# DEEP-Theory Meeting 23 October 2017

Expected at 4 pm: Elliot Eckholm, Viraj Pandya, Graham Vanbenthuysen Not coming: Vivian, James, Clayton

**GALFIT-type analysis of VELA simulations using deep learning** — Marc Huertas-Company Deep Learning GALFIT emulator vs. Haowen Zhang running them through GALFIT.

**Prolate galaxies: observation-simulation comparison** — Haowen Zhang and Vivian Tang: analysis of CANDELS b/a vs. Δa data & mocks; half-stellar-mass radius r\_0.5 vs. half-stellar-light radius r\_e from simulations.

**Deep Learning for Galaxy Environment project** — The paper by Nicolas Tejos, Aldo Rodriguez-Puebla, and me is now published in MNRAS. James Kakos, Dominic Pasquale, and Matthew Casali plan to use DL for a project to improve z and local environment for a mixture of spectroscopic and (mostly) photometric redshifts.

**Galaxy size vs. local density project** — Graham Vanbenthuysen, Viraj Pandya, Christoph Lee, Doug Hellinger, Aldo Rodriguez-Puebla, David Koo, Lin Lin — We are measuring  $\lambda$  vs. density by various methods in Aldo's mock catalogs from Bolshoi-Planck and MultiDark-Planck, and SDSS galaxy radii vs. density by the same methods. Christoph will show how  $R_s$  and  $R_s(C_{NFW}/7)^{0.4}$  depend on environmental density.

Elongated galaxies aligned with cosmic filaments? — Viraj Pandya is working on observations & mocks.

**Deep Learning for Galaxies project** — Analysis of VELA Gen3 simulations is ongoing by Raymond Simons at JHU, Christoph Lee and Sean Larkin, along with Avishai's student Tomer Nussbaum: finding all satellites. Christoph is also using the DL code that classified CANDELS images to classify VELA mock galaxy images.

Abundance matching is independent of environmental density - Radu Dragomir, Aldo, Christoph paper soon

Halo properties like concentration, accretion history, and spin are mainly determined by environmental density rather than by location within the cosmic web — we are finishing the paper led by Tze Goh

**DM halo mass loss** paper being finished — Christoph Lee, Doug Hellinger. Related work this summer on **halo radial profile** by SIP students Shawn Zhang and Peter Wu with Christoph.

Simulations of CGM & winds vs. observations — Clayton Strawn, Hassen Yesuf

**Improved Santa Cruz Semi-Analytic Model** of galaxy population evolution, including insights from high-resolution hydro simulations — Viraj Pandya, Christoph Lee, Rachel Somerville, Sandy Faber



## CNN Galfit Emulator (Marc Huertas-Company) vs. Galfit (Haowen Zhang) Applied to VELA Gen3 CANDELized Images



## CNN Galfit Emulator (Marc Huertas-Company) vs. Galfit (Haowen Zhang) Applied to VELA Gen3 CANDELized Images



## CNN Galfit Emulator (Marc Huertas-Company) vs. Galfit (Haowen Zhang) Applied to VELA Gen3 CANDELized Images



## Evolution with Redshift of Density Dependence of $C_{NFW}$ , $\lambda_B$ , $\dot{M}/M$



z = 0 Density Dependence of C<sub>NFW</sub>, Rs, Rs (C<sub>NFW</sub>/7)<sup>0.4</sup>

z = 0 Density Dependence of C<sub>NFW</sub>,  $\lambda_B$ ,  $\dot{M}/M$ 



We also need to know the dispersion in R<sub>S</sub> and R<sub>S</sub>(C<sub>NFW</sub>/7)<sup>0.4</sup>. We know that the dispersion of  $\lambda_B$  and galaxy radius are log-normal. If either R<sub>S</sub> or R<sub>S</sub>(C<sub>NFW</sub>/7)<sup>0.4</sup> control galaxy radii, as Fangzhou Jiang claims is true for NIHAO and VELA simulations, these quantities should also have log-normal dispersion in mass bins. Christoph Lee is looking at this.



### z = 0 Density Dependence of C<sub>NFW</sub>, Rs, Rs (C<sub>NFW</sub>/7)<sup>0.4</sup>

C<sub>NFW</sub>, Rs, Rs (C<sub>NFW</sub>/7)<sup>0.4</sup> Are Log-normal





#### Measuring $\lambda_B$ and $R_s$ vs. Density in Spheres of 4 and 8 h<sup>-1</sup> Mpc

Spherical Radius = 4 Mpc Graham Vanbenthuysen & Viraj Pandya Spherical Radius = 8 Mpc



#### Deep Learning for Galaxy Environment project - The SORT paper by Nicolas Teios54 16

#### published: Stochastic Qe 9 of 14 measurements, MNRAS to improve z and local e



contains more information than a less peaked one, and that having a distribution like  $\Delta z|_{\text{sort}}$  is indeed an improvement over  $\Delta z|_{\text{ph}}$ .

This fact can be partially appreciated by our previous comparison of the overall redshift distributions (see Figure 2, Section 4.2). We can observe that the original photometric sample not only loses information regarding the positions of the large-scale structure (i.e. peaks and valleys) but also produces biases in the expected number of galaxies as a function of redshift.

Another way to look at this issue is by quantifying the  $\Delta z|_{\rm obs} \equiv z_{\rm obs} - z_{\rm true}$  as a function of  $z_{\rm obs}$ . In Figure 5 we show the mean of  $\Delta z|_{\rm obs}$  as a function of  $z_{\rm obs} = z_{\rm ph}$  (left panel) and  $z_{\rm obs} = z_{\rm sort}$  (right panel). We observe that indeed the photometric sample is highly biased towards negative  $\Delta z$  values at the lowest redshift bins and somewhat biased towards positive  $\Delta z$  values at the highest redshift bins. In contrast,  $z_{\text{sort}}$  are virtually unbiased across the full redshift range which is clearly an improvement.

In the following subsections we explore how SORT performs

0.14

0.12

0.10

0.08

0.00

0.04

0.02

0.00

0.02

0.04

SORT redshifts

0.06

80.0

Ztrue

21 22 23

∩ ∩∩



 $\Delta z_{\rm lobs} \equiv z_{\rm obs} - z_{\rm true}$  as a function of  $z_{\rm obs}$ . In Figure 5 we show the mean of  $\Delta z|_{obs}$  as a function of  $z_{obs} = z_{ph}$  (left panel) and  $z_{\rm obs} = z_{\rm sort}$  (right panel). We observe that indeed the photometric sample is highly biased towards negative  $\Delta z$  values at the lowest redshift bins and somewhat biased towards positive  $\Delta z$  values at the highest redshift bins. In contrast,  $z_{\text{sort}}$  are virtually unbiased across the full redshift range which is clearly an improvement.

In the following subsections we explore how SORT performs at retrieving the three-dimensional large-scale structure distribution as well as the two-point autocorrelation function.

#### 4.3.1 Three-dimensional spatial distributions

Figure 6 shows a subvolume of the mock survey, where the redshifts come from our different estimates:  $z_{ph}$  (top left panel),  $z_{sp}$ (top right panel),  $z_{\rm sort}$  (bottom left panel) and  $z_{\rm true}$  (bottom right panel). Unsurprisingly, the original photometric sample does not provide a good description of the cosmic web. On the other hand, our SORT method is able to reproduce most of the significant largescale-structures in the volume, including voids, dense filaments, groups and clusters (compare the two bottom panels to each other). Part of this success is due to the chosen parameters for applying the method (i.e. those relevant for the cosmic web scales; see Section 4.1) in combination with having a representative reference sample (i.e.,  $z_{sp}$ ; see the top right panel of Figure 6). The recovered distribution is not perfect however and some noticeable artefacts exist. For example, the  $z_{\rm sort}$  cosmic walls are thicker and the voids Those should in principle be overcome

> s for applying the method (specially though the cosmic web seems to be whole, we also emphasize that there ies assigned to the wrong structures se contributing to the tails of the  $\Delta z$ Still, the fact that  $z_{\text{sort}}$  provides a ies assigned to the *right* structures e narrow peak in  $\Delta z$  distribution), mising for further studies aimed to l distribution of the cosmic web and alaxies using photometric redshifts.

#### unctions

ation function (2PCF) is not enough ensional spatial distribution, it is a ical measurement that contains releitial clustering. As such, it has been tragalactic questions not just related also to galaxy evolution and cosmol-

**2nd stage**: Improve the the test case. This is 4a | redshift/imaging survevers SORT to a mock CAND based on the Bolshoi (higher Vmax) galaxies, possibly also intermedia (density, cosmic web)<sup>5</sup>ar

ri and 0%, see section +.1). ritulough the cosmic web sections to be reasonably well recovered as a *whole*, we also emphasize that there will be some individual galaxies assigned to the wrong structures along the line-of-sight, i.e. those contributing to the tails of the  $\Delta z$ distribution (see Section 4.3). Still, the fact that  $z_{\text{sort}}$  provides a considerable fraction of galaxies assigned to the right structures along the line-of-sight (i.e. the narrow peak in  $\Delta z$  distribution), makes this SORT method promising for further studies aimed to quantify the three-dimensional distribution of the cosmic web and cosmological environment of galaxies using photometric redshifts.

#### 4.3.2 Two-point correlation functions

0.12 0.14

Although the two-point correlation function (2PCF) is not enough

timated with a bootstrap technique from 100 shaded area corresponds to the intrinsi $\partial \Phi_{T}$ sample variance (i.e. this is the uncertainly underlying true redshift for all the gaboies metric sample). The light-blue area correst tainty around the unbiased measurement u sample (i.e. the remaining 30% of galaxies Section 3). This figure demonstrates that  $z_{so}$ true 2PCF on scales  $\geq 4 h^{-1}$ Mpc in a some

0.10 0.12 0.14

at a higher statistical precision than that of the reference spectroscopic sample. At smaller than  $\approx 4 h^{-1}$  Mpc scales however,  $z_{\text{sort}}$ is not able to recover the 2PCF power; these scales are comparable or smaller than that of the narrow peak recovered in the redshift uncertainty distribution (see Section 4.3). On the other hand, the sente de la la la construcción de la servición de la servición

Figure 7 shows the ratio between the measured redshift space 2PCF and its true underlying value,  $\xi_{true}(s)$ , as a function of redshift space distance s. The black circles correspond to using  $z_{\text{sort}}$ 

© 2017 RAS, MNRAS 000, 1-14

#### fainter galaxies, and ining the environment





0.12

0.10

0.06

0.04

0.02

0.00

redshifts

0.02 0.04 0.06

0.08

Ztrue

0.10

14

17

18

19

20

**Deep Learning for Galaxy Environment project** — The SORT paper by Nicolas Tejos, Aldo Rodriguez-Puebla, and me is now published: Stochastic Order Redshift Technique (SORT): a simple, efficient and robust method to improve cosmological redshift measurements, MNRAS, 473. 366. James Kakos, Viraj Pandya, Dominic Pasquale, and Matthew Casali plan to use DL for a project to improve z and local environment for a mixture of spectroscopic and (mostly) photometric redshifts.



**2nd stage**: Improve the treatment in the Tejos, Rodriguez-Puebla, Primack SORT paper. This paper used a mock SDSS sample as the test case. This is a large-area survey, but the SORT method is designed for pencil-beam surveys, like CANDELS and other distant redshift/imaging surveys. (The complementary method due to Bryce Menard is designed for large-area surveys.) So let's first apply SORT to a mock CANDELS-type pencil beam survey, using the 8 mock backward light cones for each of the 5 CANDELS fields, based on the Bolshoi-Planck simulation. Also, let's assign spectroscopic redshifts (from the backward light cones) to the brighter (higher Vmax) galaxies, photometric redshifts (degraded by an appropriate Gaussian spread  $\Delta z$  (1+z) ) to the fainter galaxies, and possibly also intermediate-accuracy grism redshifts. Goal: see how much improvement we can get in determining the environment (density, cosmic web) around galaxies using SORT.

**Deep Learning**: The training set will be the same combination(s) of spec-z, photo-z, and possibly grism-z measurements as for the 2nd stage, plus the true redshifts from the simulation. We will then see how well the DL code can predict the true environments of distant galaxies. The challenge in designing the DL will be that this is a 3D point project, unlike the 2D images that face recognition and our "face recognition for galaxies" project analyzes. We hope that our Google friends will help us design an appropriate DL setup.

# Does the Galaxy-Halo Connection Vary with Environment?

Radu Dragomir<sup>1\*</sup>, Aldo Rodríguez-Puebla<sup>2,3</sup><sup>†</sup>, Joel R. Primack<sup>1</sup><sup>‡</sup>, Christoph T. Lee<sup>1</sup>§

<sup>1</sup>Physics Department, University of California, Santa Cruz, CA 95064, USA

<sup>2</sup> Department of Astronomy & Astrophysics, University of California at Santa Cruz, Santa Cruz, Ca 95064, USA

<sup>3</sup> Instituto de Astronomía, Universidad Nacional Autónoma de México, A. P. 70-264, 04510, México, D.F., México

Released 20?? Xxxxx XX

#### ABSTRACT

SubHalo Abundance Matching (SHAM) assumes that one (sub)halo property, such as mass  $M_{\rm vir}$  or peak circular velocity  $V_{\rm peak}$ , determines properties of the galaxy hosted in each (sub)halo such as its luminosity or stellar mass. This assumption implies that the dependence of Galaxy Luminosity Functions (GLFs) and the Galaxy Stellar Mass Function (GSMF) on environmental density is determined by the corresponding halo density dependence. In this paper, we test this by determining from an SDSS sample the observed dependence with environmental density of the *ugriz* GLFs and GSMF for all galaxies, and for central and satellite galaxies separately. We then show that the SHAM predictions are in remarkable agreement with these observations, even when the galaxy population is divided between central and satellite galaxies. However, we show that SHAM fails to reproduce the correct dependence between environmental density and color for all galaxies and central galaxies, although it correctly reproduces the color dependence on environmental density of satellite galaxies.

**Key words:** Galaxies: Halos - Cosmology: Large Scale Structure - Methods: Numerical



Figure 1. The global *ugriz* galaxy luminosity function. Our derived *ugriz* GLFs and GSMF are shown with the black circles with error bars. For comparison we reproduce the *ugriz* GLFs from Blanton et al. (2005a, black long dashed lines) based on the SDSS DR2; Hill et al. (2010, dotted lines) by combining the MGC, SDSS DR5 and the UKIDSS surveys; and Driver et al. (2012, short dashed lines) based on the GAMA survey. As for the stellar masses we compare with the GSMF from Baldry et al. (2012) and Wright et al. (2017), black long and short dashed lines, respectively.





**Figure 2.** Absolute magnitude in the r-band as a function of redshift for our magnitude-limited galaxy sample. The blue solid box shows our volume-limited DDP sample. Note that our DDP sample restricts to study environments for galaxies between  $0.03 \leq z \leq 0.11$  as shown by the dashed lines.

**Figure 11.** Mean density as a function of galaxy g-r color, from the SDSS DR7 (shaded regions) and the mean density predicted by SHAM based on the BolshoiP simulation, dotted lines with error bars. We present the mean density for all, central, and satellite galaxies as indicated by the labels. SHAM fails to predict the correct relationship between mean density and galaxy colors for all galaxies and central galaxies. In contrast, the SHAM prediction for satellite galaxies is in better agreement with observations.



Figure 3. Left Panel: Luminosity-to- $V_{\text{max}}$  relation from SHAM. The different colors indicate the band utilized for the match. Right Panel: Stellar mass-to- $V_{\text{max}}$  relation. Recall that SHAM assumes that these relations are valid for centrals as well as for satellites. In the case of centrals  $V_{\text{max}}$  refers to the halo maximum circular velocity, while for satellites  $V_{\text{max}}$  represents the highest maximum circular velocity ( $V_{\text{peak}}$ ) reached along the subhalo's main progenitor branch. SHAM assumes that  $V_{\text{max}}$  fully determines these statistical properties of the galaxies.



Figure 4. Two-point correlation function in five luminosity bins at z = 0.1. The solid lines show the predicted two-point correlation based on our *r*-band magnitude-to- $V_{\text{max}}$  relation from SHAM, while the circles with error bars show the same but for the SDSS DR7 (Zehavi et al. 2011).



Figure 5. Two-point correlation function in five stellar mass bins. The solid lines show the predicted two-point correlation based on our stellar mass-to- $V_{\text{max}}$  relation from SHAM, while the circles with error bars show the same but for SDSS DR7 (Yang et al. 2012).



Figure 6. Comparison between the observed SDSS DR7 ugriz GLFs and GSMF, filled circles with error bars, and the ones predicted based on the BolshoiP simulation from SHAM, shaded regions, at four environmental densities in spheres of radius 8  $h^{-1}$ Mpc. We also reproduce the best fitting Schechter functions to the *r*-band GLFs from the GAMA survey (McNaught-Roberts et al. 2014). Observe that SHAM predictions are in excellent agreement with observations, especially for the longest wavelength bands.



Figure 7. Left Panel: Comparison between the observed r-band GLF with environmental density in spheres of 8  $h^{-1}$ Mpc, filled circles with error bars, and the ones predicted based on the BolshoiP simulation from SHAM, shaded regions. The dashed lines show the best fitting Schechter functions to the *r*-band GLFs from the GAMA survey (McNaught-Roberts et al. 2014). Right Panel: Similar to the left panel but for the GSMF with environmental density. Here again the dashed lines are the best fitting Schechter functions.