continuous density field. Because the cosmological matter distribution is nearly always sampled by a discrete distribution of galaxies or *N*-body particles, the first step of the WVF is to translate this into a density field by means of the DTFE. Because the WVF involves an intrinsically morphological and topological operation, the capability of DTFE to retain the shape and morphology of the cosmic web is essential. It guarantees that within this cosmological application the intrinsic property of the WST to act independent of scale, shape and structure of a segment is retained. As a result, voids of any scale, shape and structure may be detected by WVF.

In addition to the regular WST the WVF needs to invoke various operations to suppress (discreteness) sampling noise. In addition, we extend the watershed formalism such that the WVF will be capable of analysing the hierarchy of voids in a matter distribution, i.e. identify how and which small-scale voids are embedded within a void on larger scales (Platen et al., in preparation). Markers indicating significant void centres, false segment removal by hierarchy merging and NN filtering all affect an efficient noise removal. Hierarchy merging manages to eliminate boundaries between subvoids. NN median filtering, for various orders, is an essential new ingredient for highlighting the hierarchical embedding of the void population. It allows a natural selection of voids, unbiased with respect to the scale and shape of these structures. The voids that persist over a range of scales are expected to relate to the voids that presently dominate the cosmic matter distribution. In other words, WVF preserves the void hierarchy (Sheth & van de Weygaert 2004).

The present work includes a meticulous qualitative and quantitative assessment of the WST on the basis of a set of Voronoi kinematic models. These heuristic models of spatial web-like or cellular galaxy or particle distributions facilitate the comparison between the void detection of the WVF and that of the characteristics of the cells in the original and intrinsically known Voronoi tessellation. It is found that WVF is not only successful in reproducing the qualitative cellular appearance of the Voronoi models but also in reproducing quantitative aspects like an almost perfect 1–1 match of cell size with WVF segment volume and the corresponding void size distribution.

We foresee various possible improvements of the WVF. These concern in particular the identification of significant edges. One possibility is that extension proposed by Nguyen, Worring & van den Boomgaard (2003), in which not only the 'topographic costs' but also the lengths of the contours should be minimized. The length minimization will result in smoother boundaries. Additional improvements may be found in better filtering procedures in order to facilitate studies of hierarchically structured patterns. We expect considerable improvements by anisotropic diffusion techniques (Black & Marimont 1998) and are currently implementing these in the WVF code.

Given the results of our study, we are confident for applying WVF to more elaborate and realistic simulations of cosmic structure formation and on large galaxy redshift surveys. The analysis of a set of GIF cosmological simulations will be presented in an upcoming paper.

ACKNOWLEDGMENTS

We wish to thank Miguel Aragón-Calvo for permission to use Fig. C1. We are grateful to the referee, Mark Neyrinck, for the incisive, detailed and useful comments and recommendations for improvements. We particularly thank the participants of the KNAW colloquium 'Cosmic Voids' in Amsterdam, 2006 December, for the many useful discussions and encouraging remarks.

REFERENCES

Abazajian K. et al. (the SDSS collaboration), 2003, AJ, 126, 2081
Aikio J., Mähönen P., 1998, ApJ, 497, 534
Arbabi-Bidgoli S., Müller V., 2002, MNRAS, 332, 205
Benson A. J., Hoyle F., Fernando T., Vogeley M. S., 2003, MNRAS, 340, 160
Bernardeau F., van de Weygaert R., 1996, MNRAS, 279, 693
Bertschinger E., 1985, ApJS, 58, 1
Beucher S., Lantuejoul C., 1979, in Blum H., ed., International Workshop on Image Processing. CCETT/IRISA, Rennes, France, p. 2.1
Beucher S., Meyer F., 1993, in Dougherty E. R., ed., Mathematical Morphology in Image Processing. Marcel Dekker, Inc., New York, p. 433
Black M. J., Marimont D. H., 1998, IEEE Image Process., 7, 421
Blumenthal G. R., da Costa L. N., Goldwirth D. S., Piran T., 1992, ApJ, 388, 234
Bond J. R., Cole S., Efstathiou G., Kaiser N., 1991, ApJ, 379, 440
Bond J. R., Kofman L., Pogosyan D., 1996, Nat, 380, 603
Braun J., Sambridge M., 1995, Nat, 376, 655
Ceccarelli L., Padilla N. D., Valotto C., Lambas D. G., 2006, MNRAS, 373, 1440
Chincarini G., Rood H. J., 1975, Nat, 257, 294
Colberg J. M., Sheth R. K., Diaferio A., Gao L., Yoshida N., 2005, MNRAS, 360, 216
Colless M. et al. (the 2dFGRS team), 2003, VizieR Online Data Catalog, 7226
Dekel A., Rees M. J., 1994, ApJ, 433, L1
de Lapparent V., Geller M., Huchra J. P., 1986, ApJ, 302, L1
Delaunay B. N., 1934, Bull. Acad. Sci. USSR: Clase Sci. Mat., 7, 793
Dubinski J., da Costa L. N., Goldwirth D. S., Lecar M., Piran T., 1993, ApJ, 410, 458
Einasto J., Joeveer M., Saar E., 1980, MNRAS, 193, 353
El-Ad H., Piran T., 1997, ApJ, 491, 421
Fliche H. H., Triay R., 2006, preprint (gr-qc/0607090)
Friedmann Y., Piran T., 2001, ApJ, 548, 1
Furlanetto S. R., Piran T., 2006, MNRAS, 366, 467
Goldberg D. M., Vogeley M. S., 2004, ApJ, 605, 1
Gottlöber S., Łokas E. L., Klypin A., Hoffman Y., 2003, MNRAS, 344, 715
Gregory L. A., Thompson S. A., 1978, ApJ, 222, 748
Grogin N. A., Geller M. J., 1999, BAAS, 31, 1384
Hahn O., Porciani C., Carollo C. M., Dekel A., 2007, MNRAS, 375, 489
Heijmans H. J. A. M., 1994, Morphological Image Operators. Academic Press, New York
Hoeft M., Yepes G., Gottlöber S., Springel W., 2006, MNRAS, 371, 401
Hoffman Y., Shaham J., 1982, ApJ, 262, L23
Hoffman Y., Silk J., Wyse R. F. G., 1992, ApJ, 388, L13
Hoyle F., Vogeley M., 2002, ApJ, 566, 641
Icke V., 1984, MNRAS, 206, 1
Kauffmann G., Fairall A. P., 1991, MNRAS, 248, 313
Kauffmann G., Colberg J. M., Diaferio A., White S. D. M., 1999, MNRAS, 303, 188
Kirshner R. P., Oemler A., Schechter P. L., Shectman S. A., 1981, ApJ, 248, 57
Kirshner R. P., Oemler A., Schechter P. L., Shectman S. A., 1987, ApJ, 314, 493
Kresch R., 1998, in Heijmans H. J. A. M., Roerdink J. B. T. M., eds, Mathematical Morphology and Its Applications to Image and Signal Processing. Kluwer, Dordrecht, p. 35
Lee J., Park D., 2006, ApJ, 652, 1
Little B., Weinberg D. H., 1994, MNRAS, 267, 605
Martel H., Wasserman I., 1990, ApJ, 348, 1
Matheron G., 1975, Random Sets and Integral Geometry. Wiley, New York

Mathis H., White S. D. M., 2002, MNRAS, 337, 1193
Meyer F., 1994, Signal Process., 38, 113
Meyer F., Beucher S., 1990, J. Vis. Commun. Image Represent., 1, 21
Neyrinck M. C., Gnedin N. Y., Hamilton A. J. S., 2005, MNRAS, 356, 1222
Nguyen T. N., Worring M., van den Boomgaard R., 2003, IEEE Trans. Pattern Anal. Mach. Intell., 25, 330
Novikov D., Colombi S., Dore O., 2006, MNRAS, 366, 1201
Okabe A., Boots B., Sugihara K., Chiu S. N., 2000, Spatial Tessellations: Concepts and Applications of Voronoi Diagrams, 2nd edn. Wiley, New York
Padilla N. D., Ceccarelli L., Lambas D. G., 2005, MNRAS, 363, 977
Patiri S. G., Betancort-Rijo J. E., Prada F., Klypin A., Gottlöber S., 2006a, MNRAS, 369, 335
Patiri S. G., Prada F., Holtzman J., Klypin A., Betancort-Rijo J. E., 2006b, MNRAS, 372, 1710
Peebles P. J. E., 2001, ApJ, 557, 495
Plionis M., Basilakos S., 2002, MNRAS, 330, 399
Press W. H., Schechter P., 1974, ApJ, 187, 425
Regős E., Geller M. J., 1991, ApJ, 373, 14
Roerdink J., Meijster A., 2000, Fundam. Inform., 41, 187
Rojas R. R., Vogeley M. S., Hoyle F., Brinkmann J., 2005, ApJ, 624, 571
Ryden B. S., Melott A. L., 1996, ApJ, 470, 160
Schaap W., 2007, PhD thesis, Univ. Groningen
Schaap W., van de Weygaert R., 2000, A&A, 363, L29
Schaap W., van de Weygaert R., 2007, A&A, submitted
Schmidt J. D., Ryden B. S., Melott A. L., 2001, ApJ, 546, 609
Serra J., 1983, Image Analysis and Mathematical Morphology. Academic Press, New York
Shandarin S., Feldman H. A., Heitmann K., Habib S., 2006, MNRAS, 376, 1629
Sheth R. K., 1998, MNRAS, 300, 1057
Sheth R. K., van de Weygaert R., 2004, MNRAS, 350, 517
Sukumar N., 1998, PhD thesis, Northwestern University, Evanston, IL
Szomoru A., van Gorkom J. H., Greg M. D., Strauss M. A., 1996, AJ, 111, 2150
van de Weygaert R., 1991, PhD thesis, Leiden University
van de Weygaert R., 1994, A&A, 283, 361
van de Weygaert R., 2002, in Plionis M., Cotsakis S., eds, Proc. 2nd Hellenic Cosmology Meeting, ASSL Vol. 276, Modern Theoretical and Observational Cosmology. Kluwer, Dordrecht, p. 119
van de Weygaert R., Icke V., 1989, A&A, 213, 1
van de Weygaert R., van Kampen E., 1993, MNRAS, 263, 481
van de Weygaert R., Sheth R., Platen E., 2004, in Diaferio A., ed., Proc. IAU Colloq. 195, Outskirts of Galaxy Clusters: Intense Life in the Suburbs. Cambridge Univ. Press, Cambridge, p. 58
Vincent L., Soille P., 1991, IEEE Trans. Pattern Anal. Mach. Intell., 13, 583
Voronoi G. F., 1908, J. Reine Angew. Math., 134, 198
Watson D. F., 1992, Contouring: A Guide to the Analysis and Display of Spatial Data. Pergamon Press, New York

APPENDIX A: MATHEMATICAL MORPHOLOGY

Appendices A and B provide some formal concepts and notations necessary for appreciating the WST.

The WVF formalism introduced in this study is largely based upon concepts and techniques stemming from the field of image analysis. Although they are used within the context of the morphological analysis of spatial patterns in cosmological density fields the presentation is in terms of the original image analysis nomenclature. In this we remind the cosmology reader to translate 'image' into 'density field', 'basin' into 'void interior' etc.

MM is the field of image analysis which aims at characterizing an image on the basis of its geometrical structure. It provides techniques for extracting image components which are useful for representation and description and was originally developed by G. Matheron and J. Serra. For more details we refer to Serra (1983), also see Matheron (1975), Meyer & Beucher (1990) and Heijmans (1994). It involves a set-theoretic method of image analysis and incorporates concepts from algebra (set theory, lattice theory) as well as from geometry (translation, distance, convexity). Applications of MM may be found in a large variety of scientific disciplines, including material science, medical imaging and pattern recognition.

A1 Images

A cosmological density field $f(\boldsymbol{x})$ may be mapped on to an image $\mathcal{F}$,

$$f(\boldsymbol{x}) \rightarrow \mathcal{F}(\boldsymbol{x}). \tag{A1}$$

The image $\mathcal{F}$ is a function in n-dimensional lattice space (usually $n = 3$ or $n = 2$) $\mathbb{Z}^n$,

$$\mathcal{F} : \mathbb{Z}^n \rightarrow \mathbb{Z}. \tag{A2}$$

Although in principle images may be continuous, in practice they usually attain a finite number of discrete values. Two important classes are as follows.

(i) *Binary image.*
Image with only two intensity values (Fig. A1, top left-hand frame),

$$\mathcal{F} = \begin{cases} 0, \\ 1. \end{cases} \tag{A3}$$

We follow the convention to identify the binary image by the set $X \subseteq \mathbb{Z}^n$ for which $\mathcal{F} = 1$.

(ii) *Grey-scale image.*
Image with a discrete number N of values (Fig. A1, bottom left-hand frame),

$$\mathcal{F} = \{\mathcal{F}_i, \quad i = 1, \ldots, N\}. \tag{A4}$$

MM was originally developed for binary images, and was later extended to grey-scale images.

A2 Erosion and dilation

The two basic operators of MM are erosion and dilation of a binary image X. In order to define these operators we need to invoke the translation and reflection of a set B,

$$\begin{aligned} &\text{Translation}: \ B_z \equiv \{y \mid y = b + z \quad \forall \quad b \in B\}, \\ &\text{Reflection}: \ \hat{B} \equiv \{y \mid y = -b \quad \forall \quad b \in B\}. \end{aligned} \tag{A5}$$

The dilation or erosion of a binary image X by a structuring element B identifies whether the translated set B_z has an overlap with or is contained in a certain part of X,

$$\begin{aligned} &\text{Dilation}: \quad X \oplus B \equiv \{z \mid \hat{B}_z \cap X \neq \emptyset\}, \\ &\text{Erosion}: \quad X \ominus B \equiv \{z \mid B_z \subseteq X\}. \end{aligned} \tag{A6}$$

In other words, dilation consists of the Minkowski addition ($\oplus$) of the binary image X with a structuring element B while erosion is the Minkowski substraction ($\ominus$) with B. A structuring element may be any object in $\mathbb{Z}^n$. An example is the circle which functioned as a structuring element in Fig. A1. Erosion and dilation have a number of properties:

(i) translation invariance;
(ii) global scaling invariance;

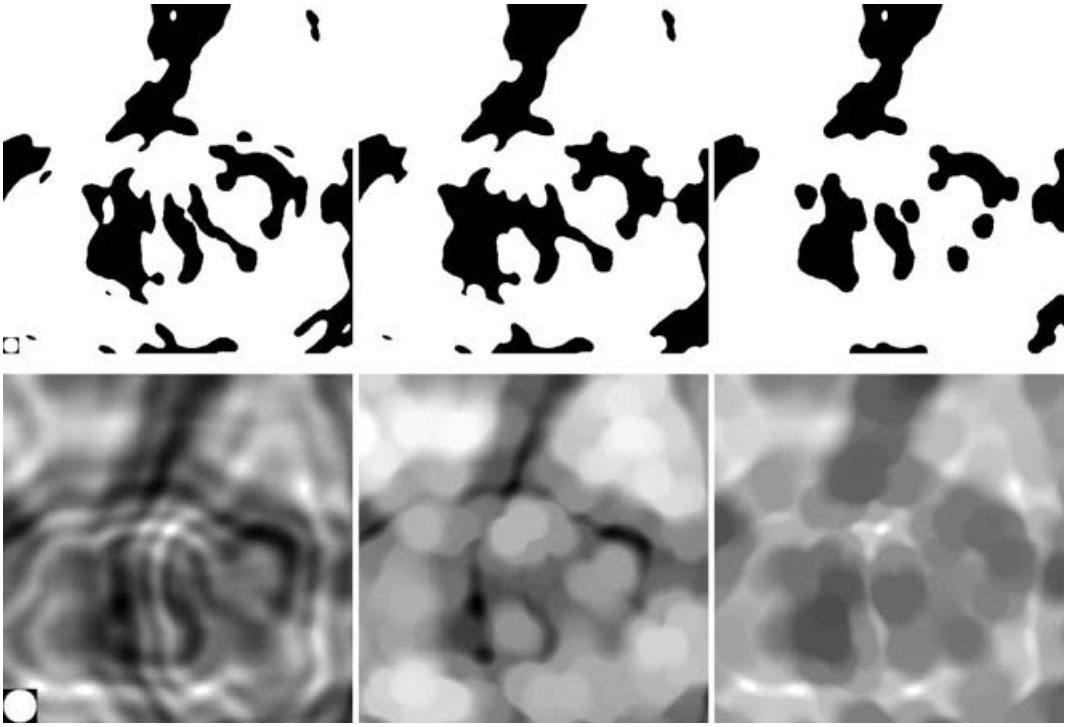

Figure A1. Illustration of the effects of a few essential operators on a binary image (top column) and a grey-scale image (bottom row). The original image is the one in the left-hand frame. The central frame contains the image following an opening operation, while the right-hand frame shows the effect of closing. The circle at the lower left-hand corner of the originals represents the circular structure element.

(iii) addition:

$$\text{if} \quad X \subseteq Y \;\rightarrow\; X \oplus A \subseteq Y \oplus A; \tag{A7}$$

(iv) complementarity:
erosion of a set is dilation of complement (and vice versa);
(v) adjunction relationship:

$$Y \oplus A \subseteq X \iff Y \subseteq X \ominus A. \tag{A8}$$

The complementarity and adjunction relationship are two aspects of the existing duality between erosion and dilation. In general erosion and dilation induce a loss of information. Erosion followed by dilation, or vice versa, will only result in a restoration of the original image if the sets X, Y and A are convex. In fact, various combinations of the erosion and dilation operators do result in new operators.

A3 Opening and closing

The most straightforward combination of dilation and erosion is that of the consecutive application of an erosion and a dilation. This introduces two new operators, the opening and closing operators. On a binary image they have the effects shown in Fig. A1 (top centre, top right-hand frame). Opening amends caps, removes small islands and opens isthmuses. Closing, on the other hand, closes channels, fills small lakes and (partly) the gulfs. An additional combination of erosion and dilation is subtraction of the first from the latter, called a morphological gradient.

Formally, an opening is an erosion followed by a dilation, a closing a dilation followed by erosion:

$$\begin{aligned} &\text{Opening}: \quad \Lambda_B \equiv [(X \ominus B) \oplus B], \\ &\text{Closing}: \quad \Phi_B \equiv [(X \oplus B) \ominus B]. \end{aligned} \tag{A9}$$

Characteristics of the opening and closing operators are the following:

(i) increasing;
(ii) idempotent:
applying the operator twice yields the same output;
(iii) opening is anti-extensive:

$$\Lambda(X) \subseteq X; \tag{A10}$$

(iv) closing is extensive:

$$X \subseteq \Phi(X). \tag{A11}$$

Note that the extensivity and/or anti-extensivity of operators define the prime conditions for a morphological operator.

A4 Grey-scale images

The morphological operators which we discussed above can be generalized to grey-scale images.

A grey-scale image is composed of subsets $S_i(\mathcal{F})$,

$$S_i(\mathcal{F}) = \{x \mid x \in \mathbb{Z}^n : \mathcal{F}(x) \geqslant i\}, \tag{A12}$$

with $S_{i+1}(\mathcal{F}) \subseteq S_i(\mathcal{F})$ and $S_1(\mathcal{F})$ the support of the full image. The erosion and dilation of a grey-scale image involves their application to each individual subset $S_i(\mathcal{F})$. Extension of the binary image definitions (equation A5) implies the following definition with respect to a grey-scale image,

$$\begin{aligned} \mathcal{F} \oplus B &\equiv \sup\{\mathcal{F}(x+b), \quad x \in X,\ b \in \hat{B}\}, \\ \mathcal{F} \ominus B &\equiv \inf\{\mathcal{F}(x+b), \quad x \in X,\ b \in B\}. \end{aligned} \tag{A13}$$

The effect of erosion on a grey-scale image is the shrinking of the bright regions. Bright spots smaller than the structuring element B disappear completely while valleys (dark) expand. Dilation has the opposite effect: dark regions shrink while bright regions expand. It illustrates the duality between erosion and dilation.

For our purposes, this formal definition translates into the following practical implementation for 2D grey-scale images. Given a grey-scale image $\mathcal{F}(A)$ with grid elements $a(i, j)$,

$$\begin{aligned} \mathcal{F} \oplus B &= \max_{(i,j)\in B} \{a[r+i, s+j] + \hat{b}[i, j]\} \\ &\qquad \forall [r, s] \in A, \\ \mathcal{F} \ominus B &= \min_{(i,j)\in B} \{a[r+i, s+j] + b[i, j]\} \\ &\qquad \forall [r, s] \in A. \end{aligned} \tag{A14}$$

As in the case of binary images new operators may be defined through combinations of erosions and dilations. The closing and opening operators are defined in exactly the same way as that for the binary images. Their effect is shown in the lower column of Fig. A1. The morphological gradient,

$$\mathcal{G} \equiv (\mathcal{F} \oplus B) \ominus (\mathcal{F} \ominus B), \tag{A15}$$

is a dilation minus erosion operation. The gradient operator is often used in object detection because an object is usually associated with a change in grey-scale with respect to the background. A variety of additional operators involving openings and closings may be defined. Interesting ones are the granulometries, a sequence of erosions with increasing scale, and distance transforms.

APPENDIX B: WATERSHED TRANSFORM

The segmentation of images is defined on the basis of a distance criterion, referring to the concept of distance between subsections of an image.

B1 Distance

For appreciating the concept of watershed segmentation we consider two distance concepts, the geodesic and topographic distance. Geodesic distances are used in the case of binary images while topographic distances form the basis for the segmentation of grey-scale images.

Let $X \subset \mathbb{Z}^n$ be a set and x and y two points in an n-dimensional lattice space $\mathbb{Z}^n$. We may define the following.

B1.1 Geodesic distance

The geodesic distance $d_X(x, y)$ is the length of the shortest (geometric) path in X connecting x and y (see left-hand frame, Fig. B1). Accordingly, the distance between two subsets A and B in X may be defined as follows. Considering the set of all paths between any of the elements of A and those of B, the distance $d_X(A, B)$ between A and B is defined to be the minimum length of any of these paths.

Based upon the concept of $d_X(A, B)$ one may formulate a distance function of a set $Y \in \mathbb{Z}^n$. For each point $y \in Y$, the distance $d(y, \bar{Y})$ to its complement $\bar{Y}$ is computed. The distance function $\mathcal{D}(y)$ is the resulting map of distance values $d_X(y, \bar{Y})$. Regions whose distance $d_X(y, \bar{Y})$ is at least r_i can be identified and equated by erosion of Y with a disc of radius r_i (defining a set $\mathcal{R}_i$). Each of these regions forms a section $\mathcal{S}_{\mathcal{D},i}$ (see Section A4),

$$\mathcal{S}_{\mathcal{D},i} = \mathcal{D}(y) \oplus \mathcal{R}_i, \tag{B1}$$

in which $\mathcal{R}_i$ is the disc of radius r_i. The map $\mathcal{D}(y)$ may be regarded as a stack of these sections. For illustration the distance transform of the binary image Fig. A1 is depicted in the central frame of Fig. B1.

B1.2 Topographic distance

The topographic distance $\mathcal{T}(x, y)$ between two points in $\mathbb{Z}^n$ is defined with respect to the image map $\mathcal{F}$. Taking the limit of a continuous map $\mathcal{F}$, the topographic distance from x to y is the path which attains the minimum length through the 'image landscape',

$$\mathcal{T}(x, y) \equiv \inf \int_\Gamma |\nabla \mathcal{F}(\gamma(s))| \, \mathrm{d}s. \tag{B2}$$

In this definition, the integral denotes the image path-length $\mathcal{F}(\gamma)$ along all paths $\gamma(s)$ in the set of all possible paths, Γ. This concept of distance is related to the geodesics of the surface $\mathcal{F}$: the path of steepest descent, specifying the track a droplet of water would follow as it flows down a mountain surface.

B2 Segmentation

Based on the specific definition of distance, we may segment a binary image $\mathcal{F}$ through the identification of the zones of influence of well-defined subsets of $\mathcal{F}$. In general the binary image $\mathcal{F}$ contains n connected subsets $Y_i (i = 1, \ldots, n)$ (the black regions of Fig. A1) and a set X (white region) which contains all points x which do not belong to any of the subsets Y_i, i.e.

$$X = \left\{ x \in \mathcal{F} | x \notin \bigcup_l Y_l \right\}. \tag{B3}$$

B2.1 Zone of influence

The geodesic zone of influence $Z_{\mathcal{F}}(Y_i)$ of a subset $Y_i \in \mathcal{F}$ is the set of all the points $x \in X$ that are strictly closer to Y_i than to any other

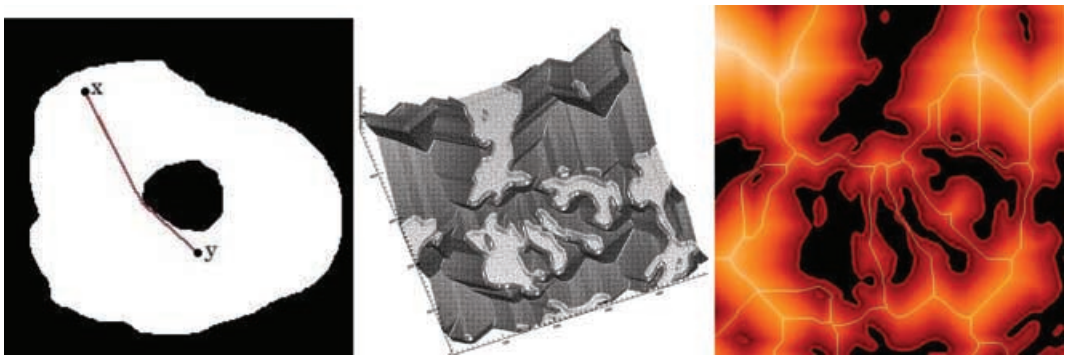

Figure B1. These three images illustrate the concept of distance and segmentation. Left: the geodesic distance between points x and y is depicted within a binary image. Centre/right: a landscape defined by a distance function (centre) and the resulting segmentation and skeleton (white line, right-hand frame).

subset $Y_j, j \neq i$,

$$Z_{\mathcal{F}}(Y_i) \equiv \left\{ x \in X \,|\, d_X(x, Y_i) < d_X(x, Y_j)\, \forall j \neq i \right\}. \quad \text{(B4)}$$

The zone of influence $\mathcal{Z}$ of $\mathcal{F}$ is the union of all influence zones of $Z_{\mathcal{F}}(Y_i)$,

$$\mathcal{Z}_{\mathcal{F}} \equiv \bigcup_n Z_{\mathcal{F}}(Y_i). \quad \text{(B5)}$$

B2.2 Skeleton

The boundary set in X consists of those points which do belong to X yet are not contained in any of the zones of influence $Z_{\mathcal{F}}(Y_i)$. These boundary points define the geodesic skeleton $\mathcal{K}$ of $\mathcal{F}$,

$$\mathcal{K}_{\mathcal{F}} \equiv X \backslash \mathcal{Z}_{\mathcal{F}}. \quad \text{(B6)}$$

In Fig. B1 (right-hand frame) the skeleton in set X is outlined by white lines. The skeleton is superimposed on its defining distance function landscape, its values indicated by a red colour gradient scheme (the corresponding landscape profile is depicted in the central frame). We should point out that here we follow the definition of MM although the name of skeleton of the large-scale cosmic matter distribution has been used for different be it related concepts (see e.g. van de Weygaert 1991; Novikov, Colombi & Dore 2006).

It is interesting to note that if we restrict the subsets Y_i to single points the skeleton naturally evolves into a (first-order) Voronoi tessellation. It is the definition of the concept of skeleton $\mathcal{K}$ within the specific context of grey-scale images which brings us to the definition of the WST (see next section).

B3 The watershed transform: algorithms

Grey-scale images consist of a finite number of discrete levels. This results in a slightly more complicated situation for its segmentation. In the case of a binary image an image is segmented on the basis of a geodesic distance. For the segmentation of grey-scale images the distance concept needs to be generalized to that of topographic distance.

Definition. Each watershed basin is the collection of points that are closer in topographic distance to the defining minimum than to any other minimum.

The literature is replete with algorithms for the construction of the WST of an image into its constituting watershed segments. They may be divided into two classes. One class simulates the watershed basin immersion process. The second aims at detecting the watershed skeleton on the basis of the distance.

B3.1 Watershed by immersion

Watershed by immersion was introduced and defined by Vincent & Soille (1991). The first step of the procedure concerns the identification of the minima. Formally, a minimum is a plateau at altitude h from which it is impossible to reach a point of lower height. Starting with the lowest grey-scale level $h_{\min}$ and recursively proceeding to the highest level $h_{\max}$ the algorithm allocates the zone of influence of each minimum by gradually filling up the surrounding catchment basin.

At a particular grey-scale level i, with altitude h_i, the algorithm has hypothetically inundated a landscape region with an altitude $\mathcal{F}(\boldsymbol{x}) \leqslant h_i$. The total of inundated area,

$$\tilde{\mathcal{S}}_i(\mathcal{F}) = \{x \in \mathbb{Z}^n \,|\, \mathcal{F}(x) \leqslant i\,\}, \quad \text{(B7)}$$

is the complement of the section $\mathcal{S}_i$. Having arrived at level i and proceeding to level $i + 1$ the algorithm has three possibilities:

1 encounter a new minimum (at level $i + 1$);
2 add new points to existing catchment basins (condition: points connected to only one existing basin);
3 encounter new points that belong to several basins.

Situation (1) signals the event that a new minimum becomes active in the image. The second option concerns the extension of an existing basin by an additional collection of points. These are points identified at level $(i + 1)$ which find themselves within the realm of a single basin existent at level i. They belong to the zone of influence of $\tilde{\mathcal{S}}_i$ and find themselves embedded in its extended counterpart $\tilde{\mathcal{S}}_{i+1}$ at level $(i + 1)$. In situation (3) more than one basin may be connected to the stack $\tilde{\mathcal{S}}_{i+1}$. Its correct subdivision is determined by computing the influence zones of all connected basins. Defining $\mathcal{B}_i(\mathcal{F})$ to be the union of all catchment basins at level i the union of catchment basins at level $(i + 1)$ becomes

$$\mathcal{B}_{i+1}(\mathcal{F}) = Z_{\tilde{\mathcal{S}}_{i+1}}(\tilde{\mathcal{S}}_i) \cup \mathcal{B}_i. \quad \text{(B8)}$$

The watershed procedure may be viewed as iteratively computing the zone of influence at each new grey scale level.

Following the rationale above the final 'immersion' definition for the watershed $\mathcal{W}$ of the image $\mathcal{F}$ within a domain X is

$$\mathcal{W}(\mathcal{F}) = X \setminus \mathcal{B}_{h_{\max}}. \quad \text{(B9)}$$

On completion of the procedure the union of points attached to every minimum m in X is equal to the union of catchment basins, $\mathcal{B}_{h_{\max}}$. The skeleton remains as the watershed segmentation.

B3.2 Watershed by topographic distance

The alternative strategy for determining the WST is that of following the strict definition of segmentation by minimum topographic distance. Roerdink & Meijster (2000) give a summary of the most notable schemes. These algorithms seek to find all points (pixels) whose topographic distance to a particular marker – i.e. a significant minimum in the density field – is the shortest amongst that to all other markers in the image.

The formalism bears some resemblance to Dijkstra's graph theoretical problem of tracing the shortest path forest in a point distribution. Based on this similarity an image is seen as a connected (di)graph in which the pixels of the image function as the nodes of the graph. Each point p is reachable from each other point p' via the graph's edges. The latter usually define a network on the basis of four connectivities or eight connectivities.

The shortest path between two points (nodes) p and p' is found by traversing the graph and keeping track of the walking cost. Critical for the procedure is the assignment of a proper measure of cost to each path. By definition it should be a non-negative increasing function and be related to the definition of topographic distance (equation B2). This suggests the use of the quantity

$$\mathcal{F}'(p, p') = \max\left\{ \frac{\mathcal{F}(p) - \mathcal{F}(p')}{d(p, p')} \right\}, \quad \text{(B10)}$$

the maximum slope linking the two pixels p and p'. This leads to the following cost function for the link between two neighbouring

pixel p and pixel p':

$$\mathcal{C}(p,p') = \begin{cases} \mathcal{F}'(p,p')\,d(p,p') & \mathcal{F}(p) > \mathcal{F}(p'), \\ \mathcal{F}'(p',p)\,d(p,p') & \mathcal{F}(p) < \mathcal{F}(p'), \\ \dfrac{\mathcal{F}'(p,p')+\mathcal{F}'(p',p)}{2}\,d(p,p') & \mathcal{F}(p) = \mathcal{F}(p'). \end{cases} \tag{B11}$$

The total cost $\mathcal{C}$ for a path $\gamma(p_1, p_2, \ldots p_n)$ connecting any two points p_1 and p_n via the points $\{p_2, \ldots, p_{n-1}\}$ will then simply be the sum

$$\mathcal{C}^{\gamma}(p_1, p_n) = \sum_{i \leqslant n} \mathcal{C}(p_i, p_{i-1}). \tag{B12}$$

The topographic distance $\mathcal{T}(p_1, p_n)$ is the infimum of $\mathcal{C}^{\gamma}(p_1, p_n)$ over all paths connecting p_1 and p_n:

$$\mathcal{T}(p_1, p_n) = \inf_{\Gamma} \mathcal{C}^{\gamma}(p_1, p_n). \tag{B13}$$

Given this definition for the topographic distance within a grey-scale image we can pursue the segmentation process described in Section B2, ultimately yielding the watershed segmentation.

B4 Ordered queues algorithm

We follow the WST algorithm by Beucher & Meyer (1993). Their method implicitly incorporates the concept of markers. These markers are the minima used as sources of the watershed flooding procedure. As such they form a select subgroup amongst all minima of an image $\mathcal{F}$.

The code for the watershed procedure involves the following steps.

(i) *Initialization.* All pixels of the cube are initialized and tagged to indicate they have not yet been processed. Each grey-scale level is allocated a queue and all pixels are attached to the queue corresponding to their level.

(ii) *Minima.* Each minimum plateau is tagged by a unique 'minimum tag'. The pixels corresponding to a minimum are inserted into the corresponding queue.

(iii) *Flooding.* All pixels in the grey-scale level queues are processed, starting at the lowest grey-scale level. Unless a pixel is surrounded by a complex of unprocessed neighbours it will be assigned to the queue of the corresponding minimum. Pixels which also border another minimum obtain a boundary tag.

(iv) *Final stage.* For any grey-scale level the flooding stops when the queue has emptied. The procedure continues with processing the pixels in the queue for the next grey-scale level. The process is finished once all level queues have been emptied.

APPENDIX C: KINEMATIC VORONOI MODELS

Voronoi clustering models are a class of heuristic models for cellular distributions of matter (van de Weygaert & Icke 1989; van de Weygaert 1991, 2002, and in preparation). They use the Voronoi tessellation as the skeleton of the cosmic matter distribution, identifying the structural frame around which matter will gradually assemble during the emergence of cosmic structure. The interior of Voronoi cells correspond to voids and the Voronoi planes with sheets of galaxies. The edges delineating the rim of each wall are identified with the filaments in the galaxy distribution. The most outstanding structural elements are the vertices, corresponding to the very dense compact nodes within the cosmic web, the rich clusters of galaxies.

We distinguish two different yet complementary approaches. One is the fully heuristic approach of Voronoi element models. They are particularly apt for studying systematic properties of spatial galaxy distributions confined to one or more structural elements of non-trivial geometric spatial patterns. The second, supplementary, approach is that of the Voronoi kinematic models, which attempts to 'simulate' foam-like galaxy distributions on the basis of simplified models of the evolution of the megaparsec scale distribution.

The Voronoi kinematic model is based upon the notion that voids play a key organizational role in the development of structure and make the Universe resemble a soapsud of expanding bubbles Icke (1984). It forms an idealized and asymptotic description of the outcome of the cosmic structure formation process within gravitational instability scenarios with voids forming around a dip in the primordial density field. For plausible structure formation scenarios, most notably the concordance ΛCDM cosmology, this evolution will proceed hierarchically. A detailed assessment of the resulting void hierarchy by Sheth & van de Weygaert (2004) demonstrated that this leads to a self-similarly evolving peaked void size distribution. By implication, most voids have comparable sizes and excess expansion rates. The geometrically interesting implication is that the asymptotic limit of the 'peaked' void distribution degenerating into one of only one characteristic void size. It yields a cosmic matter distribution consisting of equally sized and expanding spherical voids, a geometrical configuration which is precisely that of a Voronoi tessellation. This is translated into a scheme for the displacement of initially randomly distributed galaxies within the Voronoi skeleton (see Section C1 for a detailed specification). Within a void, the mean distance between galaxies increases uniformly in the course of time. When a galaxy tries to enter an adjacent cell, the velocity component perpendicular to the cell wall disappears. Thereafter, the galaxy continues to move within the wall, until it tries to enter the next cell; it then loses its velocity component towards that cell, so that the galaxy continues along a filament. Finally, it comes to rest in a node, as soon as it tries to enter a fourth neighbouring void.

C1 Initial conditions

The initial conditions for the Voronoi galaxy distribution are as follows.

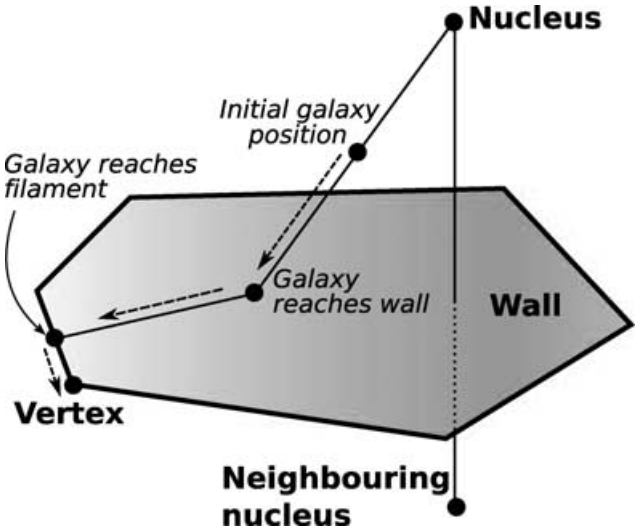

Figure C1. Schematic illustration of the Voronoi kinematic model. Courtesy: Miguel Aragón.

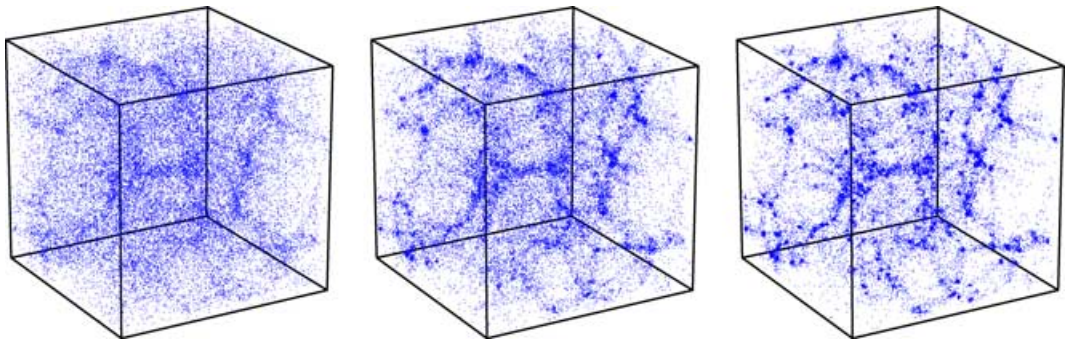

Figure C2. Evolution of galaxy distribution in the Voronoi kinematic model. A sequel of three consecutive time-steps within the kinematic Voronoi cell formation process, proceeding from left to right, and from top to bottom. The depicted boxes have a size of $100\,h^{-1}$ Mpc. Within these cubic volumes some 64 Voronoi cells with a typical size of $25\,h^{-1}$ Mpc delineate the cosmic framework around which some 32 000 galaxies have aggregated. Taken from a total (periodic) cubic 'simulation' volume of $200\,h^{-1}$ Mpc containing 268 235 'galaxies'.

(i) Distribution of M nuclei, expansion centres, within the simulation volume V. The location of nucleus m is $\boldsymbol{y}_m$.

(ii) Generate N model galaxies whose initial locations, $\boldsymbol{x}_{n0}$ ($n = 1, \ldots, N$), are randomly distributed throughout the sample volume V.

(iii) Of each model galaxy n determine the Voronoi cell $\mathcal{V}_\alpha$ in which it is located, i.e. determine the closest nucleus j_α.

All different Voronoi models are based upon the displacement of a sample of N 'model galaxies'. The initial spatial distribution of these N galaxies within the sample volume V is purely random, their initial locations $\boldsymbol{x}_{n0} (n = 1, \ldots, N)$ defined by a homogeneous Poisson process. A set of M nuclei within the volume V corresponds to the cell centres, or expansion centres driving the evolving matter distribution. The nuclei have locations $\boldsymbol{y}_m$ ($m = 1, \ldots, M$).

Following the specification of the initial positions of all galaxies, the second stage of the procedure consists of the calculation of the complete Voronoi track for each galaxy $n = 1, \ldots, N$ (Section C2). Once the Voronoi track has been determined, for any cosmic epoch t one may determine the displacement $\boldsymbol{x}_n$ that each galaxy has traversed along its path in the Voronoi tessellation (Section C2).

C2 Voronoi tracks

The first step of the formalism is the determination for each galaxy n the Voronoi cell $\mathcal{V}_\alpha$ in which it is initially located, i.e. finding the nucleus j_α which is closest to the galaxies' initial position $\boldsymbol{x}_{n0}$.

In the second step the galaxy n is moved from its initial position $\boldsymbol{x}_{n0}$ along the radial path emanating form its expansion centre j_α, i.e. along the direction defined by the unity vector $\hat{\boldsymbol{e}}_{n\alpha}$. Dependent on how far the galaxy is moved away from its initial location $\boldsymbol{x}_{n0}$ – set by the radius of expansion R_n to be specified later – the galaxies' path $\boldsymbol{x}_n$ (see Fig. C1) may be codified as

$$\begin{aligned}\boldsymbol{x}_n &= \boldsymbol{y}_\alpha + \boldsymbol{s}_{n\alpha} + \boldsymbol{s}_{n\alpha\beta} + \boldsymbol{s}_{n\alpha\beta\gamma} \\ &= \boldsymbol{y}_\alpha + s_{n\alpha}\hat{\boldsymbol{e}}_{n\alpha} + s_{n\alpha\beta}\hat{\boldsymbol{e}}_{n\alpha\beta} + s_{n\alpha\beta\gamma}\hat{\boldsymbol{e}}_{n\alpha\beta\gamma}\end{aligned} \tag{C1}$$

in which the four different components are

(i) $\hat{\boldsymbol{e}}_{n\alpha}$: unity vector path within Voronoi cell $\mathcal{V}_\alpha$;
(ii) $\hat{\boldsymbol{e}}_{n\alpha\beta}$: unity vector path within Voronoi wall $\Sigma_{\alpha\beta}$;
(iii) $\hat{\boldsymbol{e}}_{n\alpha\beta\gamma}$: unity vector path along Voronoi edge $\Lambda_{\alpha\beta\gamma}$;
(iv) Vertex $\Xi_{\alpha\beta\gamma\delta}$.

The identity of the neighbouring nuclei $j_\alpha, j_\beta, j_\gamma$ and j_δ, and therefore the identity of the cell $\mathcal{V}_\alpha$, the wall $\Sigma_{\alpha\beta}$, the edge $\Lambda_{\alpha\beta\gamma}$ and the vertex $\Xi_{\alpha\beta\gamma\delta}$, depends on the initial location $\boldsymbol{x}_{n0}$ of the galaxy, the position $\boldsymbol{y}_\alpha$ of its closest nucleus and the definition of the galaxies' path within the Voronoi skeleton.

The cosmic matter distribution at a particular cosmic epoch is obtained by calculating the individual displacement factors ($s_{n\alpha}(t)$, $s_{n\alpha\beta}(t)$, $s_{n\alpha\beta\gamma}(t)$) for each model galaxy. These are to be derived from the global 'void' expansion factor $R(t)$. This factor parametrizes the cosmic epoch and specifies the (virtual) radial path of the galaxy from its expansion centre j_α.

At first, while still within the cell's interior, the galaxy proceeds according to

$$\boldsymbol{s}_{n\alpha}(t) = \boldsymbol{x}_n(t) - \boldsymbol{y}_\alpha = R(t)\,|\boldsymbol{x}_{n0} - \boldsymbol{y}_\alpha|\,\hat{\boldsymbol{e}}_{n\alpha}. \tag{C2}$$

As a result within a void the mean distance between galaxies increases uniformly in the course of time. Once the galaxy tries to enter an adjacent cell j_β and reaches a Voronoi wall, i.e. when $R(t)\,|\boldsymbol{x}_{n0} - \boldsymbol{y}_\alpha| > \upsilon_n$, the galaxy's motion will be constrained to the radial path's component within the wall $\Sigma_{\alpha\beta}$. The galaxy moves along the wall until the displacement supersedes the extent of the path within the wall and it tries to enter a third cell j_γ, i.e. when $s_{n\alpha\beta}(t) > \sigma_n$. Subsequently, it moves along $\Lambda_{\alpha\beta\gamma}$ until it comes to rest at the node $\Xi_{\alpha\beta\gamma\delta}$ as soon as it tries to enter a fourth neighbouring void j_δ when $s_{n\alpha\beta\gamma} > \lambda_n$.

A finite thickness is assigned to all Voronoi structural elements. The walls, filaments and vertices are assumed to have a Gaussian radial density distribution specified by the widths $R_{\rm W}$ of the walls, $R_{\rm F}$ of the filaments and $R_{\rm V}$ of the vertices. Voronoi wall galaxies are displaced according to the specified Gaussian density profile in the direction perpendicular to their wall. A similar procedure is followed for the Voronoi filament galaxies and the Voronoi vertex galaxies. As a result the vertices stand out as three-dimensional (3D) Gaussian peaks.

A kinematic model time sequence is shown in Fig. C2.

APPENDIX D: DTFE RECONSTRUCTION PROCEDURE

For a detailed specification of the DTFE density field procedure we refer to Schaap (2007). In summary, the DTFE procedure for density field reconstruction from a discrete set of points consists of the following steps.

(i) *Point sample.*
Given that the point sample is supposed to represent an unbiased reflection of the underlying density field, it needs to be

a general Poisson process of the (supposed) underlying density field.

(ii) *Boundary conditions.*
The boundary conditions will determine the Delaunay and Voronoi cells that overlap the boundary of the sample volume. Dependent on the sample at hand, a variety of options exists.

+ *Empty boundary conditions:*
outside the sample volume there are no points.
+ *Periodic boundary conditions:*
the point sample is supposed to be repeated periodically in boundary boxes, defining a toroidal topology for the sample volume.
+ *Buffered boundary conditions:*
the sample volume box is surrounded by a buffer zone filled with a synthetic point sample.

(iii) *Delaunay tessellation.*
Construction of the Delaunay tessellation from the point sample. While we also still use the Voronoi–Delaunay code of van de Weygaert (1991, 1994), at present there are a number of efficient library routines available. Particularly noteworthy is the CGAL initiative, a large library of computational geometry routines.[1]

(iv) *Field values point sample.*
The estimate of the density at each sample point is the normalized inverse of the volume of its contiguous Voronoi cell $\mathcal{W}_i$ of each point i. The contiguous Voronoi cell of a point i is the union of all Delaunay tetrahedra of which point i forms one of the four vertices. We recognize two applicable situations.

− *Uniform sampling process*:
the point sample is an unbiased sample of the underlying density field. Typical example is that of N-body simulation particles. For D-dimensional space the density estimate is

$$\widehat{\rho}(\boldsymbol{x}_i) \,=\, (1+D)\,\frac{w_i}{V(\mathcal{W}_i)}, \tag{D1}$$

with w_i the weight of sample point i, usually we assume the same 'mass' for each point.

− *Systematic non-uniform sampling process*:
sampling density according to specified selection process. The non-uniform sampling process is quantified by an a priori known selection function $\psi(\boldsymbol{x})$. This situation is typical for galaxy surveys, $\psi(\boldsymbol{x})$ may encapsulate differences in sampling density $\psi(\alpha, \delta)$ as function of sky position (α, δ), as well as the radial redshift selection function $\psi(r)$ for magnitude- or flux-limited surveys. For D-dimensional space the density estimate is

$$\widehat{\rho}(\boldsymbol{x}_i) \,=\, (1+D)\,\frac{w_i}{\psi(\boldsymbol{x}_i)\,V(\mathcal{W}_i)}. \tag{D2}$$

(v) *Field gradient.*
Calculation of the field gradient estimate $\widehat{\nabla f}|_m$ in each D-dimensional Delaunay simplex m ($D = 3$: tetrahedron; $D = 2$: triangle) by solving the set of linear equations for the field values f_i at the positions $\boldsymbol{r}_i$ of the $(D + 1)$ tetrahedron vertices:

$$\widehat{\nabla f}|_m \;\Longleftarrow\; \begin{cases} f_0 \quad f_1 \quad f_2 \quad f_3, \\ \boldsymbol{r}_0 \quad \boldsymbol{r}_1 \quad \boldsymbol{r}_2 \quad \boldsymbol{r}_3. \end{cases} \tag{D3}$$

Evidently, linear interpolation for a field f is only meaningful when the field does not fluctuate strongly.

(vi) *Interpolation.*
The final basic step of the DTFE procedure is the field interpolation. The processing and post-processing steps involve numerous interpolation calculations, for each of the involved locations $\boldsymbol{x}$. Given a location $\boldsymbol{x}$, the Delaunay tetrahedron m in which it is embedded is determined. On the basis of the field gradient $\widehat{\nabla f}|_m$ the field value is computed by (linear) interpolation,

$$\widehat{f}(\boldsymbol{x}) \,=\, \widehat{f}(\boldsymbol{x}_i) \,+\, \widehat{\nabla f}\big|_m \cdot (\boldsymbol{x} - \boldsymbol{x}_i). \tag{D4}$$

In principle, higher order interpolation procedures are also possible. Two relevant procedures are
− *spline interpolation*;
− *NN interpolation*.
For NN interpolation, see Watson (1992), Braun & Sambridge (1995), Sukumar (1998) and Okabe et al. (2000). Implementation of NN interpolations is presently in progress.

(vii) *Processing.*
Though basically of the same character, for practical purposes we make a distinction between straightforward processing steps concerning the production of images and simple smoothing filtering operations and more complex post-processing. The latter are treated in the next item. Basic to the processing steps is the determination of field values following the interpolation procedure(s) outlined above. Straightforward 'first line' field operations are image reconstruction and smoothing/filtering.

+ *Image reconstruction:*
for a set of image points, usually grid points, determine the image value, formally the average field value within the corresponding grid cell. In practice a few different strategies may be followed:
− *formal geometric approach*;
− *Monte Carlo approach*;
− *singular interpolation approach*.
The choice of strategy is mainly dictated by accuracy requirements. For WVF we use the Monte Carlo approach in which the grid density value is the average of the DTFE field values at a number of randomly sampled points within the grid cell.

+ *Smoothing and filtering:*
a range of filtering operations is conceivable. Two of relevance to WVF are
− *linear filtering* of the field $\widehat{f}$:
convolution of the field $\widehat{f}$ with a filter function $W_s(\boldsymbol{x}, \boldsymbol{y})$, usually user specified,

$$f_s(\boldsymbol{x}) \,=\, \int \widehat{f}(\boldsymbol{x}')\, W_s(\boldsymbol{x}', \boldsymbol{y})\, \mathrm{d}\boldsymbol{x}'; \tag{D5}$$

− *NN rank-ordered filtering*
(see Section 3.2).

(viii) *Post-processing.*
The real potential of DTFE fields may be found in sophisticated applications, tuned towards uncovering characteristics of the reconstructed fields. An important aspect of this involves the analysis of structures in the density field. The WVF formalism developed in this study is an obvious example.

[1] CGAL is a C++ library of algorithms and data structures for computational geometry, see http://www.cgal.org/

This paper has been typeset from a TEX/LATEX file prepared by the author.