Galaxy and cluster clustering

ongoing activities and general numerical issues

Federico Marulli
Department of Physics and Astronomy
University of Bologna

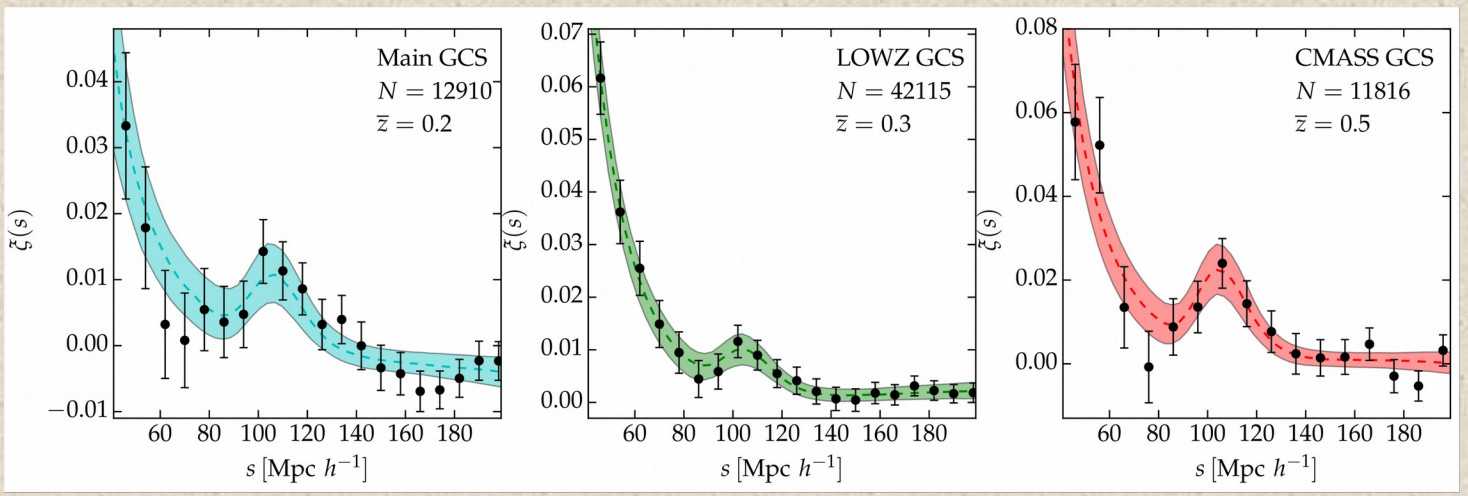
in collaboration with Enzo Branchini
and the Bologna Clustering Group
(Aricò, Cannarozzo, Cimatti, Gaspari, Moscardini, Moresco, Ronconi, Veropalumbo)

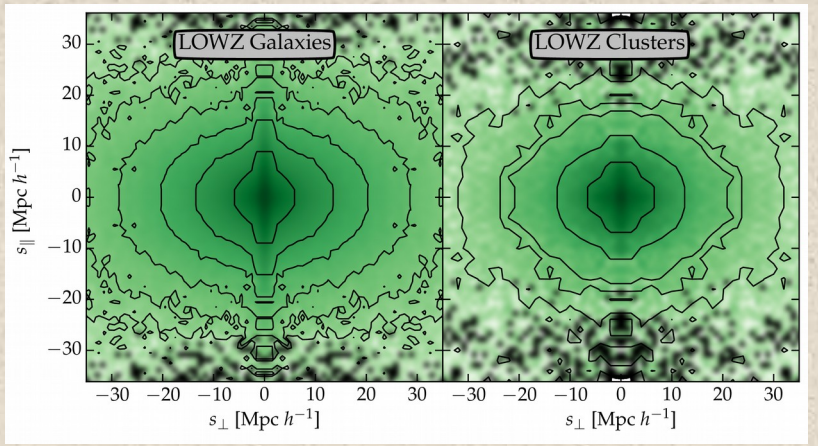
Clustering in Italy

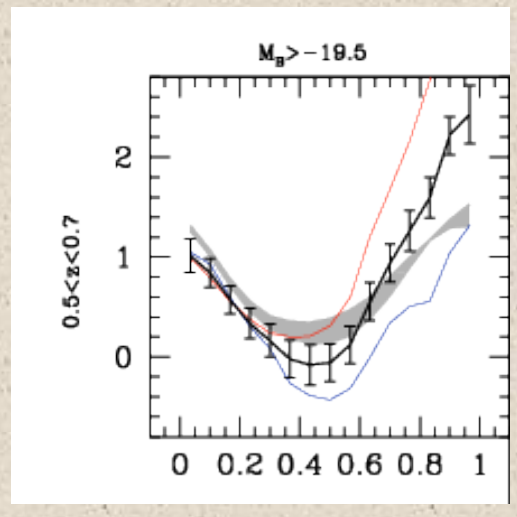
- An increasingly large community in Italy is working on clustering
- This interest has been triggered mainly by the large galaxy surveys in which Italian researchers are involved, such as VIPERS (PI Guzzo), and planned future missions, such as Euclid
- Cosmology with clustering: clustering shape, baryon acoustic oscillations, redshift-space distortions, Alcock-Paczynski test, higher-order statistics, etc.
- Galaxy evolution with clustering: constraints on semi-analytic models, bias, clustering as a function of tracer properties, AGN hosts and lifetime, etc.

2-point and 3-point correlation functions

Veropalumbo, Marulli et al. 2014-2016 ; Moresco, Marulli et al. 2016

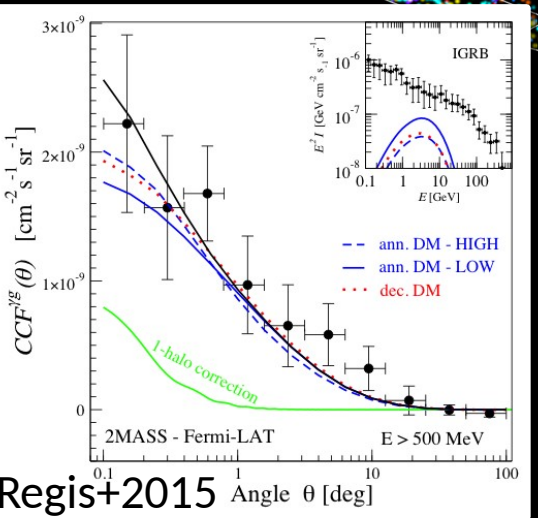






X-correlation with LSS: e.g. Fermi γ-ray background vs. 2MASS

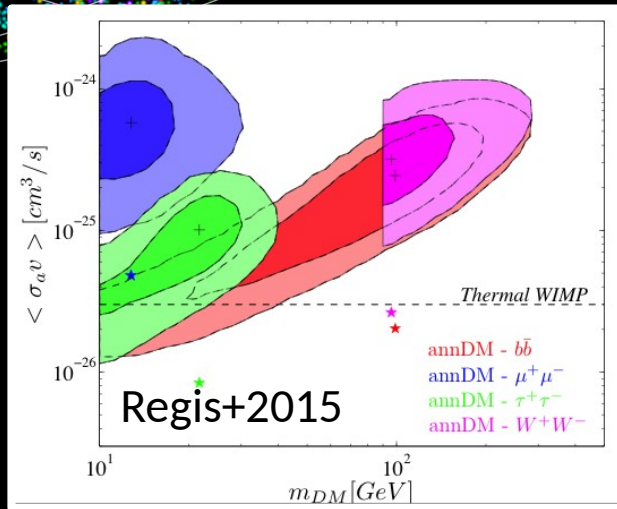
$$CCF(\vartheta) = \langle \delta I(\alpha)\delta N(\alpha+\vartheta) \rangle$$



Regis+2015

Measured x-correlation
vs. DM-annihilation model

Constraints on DM particle
Mass and x-section



Activities for the next future

- The aim is to extend these works to analyse the huge datasets that will be provided by the next-generation missions, like Euclid

- This represents a big challenge, both in terms of the implementation of statistical methods to analyse these large datasets, and for what concerns the modelling, with sufficient accuracy, of the measured quantities as a function of the cosmological model to be tested

Clustering in Euclid

Science Working Group

- Luigi Guzzo
- Will Percival
- Yun Wang

Organisation Unit - LE3

- Jean-Luc Starck
- Enzo Branchini
- Filipe Abdalla

Implementation and validation of numerical codes to measure the clustering of galaxies and clusters of galaxies

- two-point correlation function

  galaxy: de la Torre
  clusters: de la Torre, Marulli
- power spectrum

  galaxy: Balaguera-Antolinez
  clusters: Balaguera-Antolinez, Marulli
- three-point correlation function

  galaxy: Moresco
  clusters: Moresco, Marulli
- bispectrum

  galaxy: Porciani
  clusters: Porciani, Marulli
- covariance matrices of two-point correlation function and power spectrum

  galaxy: Monaco
  clusters: Hoyle

Numerical issues

In the recent past:

- one single code developer, or few (often developer = user) → no need for any specific code language, numerical libraries, operating systems, etc.
- software: few lines of code → simple linear structure, no need for object-oriented languages (e.g. Fortran was ok)
- hardware: 1 CPU → scalar codes ; no need to parallelize or aggressively optimize the algorithms
- no big issues on validation and documentation → a Readme file was enough
- one or few specific tasks for a single code → one code language was ok, no need for code integrations (no wrapper needed to convert from one language to another)

Numerical issues

Now and in the next future:

- many code developers working together → need of collaborative environments (e.g. CODEEN for Euclid) and common repositories to organise the activities (e.g. SVN, GitHub)
- software: huge number of code lines → object-oriented languages are now highly recommended (e.g. C++, Python)
- hardware: thousands of CPUs → codes have to be parallelized and optimized to run on large clusters in order to analyse increasingly large datasets
- significant efforts on validation and documentation (e.g. doxygen)
- many tasks to be accomplished → wrapper needed to connect codes written in different languages (e.g. SWIG)

What is required

→ astronomers have to develop complex numerical codes, that cannot be implemented by software engineers (scientific knowledges are required)

→ high-level computing skills are required : object-oriented languages, collaborative environments, parallelization, optimization, documentation, etc.

→ many work hours are needed for code implementation and validation

→ huge computing resources are mandatory to optimize, parallelize and validate the algorithms, and generate large mock catalogues