

Utilizing Astroinformatics to Maximize the Science Return of the Next Generation Virgo Cluster Survey

Nick Ball
Herzberg Institute of Astrophysics
Victoria, Canada



La Palma, Jun 2nd 2011 2 of 20

- g = 25.7, SB = 27.7, $M_B \sim -5$
- Photometric, 104 deg²
- 1.6x10⁴ 3.5x10⁶ galaxies
- Pl: Laura Ferrarese @ HIA
- 40 people at at 23 institutions in Canada, France, Hawaii, US, UK, and Italy.





NGVS Science Goals



4 of 20

La Palma, Jun 2nd 2011

 A definitive measurement of the faint-end shape of the cluster galaxy luminosity function

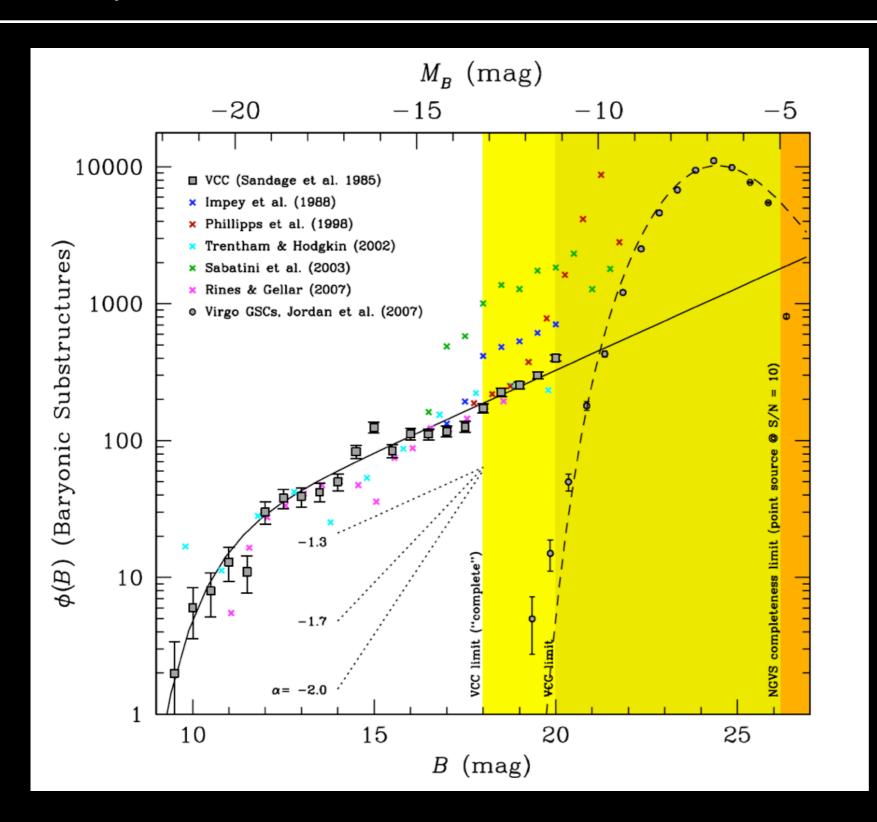
- Characterization of galaxy scaling relations over a factor 10⁷ in mass
- Stellar nuclei in galaxies and their connection to supermassive black holes
- Connection between the cluster, galaxies and the intracluster medium
- Fossil record of star formation and chemical enrichment in dense environments

NGVS Science Goals



La Palma, Jun 2nd 2011

5 of 20



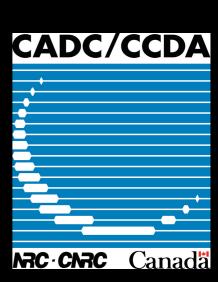
Virgo cluster luminosity function

Canadian Astronomy Data Centre



La Palma, Jun 2nd 2011 6 of 20

- CADC is one of the world's largest astronomy data centres
- ~500T of data
- Uses IVOA standards
- Staffed by astronomers and computer specialists



CANFAR



La Palma, Jun 2nd 2011 7 of 20

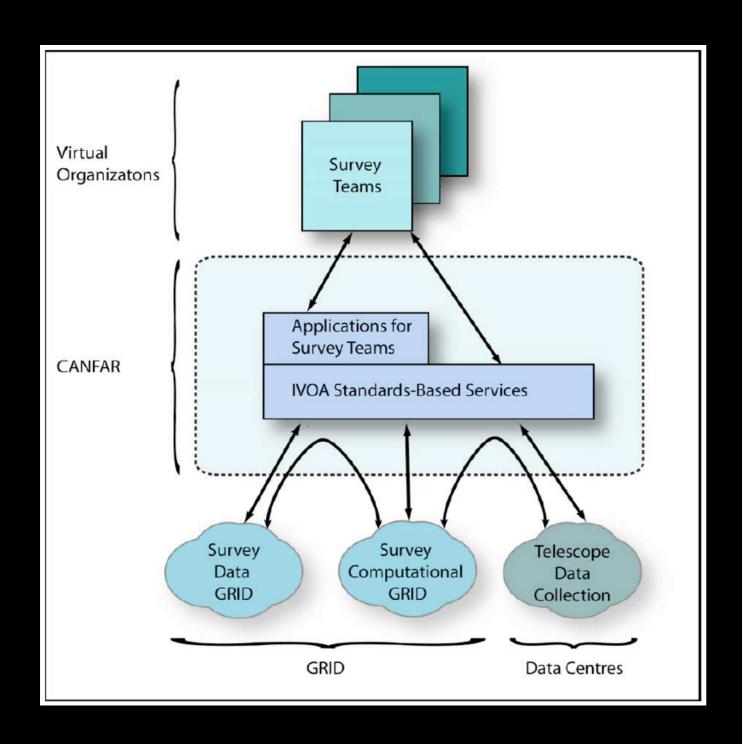
- Canadian Advanced Network for Astronomical Research
- Combines cluster job scheduling with cloud computing resources
- Users manage their own virtual machines



CANFAR



La Palma, Jun 2nd 2011



CANFAR infrastructure

Astroinformatics



La Palma, Jun 2nd 2011

- A *tool* to enable data-intensive astronomy
- Telescopes -> data -> ??? -> science
- IVOA KDD-IG guide: http://www.ivoa.net/cgi-bin/twiki/bin/view/
 IVOA/IvoaKDDguide

NGVS Data



10 of 20

La Palma, Jun 2nd 2011

Canada-France-Hawaii telescope

- Elixir preprocessing
- MegaPipe/TERAPIX processing
- 2.6T FITS files
- Catalogues
- Large, but not intractable

l.e.



La Palma, Jun 2nd 2011

- The NGVS is an ideal survey on which to use the methods of astroinformatics, without their use being the subject of the research itself
- (Although one still needs expertise in the subject!)

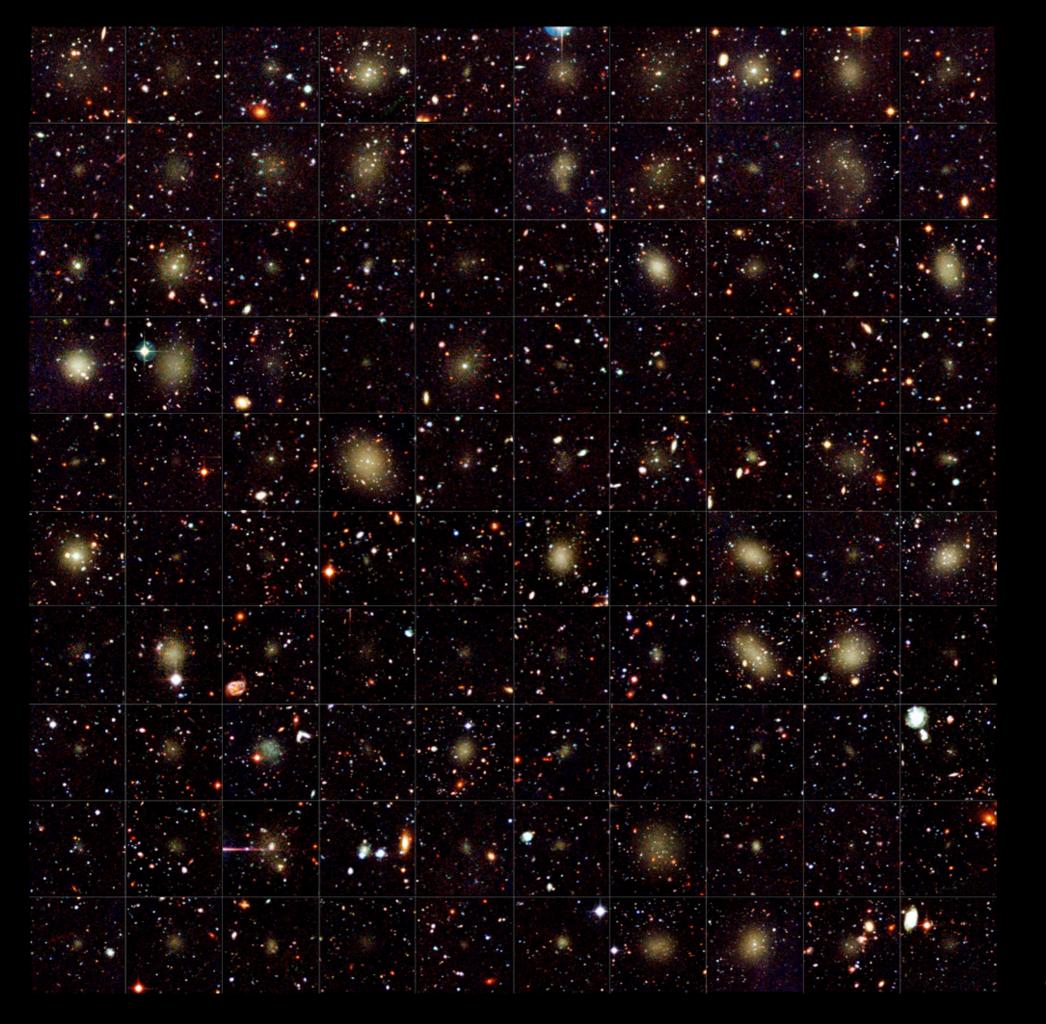
NGVS Challenges



La Palma, Jun 2nd 2011

Object detection and classification

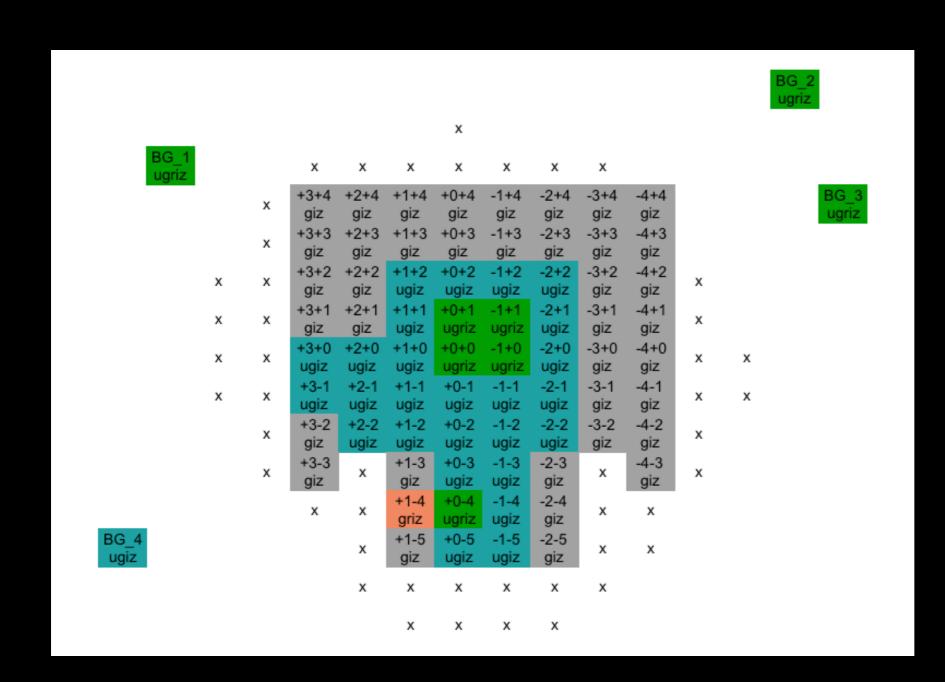
- Virgo cluster membership
- Missing data
- Photometric redshifts
- Multi-wavelength data
- Field-to-field variation
- Completeness (mag, SB, etc. etc.)



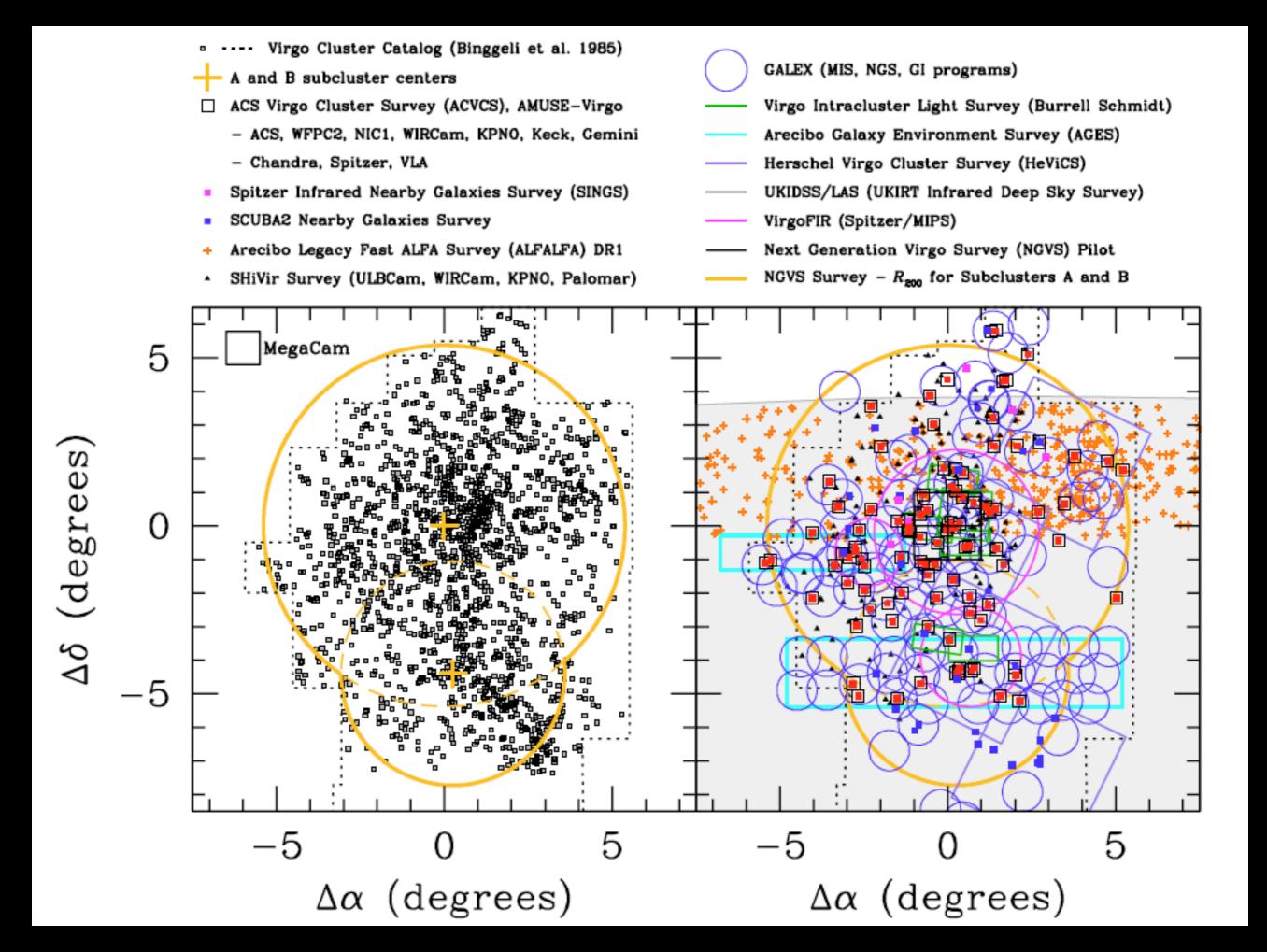
Low surface brightness galaxies







NGVS fields (not final)



Results



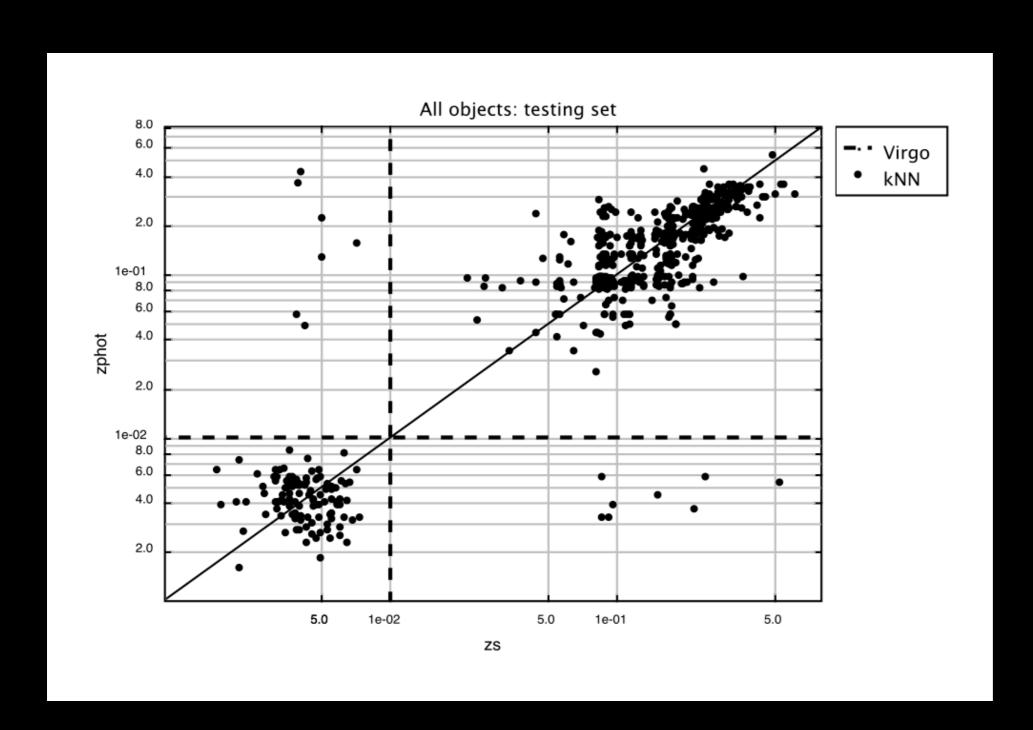
La Palma, Jun 2nd 2011

- Has been difficult to construct a catalogue
- Separating Virgo from the background
- Template-based photo-z codes fail below z ~ 0.1
- Empirical photo-z works but requires training set

Results



La Palma, Jun 2nd 2011



z_phot vs. z_spec

Results: Use of CANFAR



La Palma, Jun 2nd 2011

Rapidly speeds analysis compared to a desktop

- E.g., survey extinction corrections, simulated galaxies for the LF faint end
- Many more uses planned
- VOSpace for storing/distributing working files to collaboration

Fast Data Mining Algorithms



La Palma, Jun 2nd 2011

- Want data mining infrastructure in analogy to CANFAR
- Must be generic tools, and scale as NlogN or better
- Tests with SkyTree software confirm it scales with large astronomy data
- Plan to compare to open-source, e.g., TMVA, SPR

Conclusions



La Palma, Jun 2nd 2011 20 of 20

- NGVS is an ideal project for astroinformatics and astrostatistics
- CADC + CANFAR is an ideal setting on which to implement this
- Still need to be expert in it, but can get science using astroinformatics without it becoming the subject of the research itself