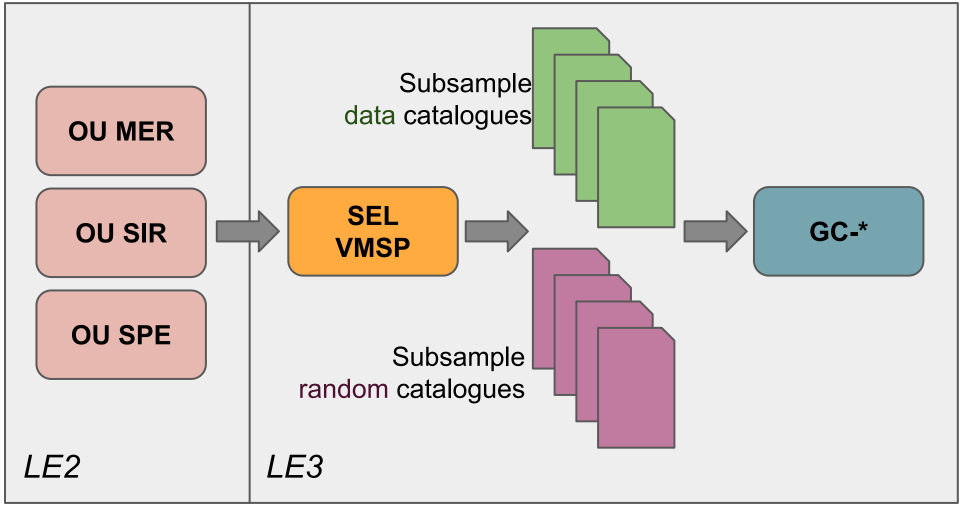
Characterizing the spectroscopic selection function for galaxy clustering and application to SC8

Ben Granett
INAF OA Brera

24 Feb 2022
5° Meeting Nazionale Collaborazione Euclid

Spectroscopic galaxy clustering pipeline

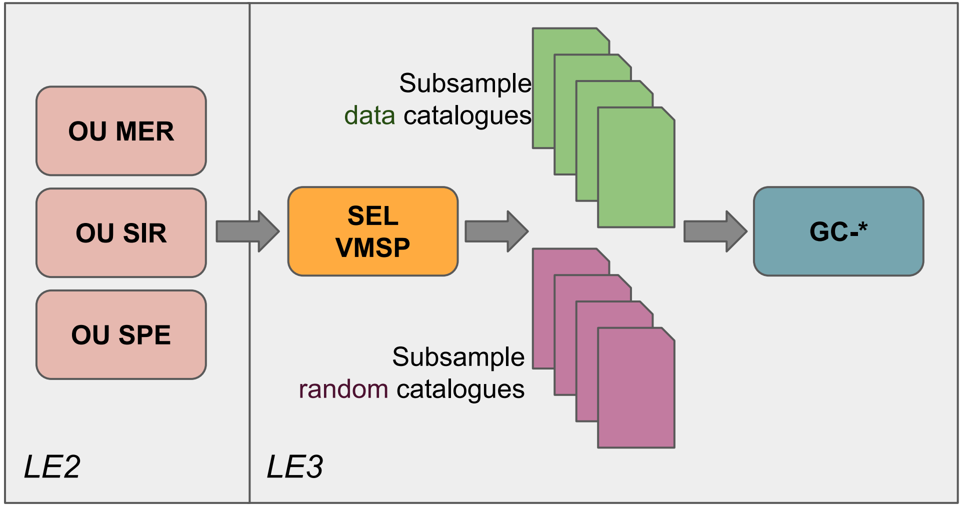
- *The **LE3 Internal data** work package* is the interface between **LE2** and **LE3**.
- The ***SEL*** processing function selects the spectroscopic samples for galaxy clustering and estimates the **purity** and **completeness** of these samples.
- ***VMSP*** (spectroscopic visibility mask) characterizes the selection function and builds the random catalog for galaxy clustering.



Benjamin.Granett@inaf.it

Internal Data SEL-VMSP team

- The SEL-VMSP processing functions are distributed between UK, Italy and USA.
  - *Manager:* Sesh Nadathur 🇬🇧
  - *Validation lead:* Lucia F. de la Bella 🇬🇧
  - *VMSP developer:* Ben Granett 🇮🇹
  - *SEL developer:* Lado Samushia 🇺🇸
  - *Active collaborators:*
    Coleman Krawczyk 🇬🇧,
    Enzo Branchini, Pierluigi Monaco,
    Samuele Galeotta 🇮🇹🇮🇹🇮🇹
- There are three related prelaunch key project papers.

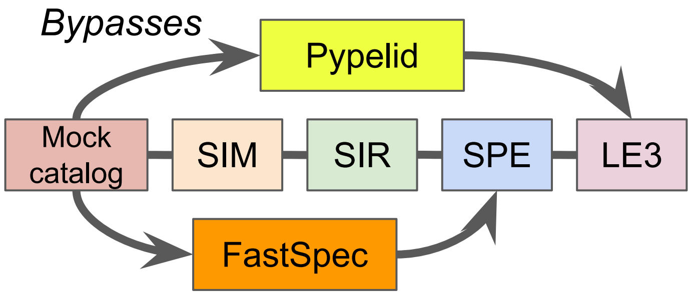


Benjamin.Granett@inaf.it

In collaboration with the SWG E2E galaxy clustering team

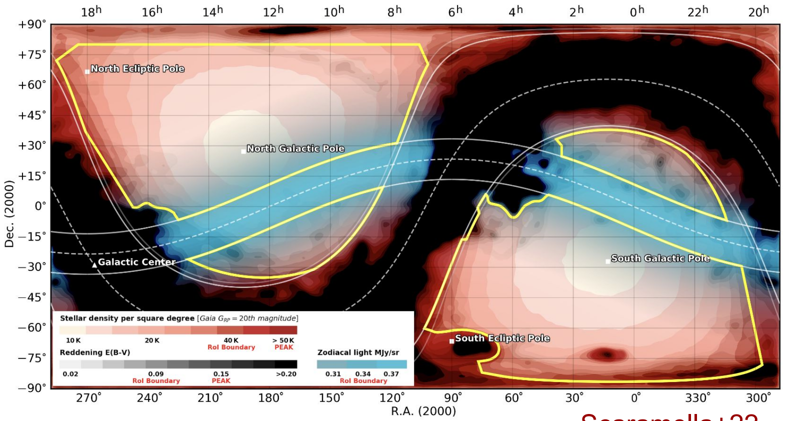
- Leads: Michele Moresco, Ben Granett, Sylvain de la Torre

Collaborators: Lucia Pozzetti, Micol Bolzonella, Bianca Garilli, Marco Scodeggio, Claudia Scarlata, Sean Bruton, Maxwell Kuschel, Matthieu Bethermin, Vincent Le Brun, Miguel Delaire, Herve Aussel, Dida Markovic



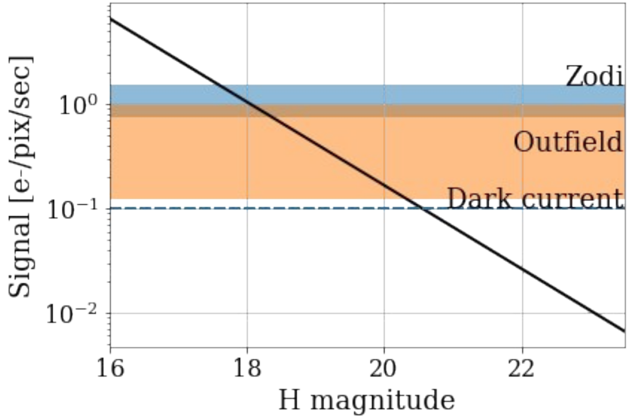
Challenges for galaxy clustering

- Slitless spectroscopy suffers from high background:
  - Zodi emission, scattered star light and MW extinction make large gradients in SNR on the sky.
  - Spectra overlaps add contamination and confusion.
- Continuum emission will not be detected in the Wide for target galaxies.
- Spectroscopic detection and redshift measurement will be based on a single emission line for most sources.
- *Our job:*
  Quantify the probability that a given galaxy would be measured successfully by Euclid if it exists at a given position on the sky.



Scaramella+22

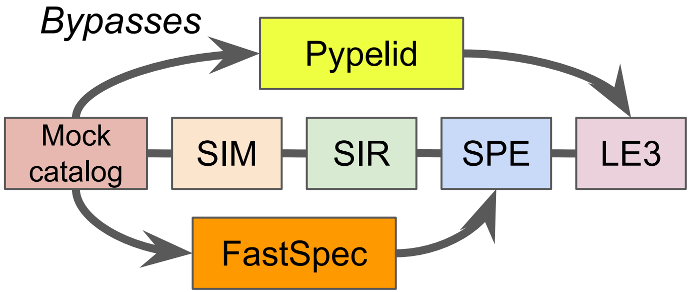
Fig. 15: The RoI outline (17 354 deg^2) with the accepted ranges of the stellar density, dust extinction, and zodiacal light.

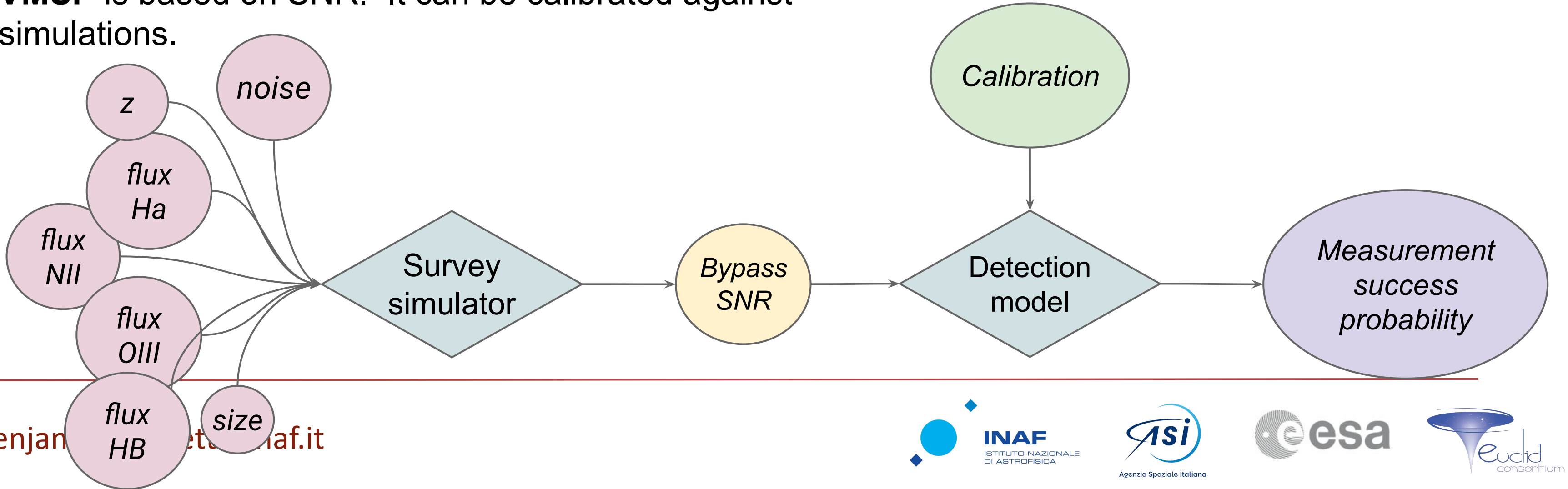


Benjamin.Granett@inaf.it

E2E simulations and bypass detection model

- End-to-end simulations of the SGS start from mock galaxy catalogs and go to redshift measurements.
- We speed up the process with bypass simulations: skip heavy image processing and work at the catalog level.
- The bypass detection model used in Pypelid and **LE3 VMSP** is based on SNR. It can be calibrated against simulations.

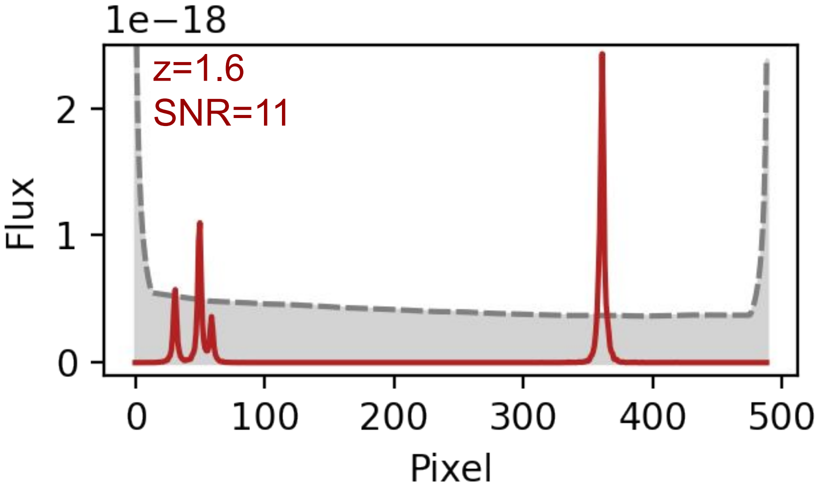




How do we define bypass SNR?

- The bypass SNR is computed on a simulated 1D spectrum without noise.
- It is the sum of the SNR in quadrature over the extracted 1D spectrum.
- The spectrum is simulated without continuum.
- *NB: This is not equivalent to the SNR estimated by SPE or specified in the requirements documents.*
- Procedure in VMSP:
    - Simulate 2D images of the emission lines assuming Gaussian profile and double Gaussian PSF model.
    - Extract 1D spectrum
    - Compute SNR

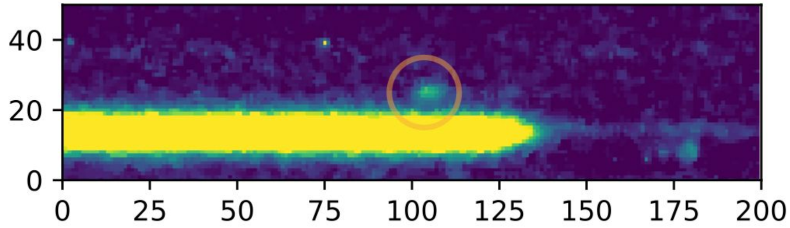
$$SNR = \sqrt{\sum_{i=1}^{n_{pix}} \frac{f_i^2}{\sigma_i^2}}$$



Benjamin.Granett@inaf.it

What about the noise?

- We measured the noise for the TU galaxies directly from the SIR science frame variance images.
- Includes all simulated backgrounds including contamination.
- The variance is looked up at the pixel coordinates red and blue of the emission line of the TU galaxies.
- The results here are the result of dedicated work with help from Maxwell Kuschel and Sean Bruton:
  - We downloaded terabytes of TU, MER, SIR, SPE Wide and Deep data products from EAS.
  - Matched the TU catalog with MER detections and SPE measurements.
  - Ran VMSP codes on the SC8 data products.



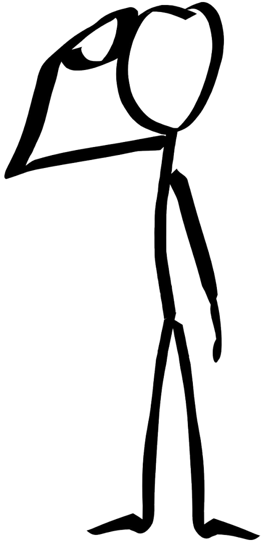
Cutout of SIR Science frame variance image (e-) with the Hα line of a TU galaxy circled.

Thanks to Marco Fumana for help with the SIR location table routines!

Benjamin.Granett@inaf.it

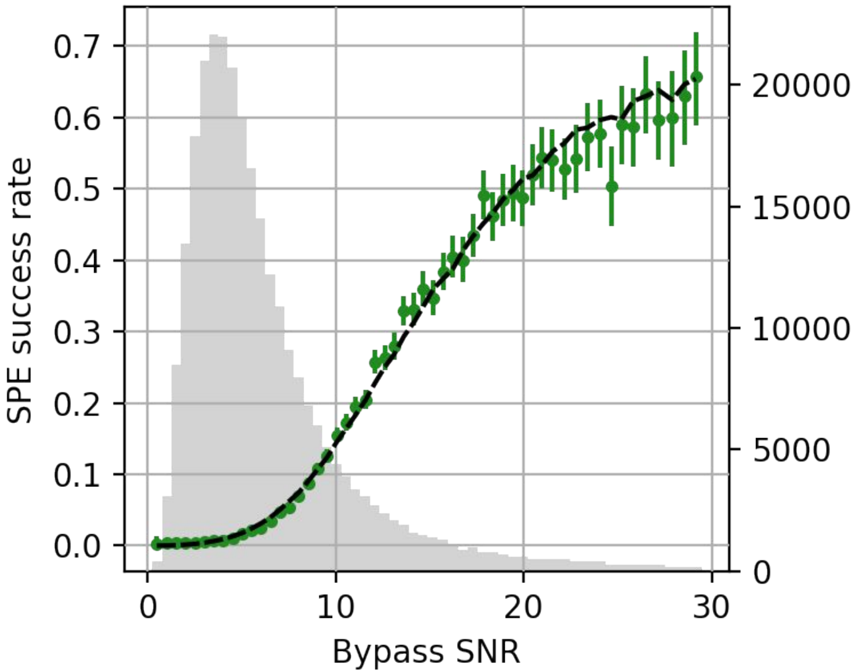
Obligatory warning

- Data from SC8 carry many caveats (see Michele's presentation, or speak to anyone working in an OU…)
- The scope here is to model the selection function.
- I am not worried by the absolute values.



Benjamin.Granett@inaf.it

Calibration of the detection model with SC8



- SPE success is defined as `SPE_Z_PROB > 0.99 & |SPE_Z - z|/(1+z) < 0.003.`

  The sample is cut at `flux_Ha > 1e-16 erg/cm2/s & 0.9 < z < 1.8`

- Fit the success rate as a function of SNR with a sigmoid.

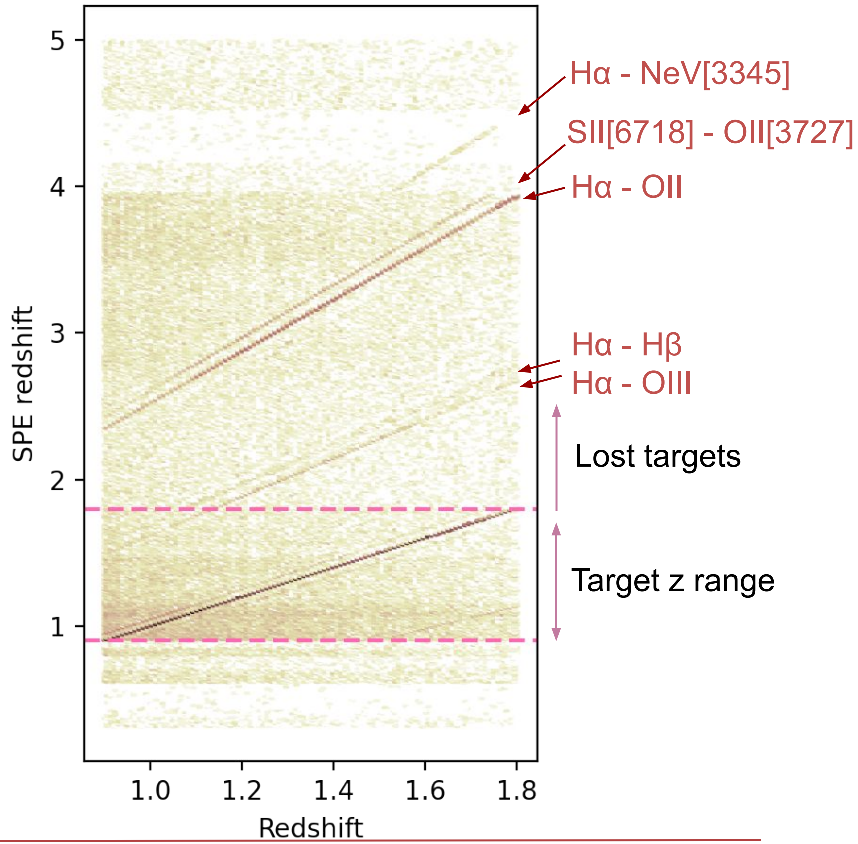
$$y = \frac{c}{1 + (x/x_0)^{-\alpha}}$$

- This is done in redshift bins to capture redshift dependence.
- Does not asymptote to 1 due to line misidentification losses.

Benjamin.Granett@inaf.it

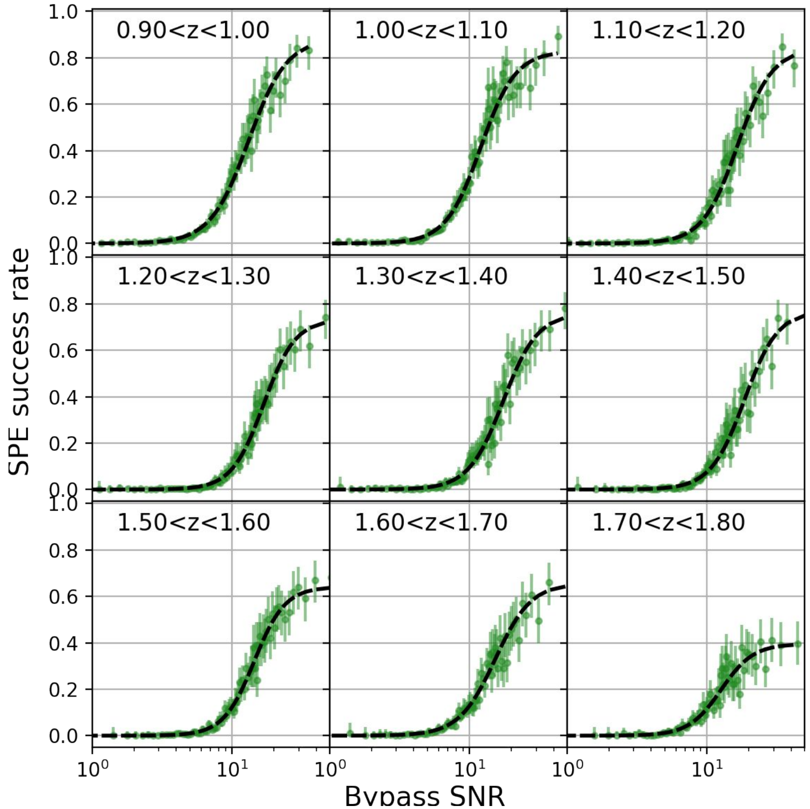
Calibration of the detection model with SC8

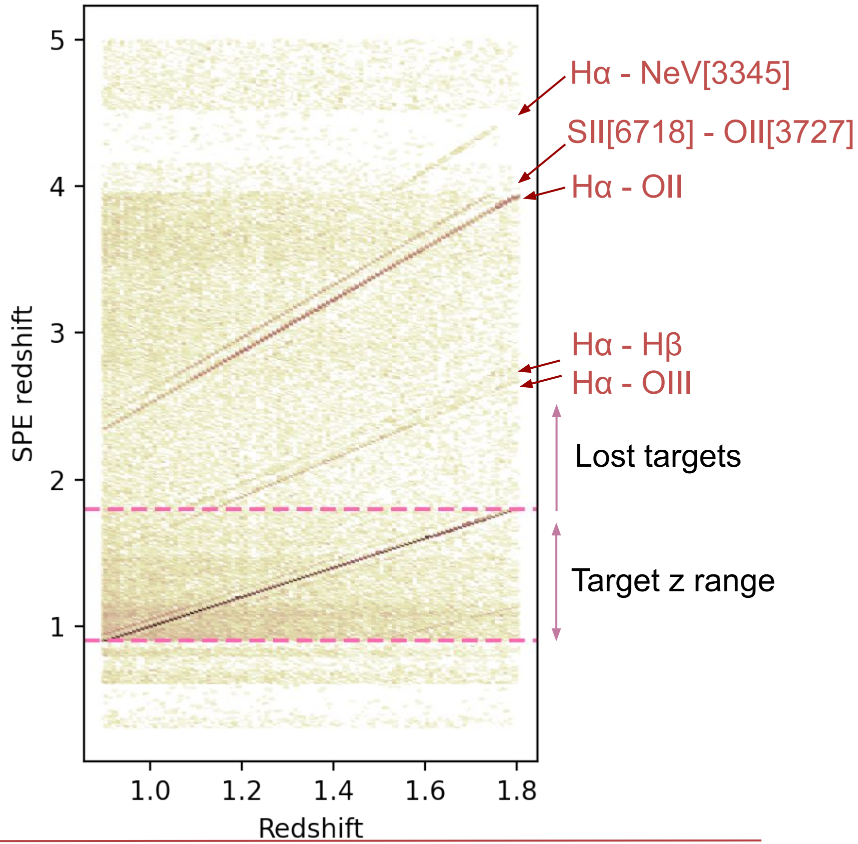
- Galaxies are lost from the sample due to line misidentification or spurious noise.
  - Hα may be mistaken for something else
  - Spurious signal from noise.
- This introduces a complex redshift dependence of the completeness.
  - Measurement success rate and misidentification/spurious rates depend on SPE template priors.



Benjamin.Granett@inaf.it

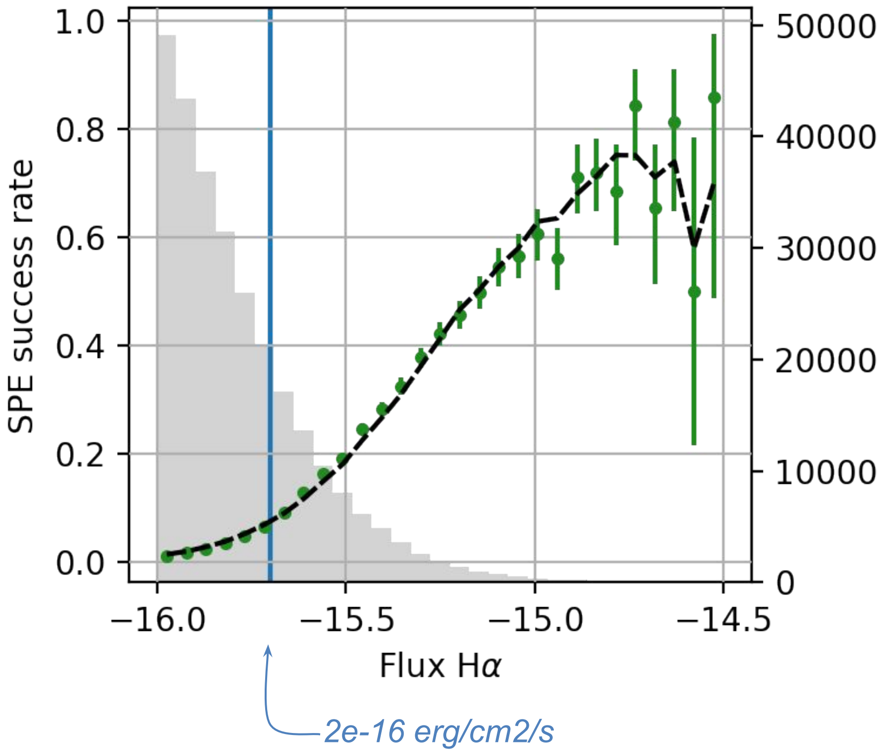
Calibration of the detection model with SC8





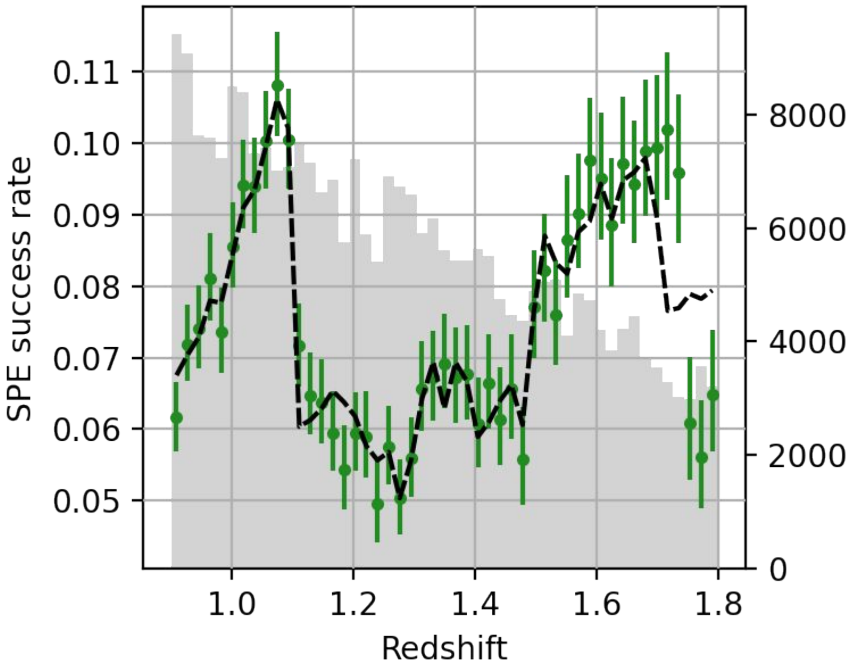
Benjamin.Granett@inaf.it

Redshift measurement trends in SC8

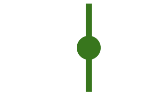


- We can measure the SPE success rate as a function of any galaxy property.
- Similarly we can use the model p(SNR, z) to predict the success rate as a function of these properties.
  - The dependencies pass through SNR which is a function of flux, size, redshift and noise.

Redshift measurement trends in SC8

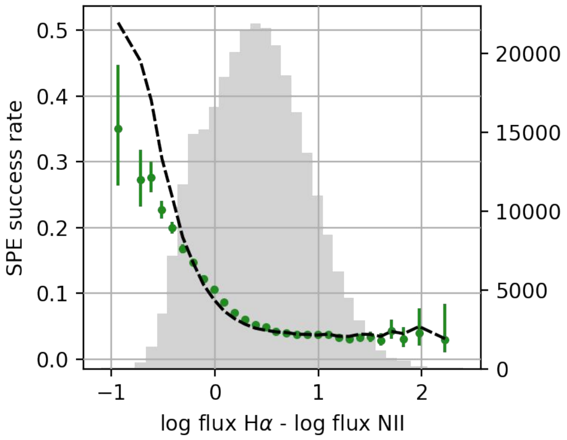


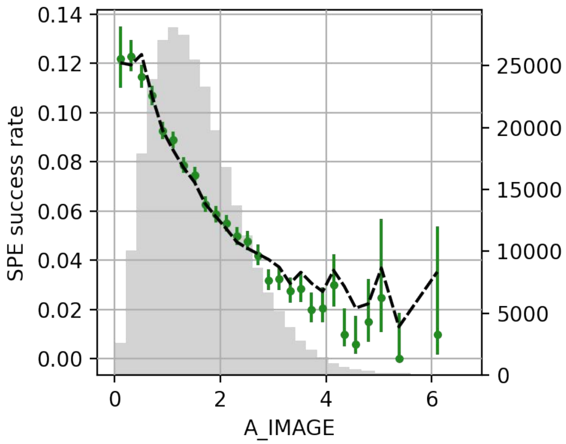
- Recall the model was fit in redshift bins of size 0.1. This was needed to reproduce the variations in measurement success rate.
- But the model successfully reproduces trends inside the bins.
- Further investigation is needed to understand poor fit at z>1.7.

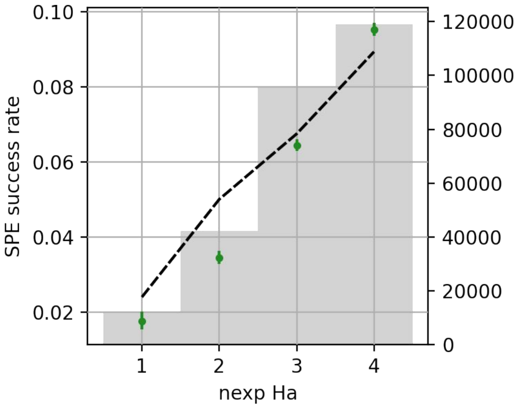


Benjamin.Granett@inaf.it

Redshift measurement trends in SC8





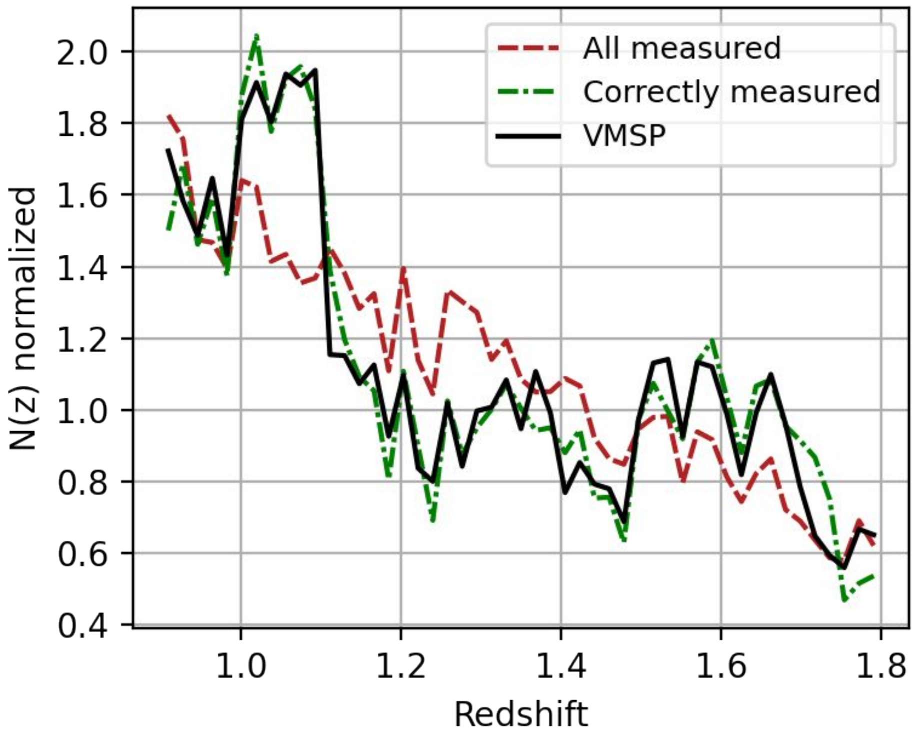


- Hα/NII flux
- Angular size
- Exposure count

The detection model reproduces the trends generally well.
A few parameters require checks and deeper investigation.

Benjamin.Granett@inaf.it

Proof of concept random catalog

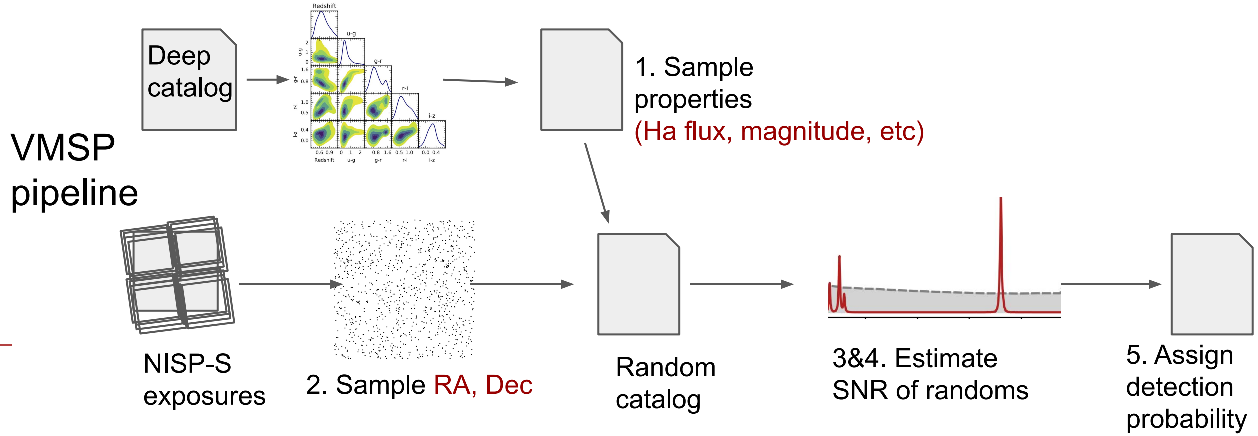


- We build a random catalog by sampling galaxies from TU in SC8 fields.
- Apply the calibrated detection model to the randoms (***VMSP*** line).
- This reproduces the N(z) of galaxies in the target sample that were measured correctly (***Correctly measured***).
- The distribution is different from the N(z) of all galaxies (***All measured***) due to incompleteness.
- The randoms are constructed to follow the N(z) of the correctly measured galaxies.

Outline of VMSP function

VMSP performs the following steps to build the random catalog for the Wide.

1. Sample galaxies from the CPC/Deep fields to build the random catalog.
2. Sample sky coordinates RA, Dec from the Wide survey area.
3. Measure noise from SIR pixels at the location of the emission lines of each random galaxy (*RA, Dec, z, wavelength*).
4. Compute the bypass SNR for each random.
5. Map bypass SNR to the detection probability.
   The detection model will be calibrated using simulations and the CPC/Deep fields.
6. Downsample the random catalog using the detection probability.



Benjamin.Granett@inaf.it

Summary

- We are developing algorithms to characterize the spectroscopic selection function for galaxy clustering and digging into SC8 data.
- Provides a proof of concept of the LE3 VMSP pipeline that builds the random catalog for galaxy clustering.
- The density of randoms is modulated by noise through the detection model.
- This work will be written up in the LE3 VMSP pre-launch key project paper.
- The algorithms will continue to be validated and improved as simulations and reduction pipelines are iterated.