**IPAM workshop: Multiscale Geometric Methods in Astronomical Data Analysis** 

UCLA, November 9, 2004

Cosmological fields and the distribution of galaxies: Multiresolution morphology

With thanks to John Peacock for some slices

### Vicent J. Martínez



In collaboration with: J.-L. Starck, D. Donoho, E. Saar, S. Paredes, and S. Reynolds





Outline of the talk: Galaxies everywhere (on the sky and in the computer) • Second-order statistics (correlation function and power spectrum) Multiresolution morphology (consistent estimation of the density field)

Outline of the talk: (on the sky and in the computer) • Second-order statistics (correlation function and power spectrum) Multiresolution morphology (consistent estimation of the density field)

# The first Hubble Deep Field









### Maps of the Galactic extinction White is misty (opaque), black is

Dust 270 270 180 80 90 90 0.33 MJy/sr 30. Log scale Coutersy of D. Schlegel, D. Finkbeiner, A. Kriegel, and M. Davis

#### J. Peacock, JENAM '04 The distribution of the galaxies



#### 1930s:

Non-Poisson (Hubble)

### 1950s:

Shane & Wirtanen spend 10 years counting 1,000,000 galaxies by eye

 filamentary patterns?

### 1980s:

Take a strip and get redshifts

### Mapping the Universe



Cosmological Distances  $D_1$ : luminosity distance  $D_m$ : Mattig distance  $D_a$ : angular diameter distance r: comoving distance





J. Peacock, JENAM '04

Redshift surveys (mid-1980s)

Inverting v = cz = Hd gives an approximate distance.

Applied to galaxies on a strip on the sky, gives a 'slice of the universe'



"The beginning of the end" or "the end of greatness" ... R. Kirshner

## 2dFGRS input catalogue from UKST/APM

### • Galaxies: $b_J \le 19.45$ from revised APM

The APM Galaxy survey Maddox Sutherland Efstathiou & Loveday

- Total area on sky ~ 2000 deg<sup>2</sup>
- 250,000 galaxies in total, 93% sampling rate
- Mean redshift  $\langle z \rangle \sim 0.1$ , almost all with z < 0.3

J. Peacock, JENAM '04



### A volume-limited sample:







#### 2dFGRS (Southern slice, volume limited sample)





 $v_r = H_0 d + v_p$ Velocity distortions can be severe:

They can expand a cluster in redshift space in the radial direction five-ten times.









# Simulating structure formation:

J. Peacock, JENAM '04

Use a supercomputer to follow the trajectories of 10 million - 1 billion imaginary particles

$$F = \frac{G m m}{r^2}$$





The Virgo consortium uses Cray, SUN & IBM supercomputers (up to 512 processors) in Edinburgh, Durham & Munich to simulate the growth of cosmological structure



J. Peacock, JENAM '04

Forming superclusters (comoving view)

redshift z=3 (1/4 present size)

redshift z=1 (1/2 present size)

Redshift z=0 (today) The dominant feature of these maps, as of all other galaxy maps of the large-scale structure of the universe, is the network of filaments of different size and contrast, along with relatively empty voids between the filaments. The filamentary network contains different scales, where smaller-scale filaments are also less prominent.



Outline of the talk: • Galaxies everywhere (on the sky and in the computer) • Second-order statistics (correlation function and power spectrum)

"The central limit theorem asserts that a density distribution is asymptotically Gaussian in the limit where the density results from the average of many independent random processes; and a Gaussian is completely characterized by its mean and variance (the 1st and 2nd irreducible moments)".

A. Hamilton

# **Correlation Analysis**

## The two-point correlation function

Infinitesimal interpretation:

 $dP_{12} = \overline{n} [1 + \xi(\mathbf{r})] dV_1 dV_2$ 

is the joint probability that in each one of the two infinitesimal volumes  $dV_1$  and  $dV_2$ , with separation vector  $\mathbf{r}$ , lies a galaxy.

• is the average number density (intensity)







A 2-D slice of the segment Cox process



Random shifts are distributed according to a power-law density probability function  $d^*(r) \propto r^{\alpha}$ .



Random shifts are performed by a three-dimensional Gaussian distributed vector with  $\sigma = 0.5$ .

The two-point correlation function does not describe the filamentary pattern of the galaxy distribution:

Small random displacements do not destroy the filaments but erase the small scale correlations


### **Power spectrum**

#### Real space

Correlation function:

$$E\left\{\rho(\mathbf{x})\rho(\mathbf{x}+\mathbf{r})\right\} = \bar{\rho}^2 \left[1 + \xi(\mathbf{r})\right].$$

Density contrast:

$$\delta(\mathbf{x}) = \frac{\rho(\mathbf{x}) - \bar{\rho}}{\bar{\rho}},$$

 $E\left\{\delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{r})\right\} = \overline{\xi(\mathbf{r})}.$ 

### Fourier space (advantages)

- It is more intuitive physically, separating processes on different scales.
- Theoretical model predictions are made in terms of power spectrum.
- The amplitudes for different wavenumbers are statistically orthogonal

$$E\left\{ \widetilde{\delta}(\mathbf{k})\widetilde{\delta}^{\star}(\mathbf{k}') \right\} = (2\pi)^{3}\delta_{D}(\mathbf{k} - \mathbf{k}')P(\mathbf{k}).$$

is the Fourier amplitude of the overdensity field  $\delta$  at a wavenumber k

#### Power spectrum — correlation function:

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

Fourier transform pair

$$\xi(\mathbf{r}) = \int P(\mathbf{k})e^{-i\mathbf{k}\cdot\mathbf{r}} \frac{d^3k}{(2\pi)^3},$$

Isotropy:

$$\xi(r) = 4\pi \int_0^\infty P(k) \frac{\sin(kr)}{kr} \frac{k^2 \, dk}{(2\pi)^3}.$$

Delta-power:

$$\Delta^2(k) = \frac{1}{2\pi^2} P(k)k^3,$$



The density contrast power spectrum in a CDM cosmology

## 2dFGRS power-spectrum results

Dimensionless power:

d (fractional variance in density) / d ln k

Note no large oscillations: pure baryon universe disfavoured

Percival et al. MNRAS 327, 1279 (2001)

> J. Peacock, JENAM '04



Outline of the talk: Galaxies everywhere (on the sky and in the computer) • Second-order statistics (correlation function and power spectrum) (consistent estimation of the density field)

# Other clustering measures to describre the morphology beyond second order statistics:

- Higher order correlation functions
- Topology and other morphological descriptors provide information about the phase correlation of the density fluctuations in *k*-space
- Shape finders and integral transforms (MGA) are useful to describe the cosmic web

Sergei Shandarin's talk this morning

Minkowski Functionals Complete morphological description of scalar fields is given by Minkowski functionals (N + 1for *N*-dimensional space). We start for a:

- point distribution decorating the points with balls(*R*),
- continous distribution slicing by density isolevels.
- a

<sup>a</sup>Minkowski functionals: read K.R. Mecke, T. Buchert, H. Wagner, "Robust morphological measures for large-scale structure in the Universe", Astron. Astrophys. 288, 697-704 (1994).

### Minkowski functionals

Mecke, Buchert, and Wagner (1994), A&A, 288, 697



#### The germ-grain

model

 $A_r = \bigcup_{i=1}^N B_r(\mathbf{x}_i)$  for the diagnostic parameter r, where  $\{\mathbf{x}_i\}_{i=1}^N$  represents the galaxy positions and  $B_r(\mathbf{x}_i)$  is a ball of radius r centered at point  $\mathbf{x}_i$ . In  $\mathbb{R}^3$  there are four functionals: the volume V, the surface area A, the integral mean curvature H, and the Euler-Poincaré characteristic  $\chi$ , related with the genus of the boundary of  $A_r$  by

 $\chi = 1 - g.$ 



Kerscher & Martínez (1998), Bull. Int. Statist. Inst. 57-2, 363

![](_page_46_Picture_0.jpeg)

FIG. 1.— Spatial distribution of the low- (1) Auting Zret al. 2004 2.— Spatial distribution of the low- (left column) and high density (right column) regions for a realization of a Gaussian random field, with comparatively little smoothing. The upper pair shows the 7% low, 93% high density regions, the middle pair stands for 50%–50%, and the lower pair shows the 93% low-density, 7% high-density case.

case.

### Gaussian predictions for densities:

• 
$$v_0(\nu) = \frac{1}{2} - \frac{1}{2} \Phi\left(\frac{\nu}{\sqrt{2}}\right)$$
,  
•  $v_1(\nu) = \frac{2}{3} \frac{\lambda}{\sqrt{2\pi}} \exp\left(-\frac{\nu^2}{2}\right)$ ,  
•  $v_2(\nu) = \frac{2}{3} \frac{\lambda^2}{\sqrt{2\pi}} \exp\left(-\frac{\nu^2}{2}\right)$ ,  
•  $v_3(\nu) = \frac{\lambda^3}{\sqrt{2\pi}} (\nu^2 - 1) \exp\left(-\frac{\nu^2}{2}\right)$ ,  
• where  $\lambda^2 = \frac{1}{2\pi} \frac{-\xi''(0)}{\xi(0)}$ .

#### Topological genus (Euler-Poincaré characteristic)

![](_page_48_Figure_2.jpeg)

FIG. 3.— The average genus curve for 50 realizations of a Gaussian random field with  $P(k) = \sim k^{-1}$  together with the expected analytical result (solid line). Error bars are 1  $\sigma$  deviations.

Minkowski Lefunction the excursion of a more set of a points where  $\phi(\mathbf{x}) \geq \phi$ ). For a 3-D field the Minkowski functionals are: •  $V_0(\phi) = \int_{F_{\phi}} d^3x$  (volume), •  $V_1(\phi) = \frac{1}{6} \int_{\delta F_{\phi}} dS(\mathbf{x})$  (surface area), •  $V_2(\phi) = \frac{1}{6\pi} \int_{\delta F_{\phi}} \left( \frac{1}{R_1(\mathbf{x})} + \frac{1}{R_2(\mathbf{x})} \right) dS(\mathbf{x})$ 

(integrated mean curvature of the boundary),

• 
$$V_3(\phi) = \frac{1}{4\pi} \int_{\delta F_{\phi}} \frac{1}{R_1(\mathbf{x})R_2(\mathbf{x})} dS(\mathbf{x})$$
 (integrated Gaussian curvature of the boundary).

•  $V_3 = \chi = (1 - g) =$ # of isolated regions - # of holes (g is the topological genus).

Arguments:

- density deviation  $\nu_{\sigma} = \frac{\phi_c \bar{\phi}}{\sigma_{\phi}}$ ,
- Gaussianized volume fraction  $\nu_G$ :  $v_f = 1 - \frac{1}{\sqrt{2\pi}} \int_{\nu_G}^{\infty} \exp(-x^2/2) dx.$

Lattice algorithm (Kendrick invariants): a lattice of a step a, with N vertices. In the excursion set:  $N_0$  vertices,  $N_1$  segments,  $N_2$ faces,  $N_3$  cells.

![](_page_51_Figure_1.jpeg)

### Smoothing

$$\rho(\mathbf{x}) = \frac{1}{N} \sum_{i}^{N} K_h\left(\frac{\mathbf{x} - \mathbf{x}_i}{h}\right),$$

where  $K(\mathbf{x})$  is a distribution function:

$$\int K(\mathbf{x}) \, d\mathbf{x} = 1, \qquad \int \mathbf{x} K(\mathbf{x}) \, d\mathbf{x} = 0.$$

#### To choose of the band with h is an art!

### Techniques:

### • sample point, sandbox, gather estimators: $h = h(\mathbf{x}_i)$

• balloon, scatter estimators:

$$h = h(\mathbf{x})$$

### • choose *h* to minimize $MSE = E[\tilde{f}(\mathbf{x}) - f(\mathbf{x})]^2 = Var(\tilde{f}(\mathbf{x})) + [Bias(\tilde{f}(\mathbf{x}))]^2$

## The à trous wavelet transform

One-dimensional signals s(l):

Decomposition (and reconstruction)

Already explained in Jean-Luc's talk

$$s(l) = c_{0,l} = c_{J,l} + \sum_{j=1}^{J} w_{j,l}.$$

The scaling function

$$\frac{1}{2}\phi(\frac{x}{2}) = \sum_{k} h(k)\phi(x-k).$$

Define the wavelet:

$$\frac{1}{2}\psi(\frac{x}{2}) = \phi(x) - \frac{1}{2}\phi(\frac{x}{2}).$$

![](_page_55_Figure_0.jpeg)

$$w_{j+1,l} = c_{j,l} - c_{j+1,l}.$$

#### The scaling function

$$\phi(x, y, z) = \phi(x)\phi(y)\phi(z),$$

decomposition (and reconstruction)

$$s(l,m,n) = c_{0;l,m,n} = c_{J;l,m,n} + \sum_{j=1}^{J} w_{j;l,m,n}.$$

• Smooth:

$$c_{j+1;l,m,n} = \sum_{i} \sum_{k} h(i)h(k)h(o)c_{j;l+2^{j}i,m+2^{j}k,n+2^{j}o}$$

J data sets for all dimensions.

## The $B_3$ spline

![](_page_57_Figure_1.jpeg)

![](_page_57_Figure_2.jpeg)

![](_page_58_Figure_0.jpeg)

![](_page_59_Picture_0.jpeg)

Data – NGC 2997

![](_page_60_Picture_0.jpeg)

À trous transform of the NGC 2997 data

![](_page_61_Picture_0.jpeg)

À trous transform, wavelet amplitudes

# PixInsight

• A comercial software made and tested by amateur astronomers

![](_page_63_Figure_0.jpeg)

![](_page_64_Figure_0.jpeg)

Releiades PixInsight 1.0 File Edit Project Image Preview Mask Process Tools Windo	W Help			
□       □       □       □       □       □       □       □       ■ RGB       ▼         □       ■       ■       ■       □       □       ■<	사업╫수☆ヸ゠゠゚゚゚ヿヿ゚゚゚゚ヷ゙゙゙゙゙゙゙ヷ゚゚゚ヹ゚゚゚゚ヹ゚゚ヹ゚゚ヹ゚゚゚゚゚゚゚ヹ゚゚゚゚゚゚゚゚			
RGB 2:1 Image01->Preview01 E:\home\PixInsight\seminario	Process Container         V 1       Image/dentifier       2044 out 52 82 0.0017       0.02       gustamenters         V 13       Image/dentifier       2044 out 52 82 0.0017       0.02       gustamenters         V 13       Image/dentifier       2044 out 52 82 0.0017       0.02       gustamenters         V 14       Convolution       2044 out 52 82 0.0017       0.02       gustament-tsignen10.00.orde=6.00, resamply         V 15       HistogramsTransform       2044 out 52 82 0.0017       0.016       ////////////////////////////////////			
Processing planetary images with <i>à trous</i> wavelet transforms.				
x186.00 · y:107.50	w:180 · h:180 · BGB(global) · 0.371 MB			
🎒 Inicio 👔 📽 😂 📽 🗞 🚥 ۂ 🔣 🔡 🔛 Philasight 1.0				

![](_page_66_Picture_0.jpeg)

**3.** Large dimensional scales enhanced with à trous wavelet transform. Stars and small-scale structures re-inserted after wavelets processing.

Barnard 142/143 Region DSS Image Processing in PixInsight by Vicent Peris (PTeam) removed in à trous wavelet

![](_page_66_Picture_4.jpeg)

![](_page_66_Picture_5.jpeg)

![](_page_66_Picture_6.jpeg)

![](_page_66_Picture_7.jpeg)

#### NGC 7662 image by Vicent Peris (PTeam) / Processing in PixInsight TROBAR 0.6 m Ritchey-Chretien Telescope @ f/25 / Canon 300D digital SLR camera

![](_page_67_Picture_1.jpeg)

1. Original RGB image

![](_page_67_Picture_3.jpeg)

**2.** Small-scale noise reduction with *à trous* wavelet transform

🕌 À Trous Wavelets				
Wavelet	Layers	Scaling Function		
Level	Scale	Parameters	Preview	
× 1	1		,	
× 2	2			
✓ 3	3	S:(0.79,1,5)		
🗸 B	8258			
			Reset	
◯ Dyadic 🔹 Linear: 1 🚦 Count: 3 💌 🔽 Use luminance				
Detail L	ayer 3/3		Reset	
Bias: 0.000				
Noise Reduction				
Amount: 0.79 n: 1 v k: 5 v				
Deringing				
Amo	ount: 0.0	D / <del></del>		
Thres	hold: 0.0	50 · /		

Suppression of small-scale wavelet layers. Adaptive noise reduction on the third wavelet layer.

![](_page_67_Picture_7.jpeg)

**3.** Regularized Richardson-Lucy deconvolution / Color enhancement processing.

Comparison with HST image.

•PixInsight is a high-performance, flexible and modular image processing software specialized in astrophotography.

•It consist of a low-level processing engine, a sophisticated graphical user interface, and a programming environment to build high-level processing tools and their associated user interfaces.

•A limited edition of PixInsight (LE) has been released as a freeware application.

•PixInsight Home Page: <a href="http://pleiades-astrophoto.com/pixinsight/">http://pleiades-astrophoto.com/pixinsight/</a>

### Wavelet smoothing

![](_page_69_Figure_1.jpeg)

van de Weygaert talk also dealed with an adaptive method to find the density field using Delaunay tessellations

![](_page_70_Picture_0.jpeg)

Á trous density (slices) for a Voronoi walls sample (square root scale).

#### The multifesolution world

![](_page_71_Picture_1.jpeg)

Á trous density slices for an N-body model structure (logarithmic scale).


Á trous potential (slices) for the N-body model above (linear scale).



#### N- body simulation

Voronoi model – populating filaments

Swiss-Cheese model dominated by huge voids

About 15000 points







The wavelet density field keeps information at all scales due to its multiscale nature, and represents an unique reconstruction, which does not depend on a given band width.

## Wavelet cleaning





 $\begin{array}{l} \mbox{Left-Gaussian } \sigma = 1 \ \mbox{Mpc smoothing, middle-wavelet cleaning,} \\ \mbox{right-Gaussian } \sigma = 3 \ \mbox{Mpc smoothing.} \end{array}$ 



Wavelet cleaning of a model galaxy distribution (left – a successful attempt, right – overcleaning).

Denosing technique explained in Jean-Luc's talk **Ringing artifacts** 

#### Gaussian smoothing (n body)







#### Wavelet smoothing (n-body)









The genus curve of this adaptive reconstructed density field is much more informative because it is unique and does not depend of the particular choice of the filter radius. Additionally, the genus curves of Gaussian-smoothed density fields mimic those of Gaussian random fields, describing thus more the properties of the filter than the real morphology of the density distribution.

#### Voronoi filaments









#### Swiss Cheese









#### 2dfGRS northern slice













We have seen that Gaussian smoothing, as usually applied, introduces a strong Gaussian signal for any kind of distribution, erasing small scale non-Gaussian features

Coles & Lucchin pointed out this idea already in 1995

We see that the multiscale representation retains the non-Gaussian nature of the initial samples; the Minkowski functionals are far from Gaussian

In contrast, Gaussian smoothing smears matter into low-density regions and generates density distributions, which resemble Gaussian random fields, as shown by the Minkowski functionals shown in the previous sections.

### Conclusions

- Quantitative descriptors -being reliable, robust, unbiased, and physically interpretable- are needed to extract cosmological information from the data.
- We have shown a consistent method to estimate the density field in an optimal way that ensure that the topology of the reconstructed field reflects the true underlying topology of the point process.
- The method is suitable for the galaxy distribution which has a genuinely multiscale structure, with non-linear structures like filaments, sheets and prominent clusters.

# If you want to know more...

