Astrophysical Data Analysis and Visualization Toolkit

Daniela Ferro
Dipartimento di Fisica e Astronomia, via S.Sofia, 78,
95100, Catania, ITALY
woac.ct.astro.it
dferro@ct.astro.it

Ugo Becciani, Vincenzo Antonuccio-Delogu and Angela Germaná
INAF, via S.Sofia, 78,
95100, Catania, ITALY
woac.ct.astro.it
ube@ct.astro.it,van@ct.astro.it,age@ct.astro.it

Claudio Gheller
CINECA, via Magnanelli, 6-3,
40100, Casalecchio di Reno (BO), ITALY
www.cineca.it
c.gheller@cineca.it

Abstract

AstroMD is a visualization software offering several capabilities specifically oriented to the cosmological analysis of three-dimensional structures. It was developed within the framework of the Cosmo.Lab project, financially supported by the European Community, which involves several European Astrophysical Institutions and the CINECA. This tool gives a 3D graphic representation of data, exploiting the most advanced visualization technology based on virtual reality, and has several built-in-tools which allow the user an efficient manipulation and analysis of data, in order to build a leading edge instrument for scientific research. It was developed using the Visualization Toolkit (VTK) by Kitware, a freely available visualization library portable on many platforms. AstroMD is an open source freely available code at the address http://cosmolab.cineca.it/.

Keywords: Scientific Visualization, Astrophysical Data Analysis

1. INTRODUCTION

Since the beginning of modern astronomy, the scientific community has expressed a great interest in scientific visualization tools. The improvement of technology and the availability of super-computing multiprocessor systems has led to a dramatic increase in the volume of data coming from observations of astronomical sources and from cosmological simulations. Today, numerical simulations produce many gigabytes of data which have to be efficiently visualized and analyzed. Visualization is the most intuitive approach to data and basic information can be obtained just *at a glance*. Then the possibility of moving inside the data allows the scientist to focus on regions of interest and there to perform quantitative calculations.

AstroMD provides an effective and intuitive way of managing and analysing large amounts of astrophysical data. It can find structures having a not well defined shape or symmetry, and perform quantitative calculations on a selected region or structure. The most advanced visualization techniques are exploited, in particular stereographics visualization and virtual reality.

AstroMD also contains functionalities for the calculation of quantities of cosmological interest, such as the *two-point Correlation Function*, the calculation of the *Gravitational Field*, the *Power Spectrum* of the density fluctuations, the *Minkowski Functionals* and includes the *Friend-of-friend* group finding algorithm. The most common data formats used in cosmological simulations (i.e.
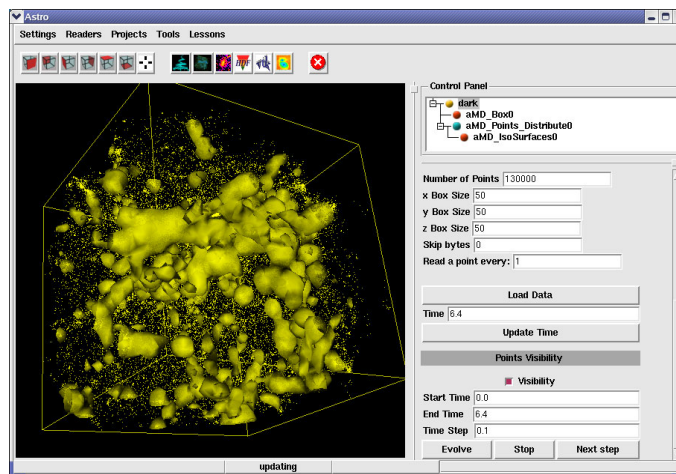
FIGURE 1: The main Graphic User Interface of AstroMD allows the user to read and manipulate different data format files. Isosurfaces characterize regions with density over the user-defined threshold.

Tipsy,*HDF5*) and in observational fields (i.e. *Fits*, *ASCII*) are loaded by AstroMD. The common unformatted *C* standard, written in a continuous sequence of *x*, *y*, *z* coordinates followed by any scalar field, is also accepted by AstroMD.

The implementation of the software was mainly made in the framework of the two-year Cosmo.Lab project, financially supported by the European Community. The project involves the Astrophysical Observatory of Catania (*INAF*), the Institute of Radio Astronomy of Bologna (*IRA*), the Institute of Cosmic Physics G.Occhialini of Milan (*IFCTR*), the Netherlands Foundation for Research in Astronomy (*NFRA*) and the *CINECA*.

2. THE SOFTWARE

In order to build a widely used product it was chosen to use a low cost software, the Visualization Toolkit (*VTK*) by Kitware [1, 2, 3]. *VTK* is an open source, freely available and portable (runs on Windows, Macintosh, and Unix platforms) software system for *3D* computer graphics, scientific visualization and image processing. The design and implementation of the library is based on the Object-Oriented paradigm. It consists of two basic subsystems: a compiled *C++* class library and an interpreted wrapper layer for manipulating the compiled classes. Data analysis functionalities inside AstroMD consist in efficient (in both CPU and memory) algorithms, specifically built and inserted in the *VTK* architecture. These built-in functionalities are controlled by a specific Graphic User Interface (see Figure 1), written in *incrTcl/Tk* [4, 5] that supports the Object-Oriented programming structure. *VTK* supports a wide variety of visualization algorithms including scalar, vector, tensor, texture and volumetric methods and advanced modelling techniques. It supports stereographic rendering and can be used for virtual reality visualization. The stereographic capabilities of AstroMD are presently implemented at the Virtual Theatre of CINECA [6].

2.1. Basic Functionalities

AstroMD allows the user to treat both particles (*unstructured data*) and continuous fields discretely represented over a computational mesh (*structured data*). The whole set of particles can be visualized but it is also possible to use a statistically significative sampling of the original data in order to get a faster and easier visualization. Data are visualized compared to a cubic box which describes the computational region. A cubic or spherical sub-region can be interactively selected inside the parent box with a different spatial resolution, in order to focus on the most interesting regions. Data inside the sampler can be studied with the analysed tools or can be saved in specific files for an off-line analysis. Boxes can be translated, rotated, zoomed in and out compared to selected positions. Colours and luminosities can be chosen by the user. The opacity of the particles can be increased, low density regions are more easily detectable, or decreased, and the details of the high density regions substructures are shown. Different particles species
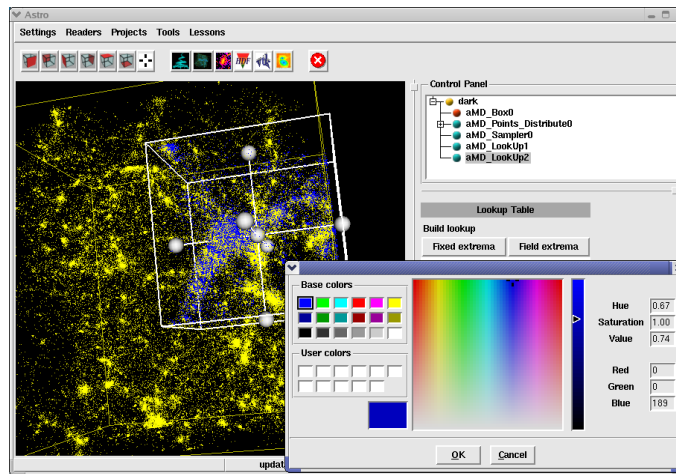
FIGURE 2: The look-up table associates scalar values of a field with a colour map. This image visualizes the course of the temperature in a galaxy.

(e.g. dark matter and baryons) can be visualized at the same time using different colours. A look-up table allows the user to associate scalar values of a field with a colour map (see figure 2). Other particle-related continuous quantities, like density fields, can be calculated as typical grid-based fields and visualized as isosurfaces or volumes. Scalar fields can be visualized with isosurfaces or with volume rendering. The volume rendering can be calculated using both the texture mapping and the ray tracing technique, depending on the specific hardware used. Different time frames can be shown in a sequence. When the particle representation is used, the position of particles is interpolated between one time-step and the following. This improves the quality of the *animation* giving a fluidity otherwise unachievable. Both the single images and the whole sequence can be saved in bitmap or jpeg format to prepare an animation of the evolution. Enabling the steady-cam, the system can be rotated in Azimuth and Elevation during its evolution. Zoom-in and zoom-out possibilities are also offered.

3. DATA ANALYSIS FUNCTIONALITIES

With the aid of the visualization, AstroMD allows the user to do several types of analysis to observational or simulated data. The scientist can quickly select the region of interest, using the sub-region selector of AstroMD, and the analysis modules can be applied to one. Moreover the analysis is done on *click-of-mouse*: the results are displayed in a few seconds.
The main following implemented functionalities are specifically directed at cosmological simulation data outputs. Some of them are statistical tools that one can deploy in order to discriminate among different cosmological models.

Particles Mass Density - By using an *eight points Cloud-in-Cell* smoothing algorithm [7], AstroMD distributes the mass of each particle over the computational mesh and calculates the mass density field $\rho(\mathbf{x})$associated to the particle distribution. The computation can be done with the maximum accuracy, that is by considering all the particles over a uniform high resolution mesh. The user can also use a sample of the whole set of particles thus reducing the CPU time consumption and the memory request. The smoothing of the masses can be generally performed by using a coarse grid that can be refined where high resolution is necessary. The same tool can be used to calculate other fields related to quantities associated to particles like, for example, the thermal energy density field or the X-ray luminosity field.

Gravitational Field Calculation - By using the Fourier Transform procedure it is possible to calculate the gravitational field with Poisson's equation:

$$\phi(\mathbf{x}) \propto \nabla^2 \rho(\mathbf{x}), \tag{1}$$

where $\phi(\mathbf{x})$ is the gravitational potential and $\rho(\mathbf{x})$ is the mass density defined over the computational mesh as above. If Poisson's equation is represented in terms of wave vectors, the equation is reduced to a simpler algebric operation:

$$\phi(\mathbf{k}) \propto \frac{1}{\mid \mathbf{k}^2 \mid}\rho(\mathbf{k}), \tag{2}$$

where $\phi(\mathbf{k})$ and $\rho(\mathbf{k})$ are the Fourier images of the potential and of the density respectively, and $\mid \mathbf{k}^2 \mid$ is the square module of the wavenumber. Finally, the potential is transformed into physical space using an inverse *FFT* [8].

Fourier decomposition, Power Spectrum and Correlation Function - The fluctuation of density $\delta(\mathbf{x}) = [\rho(\mathbf{x}) - \langle\rho\rangle]/\langle\rho\rangle$ is expressible as a Fourier series:

$$\delta(\mathbf{x}) = \sum_{\mathbf{k}} \delta_{\mathbf{k}}\mathbf{exp}(\mathbf{ik} \cdot \mathbf{x}) = \sum_{\mathbf{k}} \delta_{\mathbf{k}}^{*}\mathbf{exp}(-\mathbf{ik} \cdot \mathbf{x}). \tag{3}$$

For a statistically homogeneous and isotropic density field, the variance across the statistical ensemble is:

$$\sigma^2 = \langle\delta^2\rangle = \frac{1}{V_u}\sum_{\mathbf{k}} \delta_k^2 = \frac{1}{2\pi^2}\int_0^{\infty} P(k)k^2 dk, \tag{4}$$

where, in the limit $V_u \to \infty$, the quantity $\delta_k^2 = P(k)$ is called Power Spectrum [9].
The covariance function can be defined in terms of density field $\rho(\mathbf{x})$:

$$\xi(r) = \frac{\langle[\rho(\mathbf{x}) - \langle\rho\rangle][\rho(\mathbf{x} + \mathbf{r}) - \langle\rho\rangle]\rangle}{\langle\rho\rangle^2} = \langle\delta(\mathbf{x})\delta(\mathbf{x} + \mathbf{r})\rangle \tag{5}$$

and, passing to the limit $V_u \to \infty$, we can see that the Power Spectrum is the Fourier Transform of the Covariance Function (*Wiener-Khintchine theorem*):

$$P(k) = \int \xi(\mathbf{r})exp(i\mathbf{k} \cdot \mathbf{r})d\mathbf{r}. \tag{6}$$

Six modules were specifically implemented for the computation of the two-point Correlation Function and the Power Spectrum of a point distribution. The two-point Correlation Function of AstroMD is based on the three-dimensional counterpart of the *Peebles & Hauser* estimator (1974):

$$\xi_{PH} = \frac{DD(r)}{RR(r)}(\frac{N_{rd}}{N}) - 1, \tag{7}$$

where N_{rd} is the number of points of an auxiliary random sample, $DD(r)$ is the number of all pairs of points with separation inside the interval $[r - dr/2, r + dr/2]$, and $RR(r)$ is the number of pairs between the data and the random sample with separation in the same interval. For a comparison among most estimators of this statistical quantity, see [10]. The calculation of RR and DD is performed through Monte Carlo integration.
AstroMD allows the user to generate the auxiliary random sample and to extract one or more sub-samplers where the estimator is computed. The method is generally applied to a subset of particles with few thousands of particles. Figures 3 and 4 are about some results of a ΛCDM model with $\Omega_\Lambda = 0.7$ and initial redshift $z = 10$. The spatial distributions shown on top of these figures regard the visualization of the output data from the FLY code [11], for a box of size $50h^{-1}$ Mpc and with 150000 bodies, randomly extracted from a file with 256^3 particles. Using the *AstroMD Sample Tool*, the user can select a specific cubic or spherical region where the function is computed. X-Y plotting data (Figures 3 and 4, in middle) can be saved in a text file for an off-line analysis. Figures 3 and 4 (bottom) represent an example to estimate the exponent γ, that characterizes the power-law shape of $\xi(r)$, with external functionalities.
The code for the calculation of the Power Spectrum gets the coordinates of a point set from a cosmological simulation or from a catalogue as input, and computes the limits of the bounding box. The user can define the number of binning intervals.
Figures 6 and 7 show the graphical outputs of the power spectral density of a point distribution for
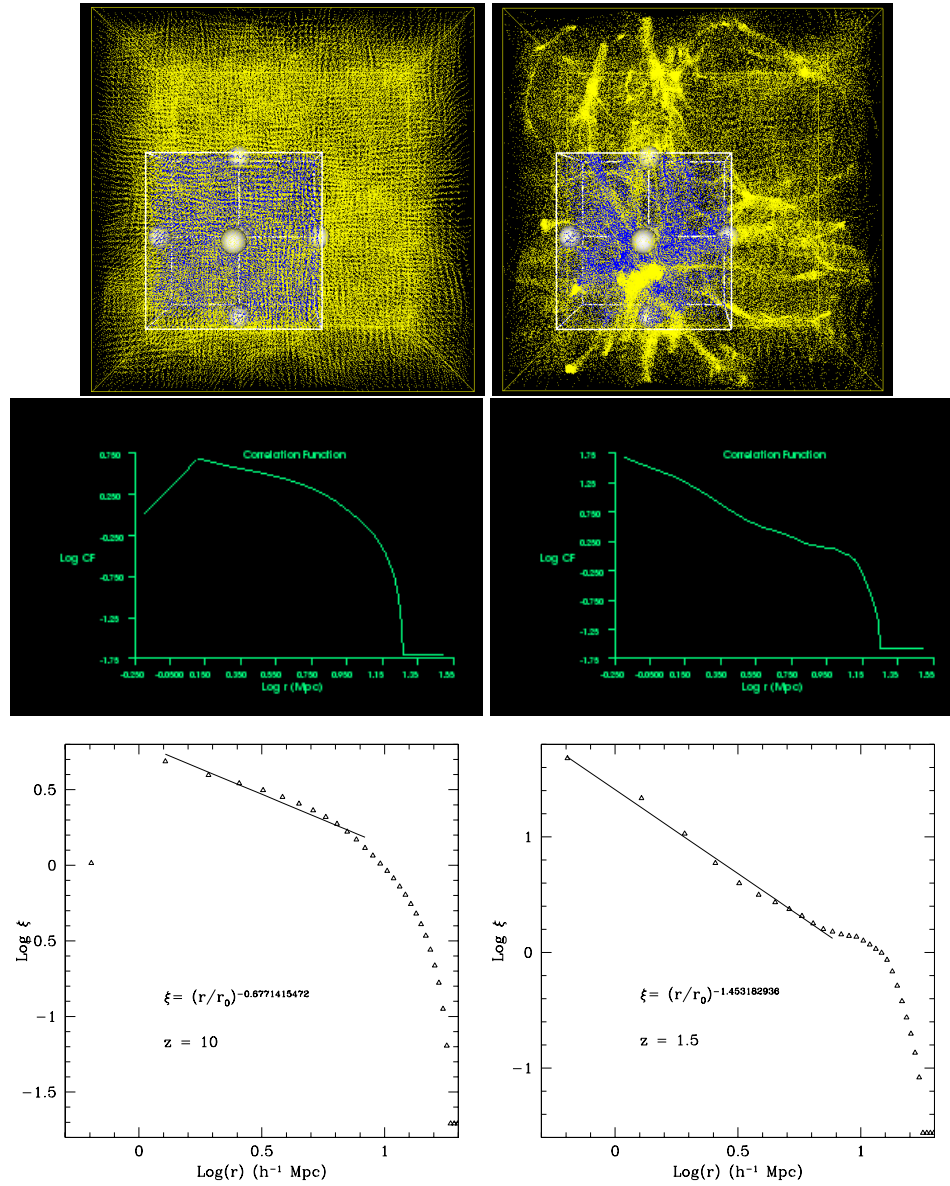
FIGURE 3: Some applications of the analysis tools of AstroMD to output data from FLY code [11] at different redshifts. From left to right, $z = 10$ and $z = 1.5$. X-Y plotting data can be studied with external analysis tools. Bottom figures show an approach for characterizing the slope of the curves.
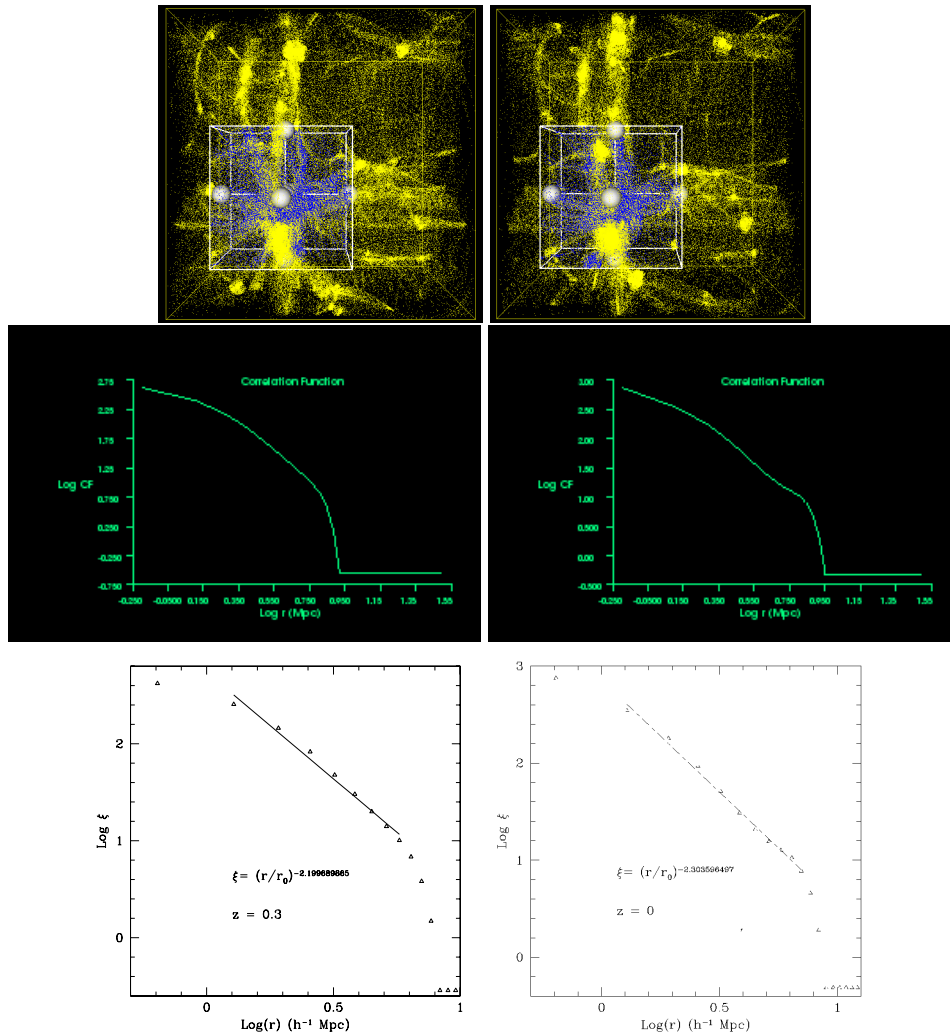
FIGURE 4: Examples of the two-point Correlation Function obtained at different redshifts for a ΛCDM simulation model.
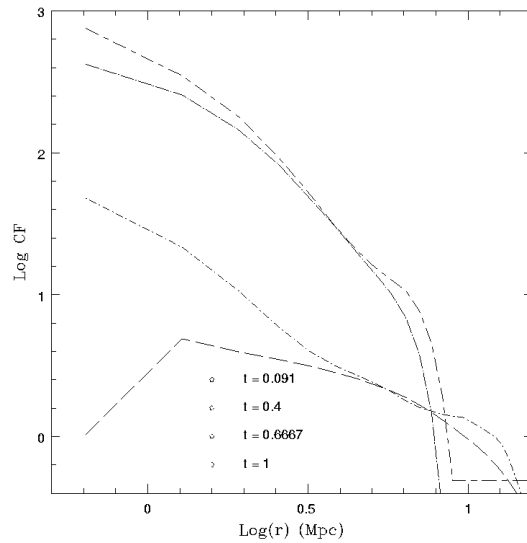

FIGURE 5: Composition of the two-point Correlation Function in a logarithmic plot for the whole box (size $50h^{-1}$Mpc) at different redshifts for the same cosmological model. The unit of time t, in this case, is a fraction of the Hubble time.
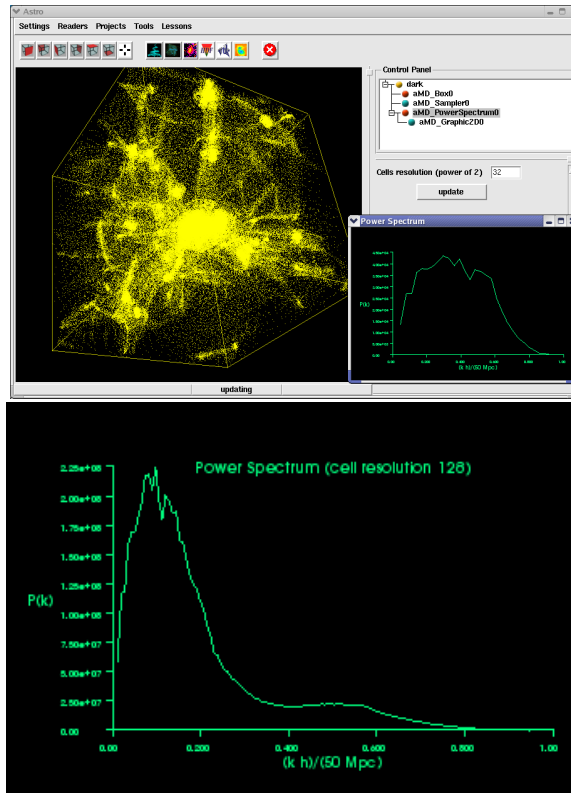
FIGURE 6: Examples of the computation of the Power Spectral Density (*PDS*) with AstroMD, applied to data by FLY [11] at a redshift of about zero, for different cell resolution values. The spectral index is approximately 2 up to the turn-over, so power spectrum converges at large scale (small k). After the turn-over the *PDS* decades as a power law with an exponent of approximately -4.
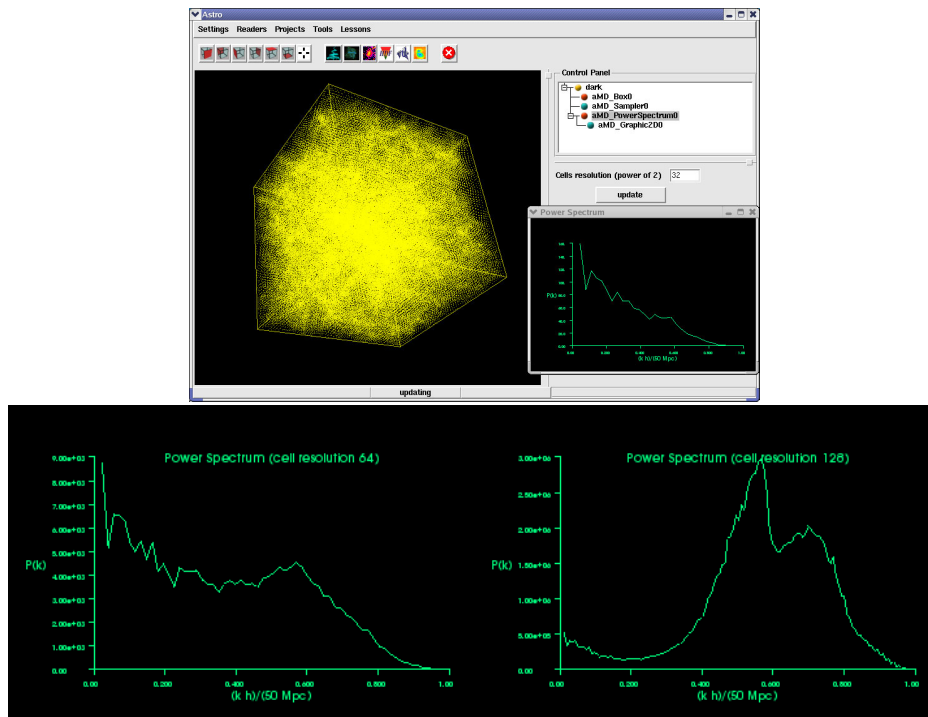

FIGURE 7: Graphical outputs of the Power Spectrum of data from FLY at a redshift of approximately 10.

TABLE 1: Geometric interpretation of the Minkowski Functionals. The first Functional equals the volume V of the body, the second one is the surface area A. The third Functional corresponds to the integral mean curvature H of the body's surface and provides information about the shape.

geometric quantity	μ	M_μ	V_μ
V volume	0	V	V
A surface	1	$A/8$	$A/6$
H mean curvature	2	$H/2\pi^2$	$H/3\pi$
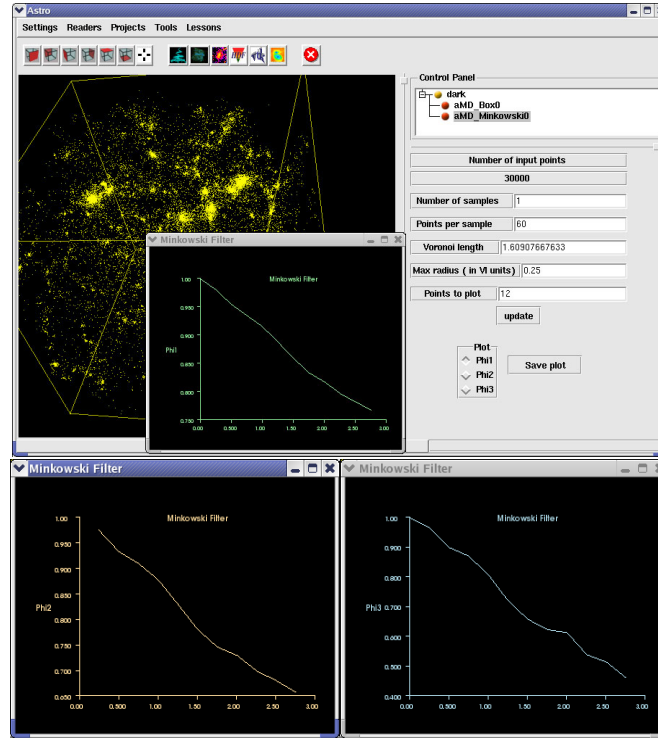
FIGURE 8: Reduced numerical Minkowski Functional Φ_1 (top), Φ_2 and Φ_3 (bottom), obtained with AstroMD, for a ΛCDM simulation at $z = 0$, in a cubic box with side length 50 Mpc with 150000 particles, plotted against the radius of balls measured in Mpc unit.

a sub-set with 262144 particles, extracted from a dataset of 256^3 bodies produced by FLY [11]. It was considered the same cosmological model used before at two different redshifts.

Minkowski Functionals The Minkowski Functionals (*MF*) provide a novel tool to characterize the Large Scale Structure of the Universe. They describe the Geometry, the Curvature and the Topology of a point-set [12].

In a three-dimensional Euclidean space, these functionals have a direct geometric interpretation as listed in Table 1 [13]. The *MF* algorithm inside AstroMD associates a ball of radius r to each point of the point distribution. The size, the shape and the connectivity of the spatial pattern, formed by the union-set of these balls, change with the radius, which can be employed as a diagnostic parameter [13]. In particular, astroMD computes the reduced values of the Minkowski Functionals, $\Phi(\mu)$ with $\mu = 1, 2, 3$, that are the ratio of *MF* of the actual distribution to the *MF* of the same number of disjoint convex bodies. So their values start always from unity because, for small radii, all covering balls are disjoint. In the 3-rd functional $\Phi_3(r)$, the first zero provides an estimate of the percolation threshold [14]. A spongy structure like a Poisson distribution gives lower values for $\Phi_3(r)$, while higher values indicate structures with few but big filaments or tunnels.

Figures 9 (top) compare the Minkowski Functionals, computed by AstroMD, for an uniform ($z = 50$) and a clustered ($z = 0$) point distribution. Two cubic samples with 1251 points have been extracted from the whole simulation, while the uniform one has 1070 points.
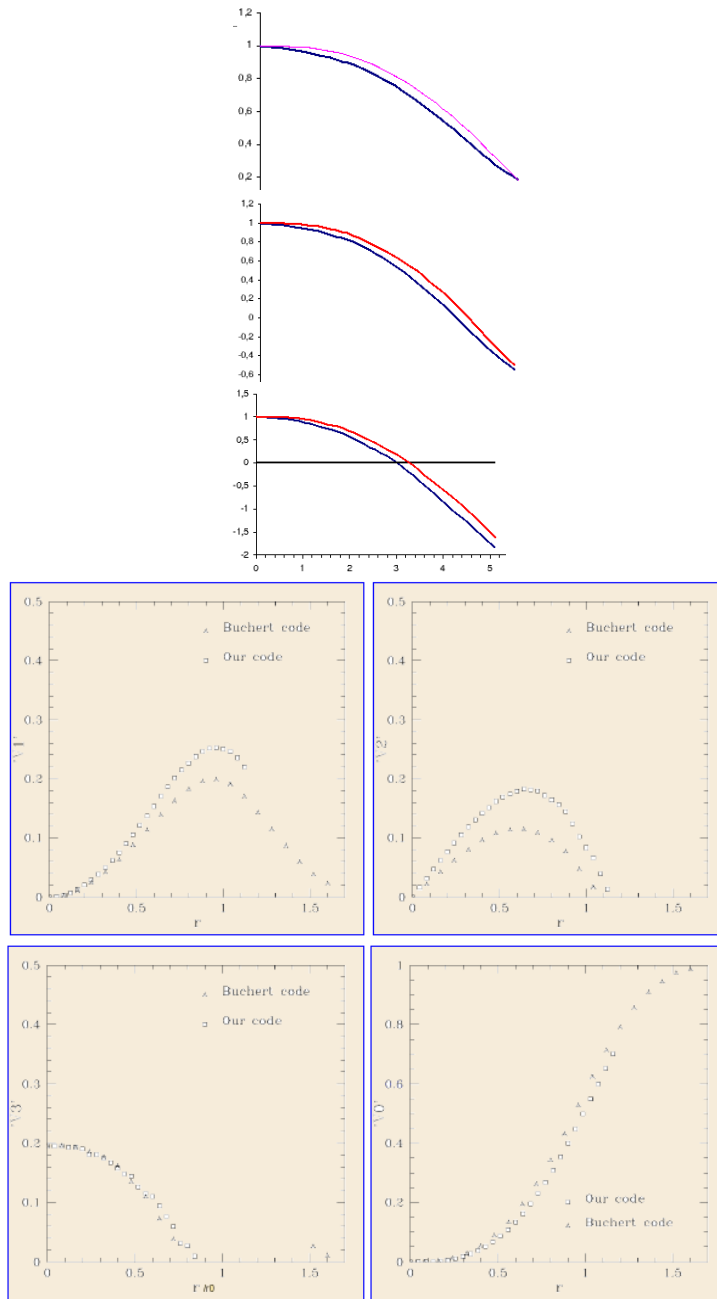
FIGURE 9: (Top) The MF computed by AstroMD, for a uniform ($z = 50$) and a clustered ($z = 0$) point distribution (upper line). (Bottom) The graphical outputs of *MF* with AstroMD are in good accordance with those of the Buchert code [12].
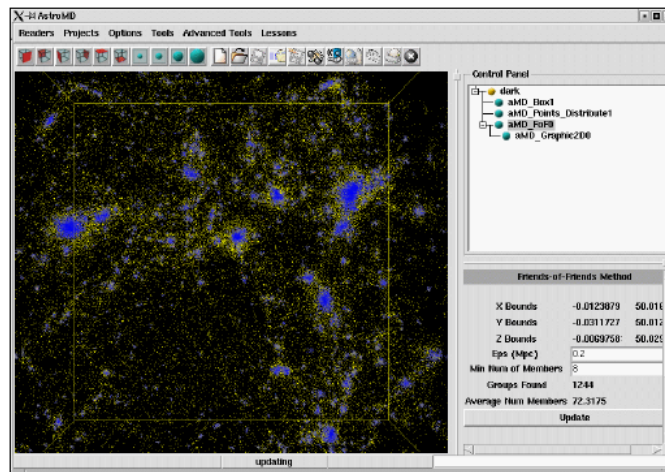
FIGURE 10: The *FoF* group finding algorithm inside AstroMD.

Friend-of-Friend Algorithm Dynamical studies of groups of galaxies are an important method for estimating galaxy masses. The most common group-finding algorithm is known as *Friend-of-Friend* (*FoF*). This technique was first used by Huchra & Geller [15].

A particle belongs to a *FoF* group if it lies within some linking length ϵ of any other particle in the group. The identification of groups does not depend on the Hubble constant H_0 used. If no companions are found, the group is entered in a list of *isolated* elements. All companions found are added to the list of group members. The surroundings of each companion are then searched. This process is repeated until no further members can be found. After all such groups are found, those with less than a specified minimum number of group members $num_{Members}$ are rejected.

AstroMD visualizes the positions of all grouped particles (each of them are marked with their group identifier) and the positions of all centers of mass, that are associated with the number of particles in the group (see figure 10); the radius, velocity (if present in input) and total mass are also visualized. We obtain two kinds of graphical output: the grouped particles with the center of mass of each group and, in another window, the plot of the fraction of grouped particles versus the number of components of the group, or the mass function (number of groups versus mass of group). *FoF Tool* runs more and more as the parameters (ϵ and $num_{Members}$) are changed by the user, and updates all its ouputs.

4. CONCLUSIONS

AstroMD represents the first experience of an immersive Visualisation and Data Analysis Tool in astrophysics. The AstroMD Graphic User Interface is an Object-Oriented interpreted language software and its architecture is based on the newest visualization and GUI technologies. The internal structure is based on the VTK library, written in C++, and all the analysis tools consist in specific modules that we have implemented in an *ad hoc* manner.

AstroMD is a valuable tool that allows the scientific groups to interact efficiently with large amounts of data, easily *navigating* inside them, calculating their statistical properties and reconstructing their three dimensional shapes and features, on *click-of-mouse*.

This leads to a deeper understanding of scientific problems and to their easier solution.

We would like to thank L.Santagati for English revision of the text.

REFERENCES

[1] Schroeder W, Martin K and Lorensen B. *The Visualization Toolkit* (2nd edn). Prentice Hall PTR.

[2] Schroeder W, Avila L, Martin K, Hoffman W and Law C. *The Visualization Toolkit. User's Guide.* Addison-Wesley.

[3]   *http://www.kitware.com/*

[4]   Welch B. *Practical Programming in Tcl and Tk.* Prentice Hall PTR.

[5]   *www.tcltk.com/itcl/.*

[6]   *http://www.cineca.it/HPSystems/Vis.I.T/.*

[7]   Hockney R and Eastwood J. *Computer Simulation using particles.* New York, Mac Graw Hill HE.

[8]   Demmel J. *Solving the Discrete Poisson Equation using Jacobi, SOR, Conjugate gradients and the FFT.* CS267 Lectures, 1996, 15-16.

[9]   Coles P and Lucchin F. *Cosmology. The Origin and Evolution of Cosmic Structure.* Wiley.

[10] Pons-Borderiá M, Martińez V, Stoyan D, Stoyan H and Saar E. *Comparing Estimators of the Galaxy Correlation Function.* ApJ 523 480.

[11] Antonuccio-Delogu V, Becciani U and Ferro D. *FLY. A parallel tree N-body code for cosmological simulations. Reference Guide.* astro-ph/0310786.

[12] Platzöder M, and Buchert T. *Applications of Minkowski-functionals to the Statistical Analysis of Dark Matter Models.* astro-ph/9509014.

[13] Schmalzing J, Kerscher M and Buchert T. *Minkowski Functionals in Cosmology.* astro-ph/9508154.

[14] Mecke K and Wagner H. 1991, J. Stat. Phys., 64, 843.

[15] Huchra J and Geller M. *Groups of Galaxies I. Nearby Groups.* ApJ 257 423.