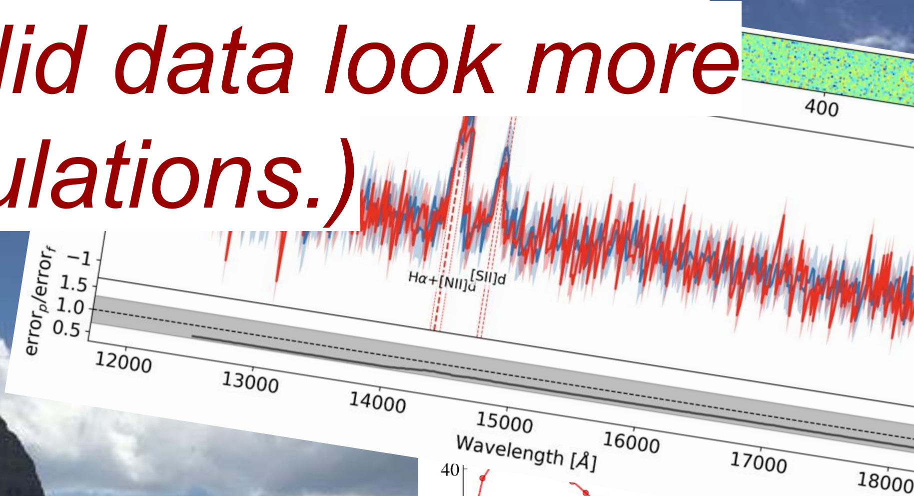
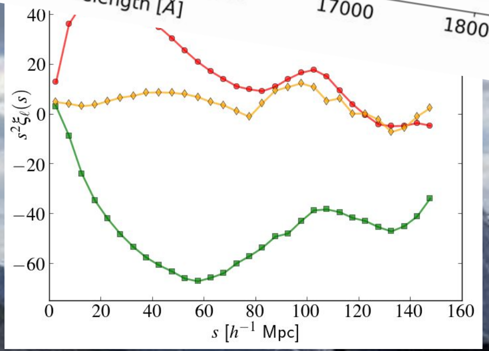
Lessons from end-to-end simulations *(or how to make Euclid data look more like the simulations.)*



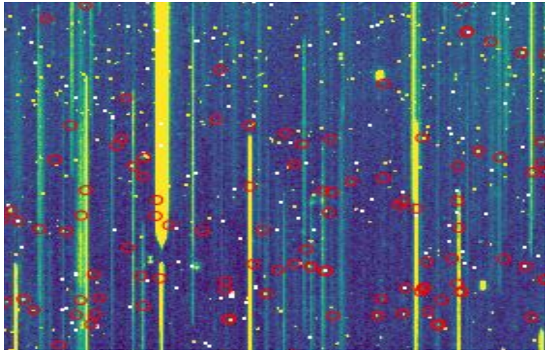


Ben Granett
INAF Brera Observatory (Milan)

Sexten Workshop
14 July 2025

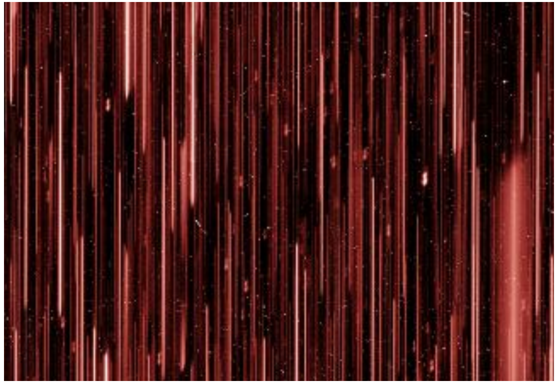
Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
INAF ISTITUTO NAZIONALE DI ASTROFISICA
euclid consortium

Credits: B. Granett & SGS



Euclid NISP-S simulated exposure with Halpha lines marked

simulated



real!

Euclid slitless spectroscopy

- All photons pass the grism (no slits or fibers)
  - No targeting required
  - Efficiency loss due to higher background
  - Emission line galaxies are main targets
- Euclid is the first large-scale application of this technique
- Slitless spectroscopy is technically simpler, but the resulting selection function is complex: confusion of adjacent spectra makes measuring redshifts more difficult in crowded areas
- Slitless spectroscopy means that almost all spectra are contaminated
- Contamination (or confusion) is biggest source of redshift failures

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
INAF ISTITUTO NAZIONALE DI ASTROFISICA
Euclid consortium

Why do end-to-end simulations?

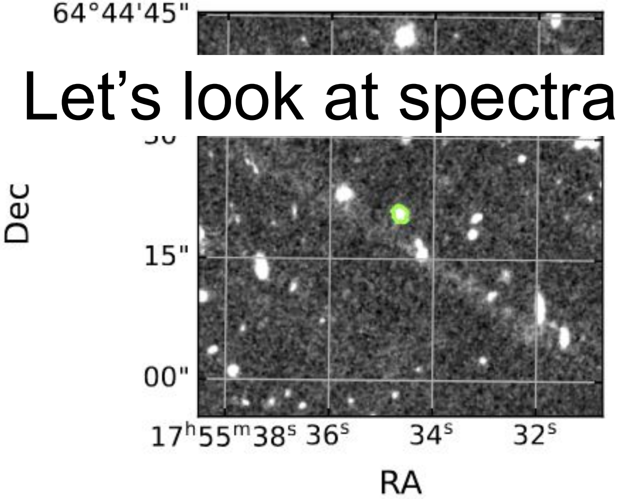
Assess survey performance and systematics (Ilaria and Francesca's talks)

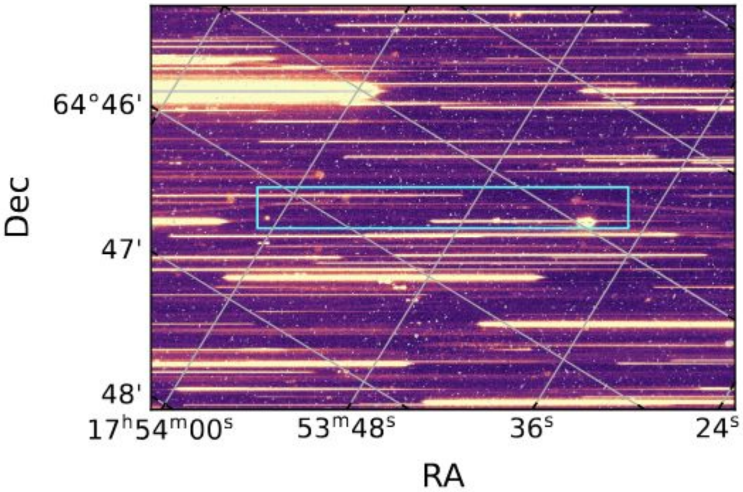
Forward model to the selection function / visibility mask (Antonio's talk)

E2E simulations

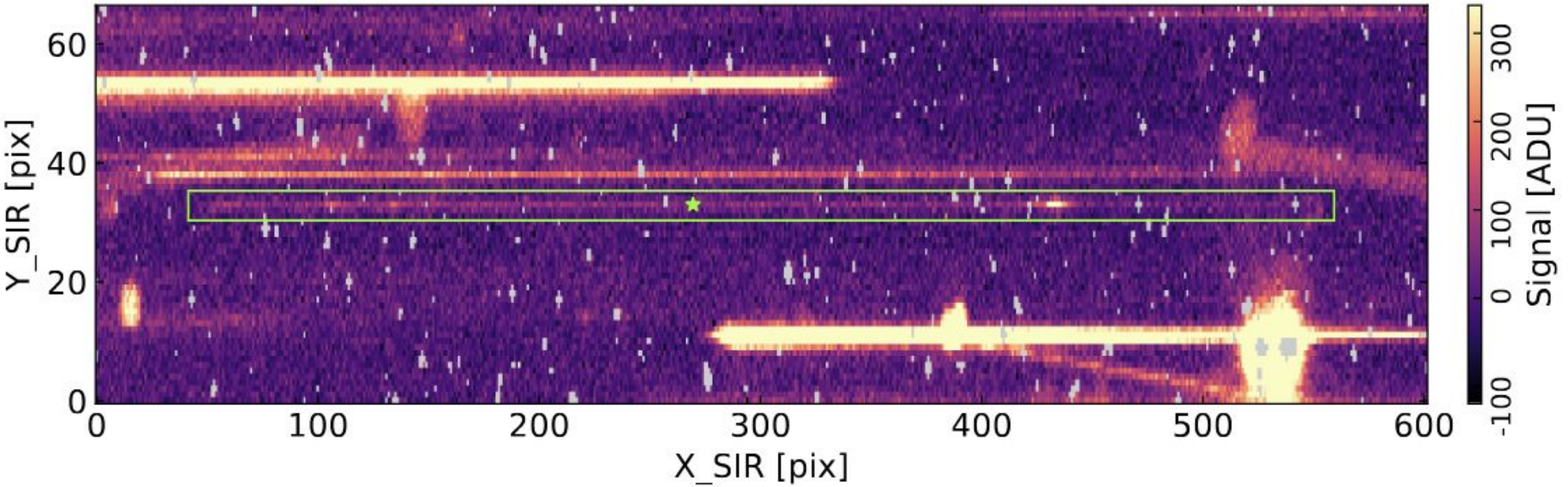
Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
INAF ISTITUTO NAZIONALE DI ASTROFISICA
euclid consortium

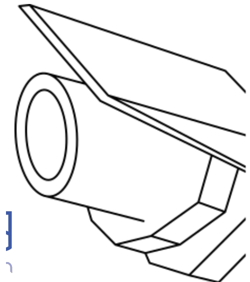
Let’s look at spectra





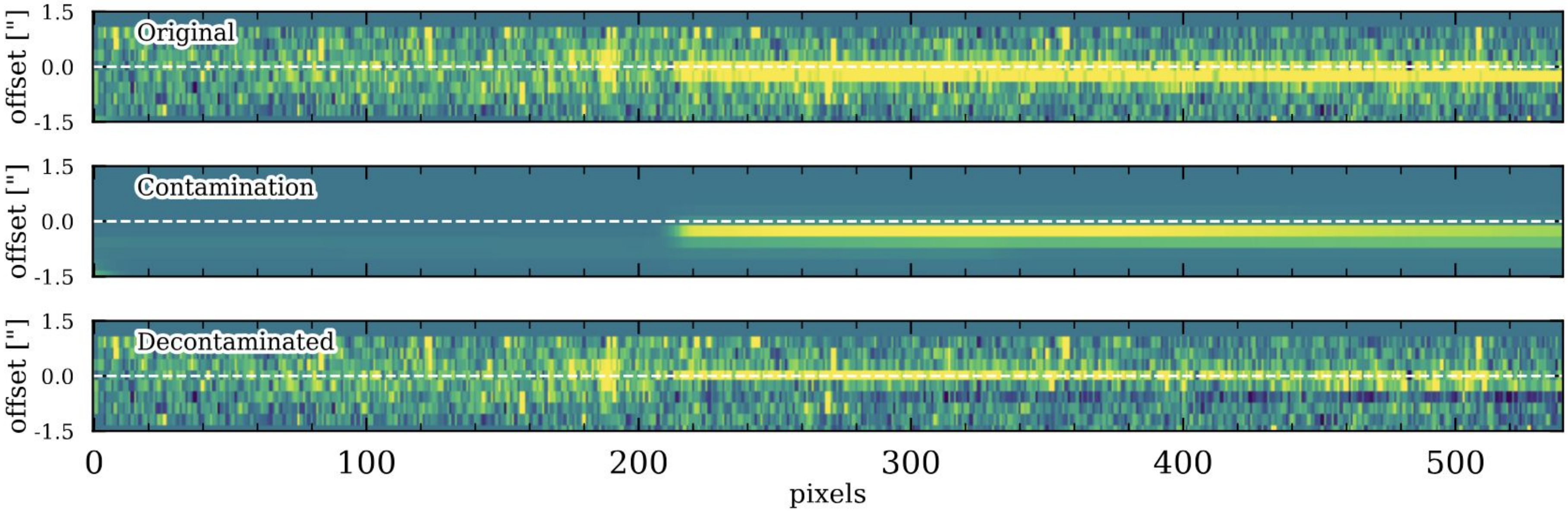
Copin+2025 Q1 SIR spectral reduction.



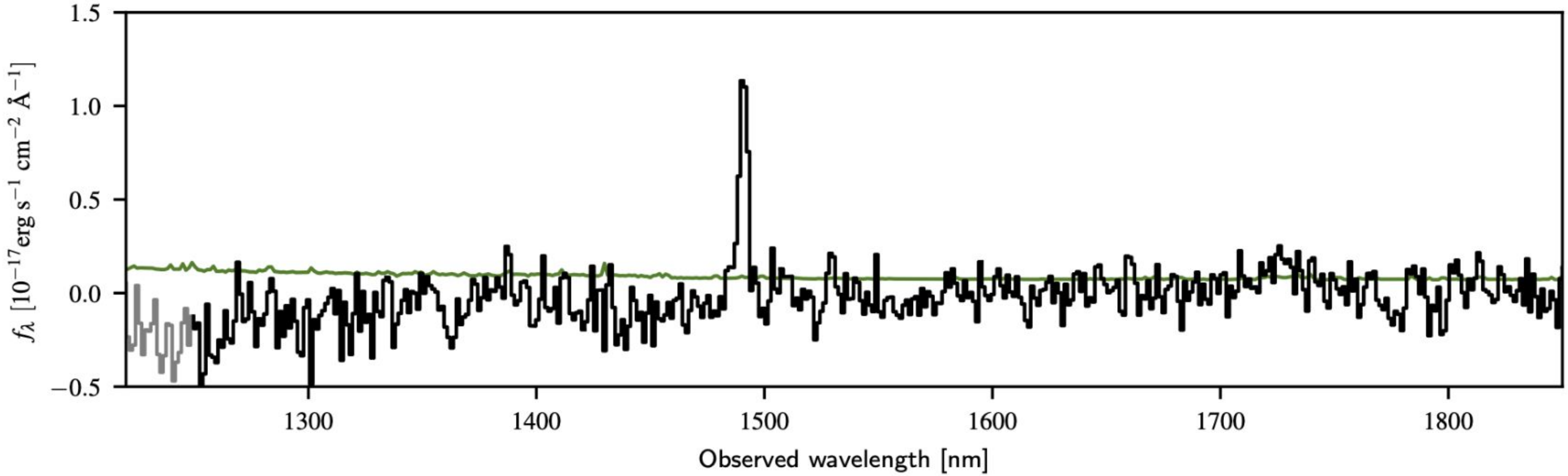


Modeling contaminants

Copin+2025

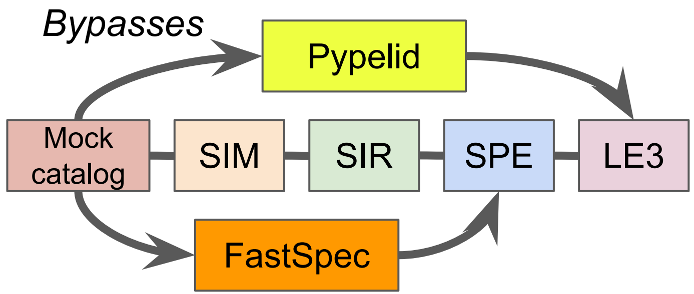


A typical spectrum



A bright Halpha line at z=1.271,
Le Brun+2025

Euclid’s end-to-end simulations



End-to-end simulations are fundamental for characterizing the survey selection function.

The Euclid consortium's **OU SIM** provides pixel-level simulations of imaging and spectroscopy.

For galaxy clustering we have developed bypass codes:

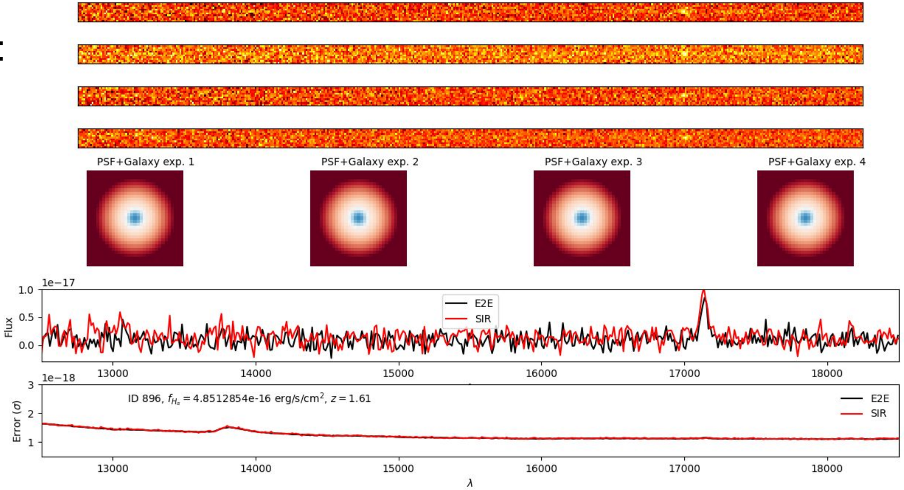
Pypelid

- ✓ Simulates emission line spectra;
- ✓ Inhomogeneous noise from survey exposures;
- ✓ Measures redshift;
- ✓ Maps SNR to detection probability;

FastSpec

- ✓ Simulates 1D/2D spectra from galaxy properties, instrument & survey configurations;
- ✓ Assumes constant background noise.

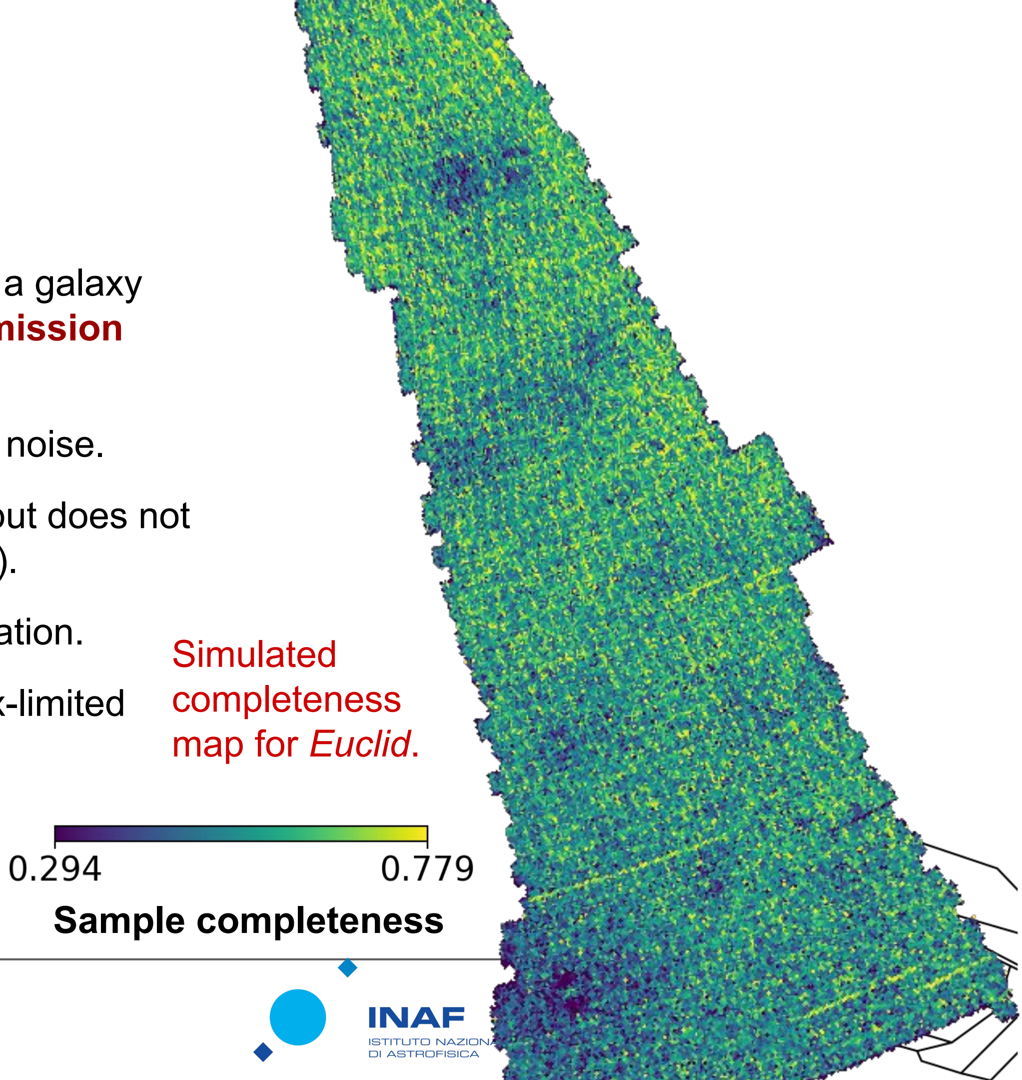
FastSpec



Pypelid in a nutshell

- The detection and redshift measurement of a galaxy depends on the **signal-to-noise ratio of emission lines**.
- Continuum emission only contributes to the noise.
- Flux from overlapping spectra adds noise, but does not change the signal (perfect decontamination).
- Python and cython code with MPI parallelization.
- The code scales to full-sky surveys with flux-limited mock galaxy samples.

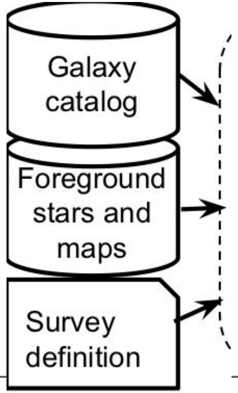
Simulated completeness map for *Euclid*.



Finanziato dall'Unione europea
NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA
INAF
ISTITUTO NAZIONA
DI ASTROFISICA

Pypelid processing steps

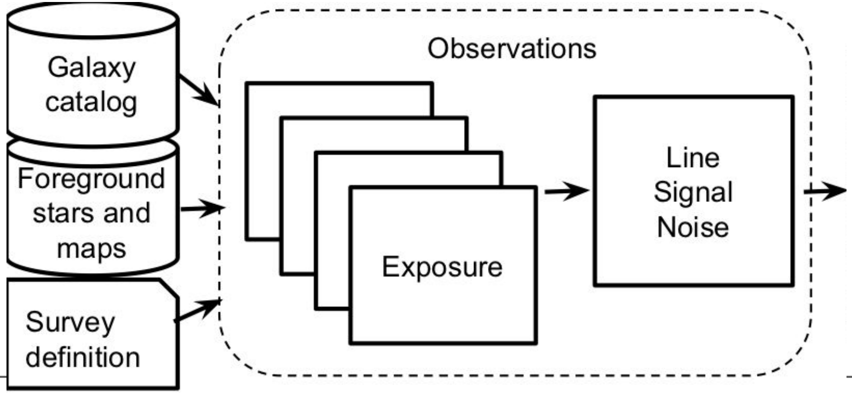
1. **Inputs:**
   a. Mock galaxy catalog is pre-processed with ***prepelid***.
   b. Star catalog may be supplied for noise contribution.
   c. List of exposures (RA, Dec, PA).
   d. Foreground maps (Zodi, straylight, Galactic extinction)
   e. Instrument configuration: transmission, PSF, spec dispersion, focal plane geometry.



Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
INAF ISTITUTO NAZIONALE DI ASTROFISICA
euclid consortium

Pypelid processing steps

2. **Simulate observations:**
   a. For each exposure, determine the position of the source in imaging and spectroscopy on the detector.
   b. Read the noise at the detector positions.
   c. Compute noise from the continuum emission of overlapping sources.
   d. Compute the signal accounting for Galactic extinction.
   e. Count number of exposures made on each emission line, and sum signal and noises.



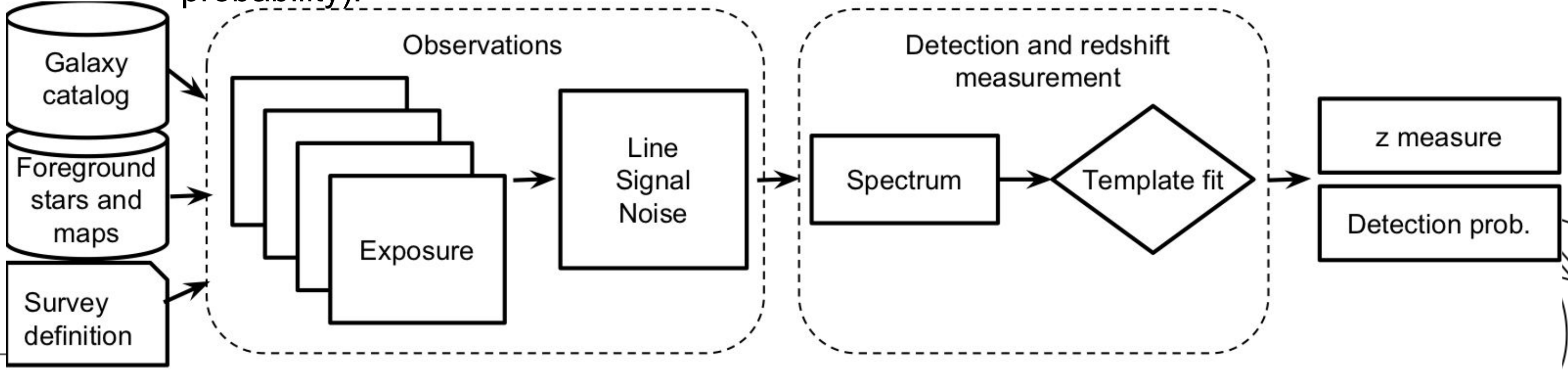
Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
INAF ISTITUTO NAZIONALE DI ASTROFISICA
euclid consortium

Pypelid processing steps

3. **Make redshift measurements (two modes implemented):**

   a. ***Template fit***: Generate a realization of the noisy spectrum and run a template fit over a redshift grid. Gives p(z) distribution and minimum chi2 solution (template fit amplitude).

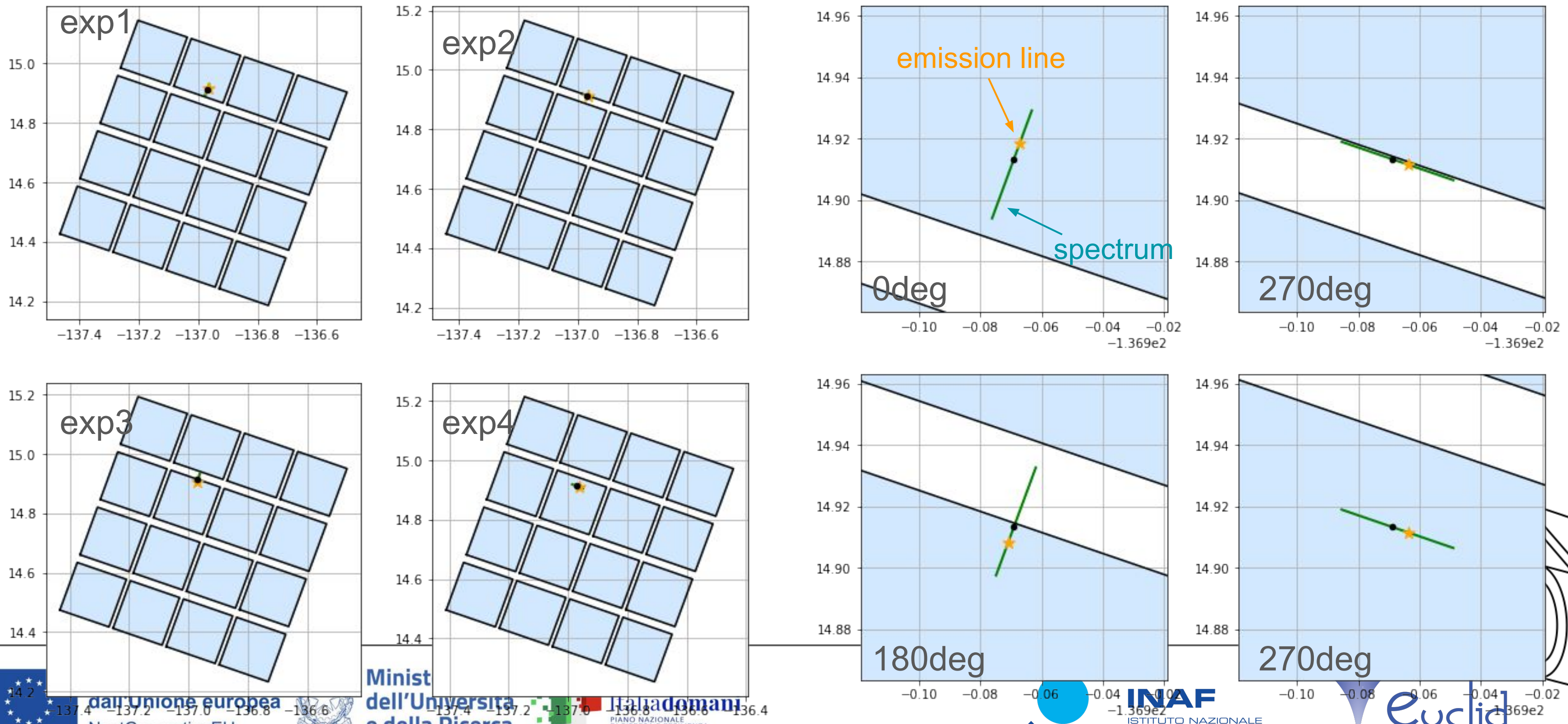
   b. ***Probabilistic***: Generate a noiseless 1D spectrum and compute SNR. Apply a calibrated detection model that maps SNR to measurement success probability (and wrong z, interloper probability).

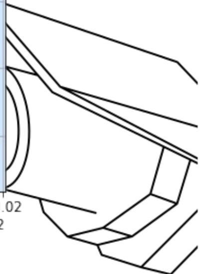


Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
INAF ISTITUTO NAZIONALE DI ASTROFISICA
Euclid consortium

Focal plane geometry

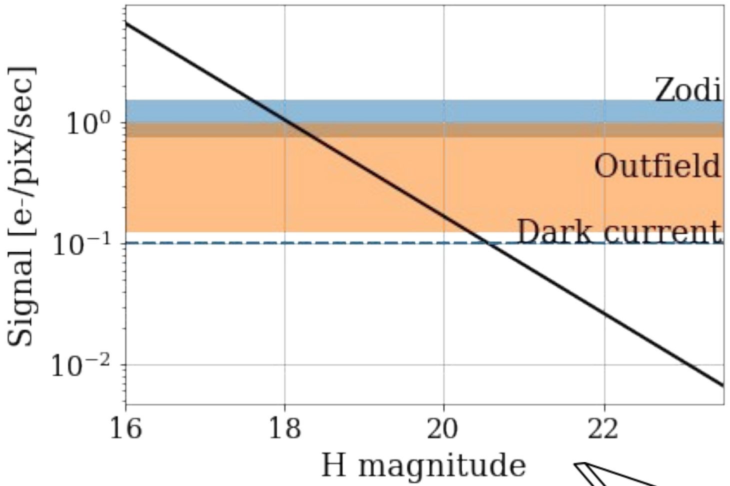
The detector plane coordinates of galaxy emission lines are determined for each exposure with a linear dispersion model.

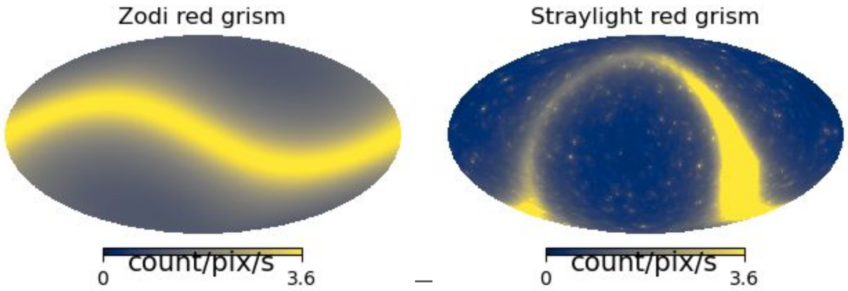




Noise sources

- The dominant backgrounds in the near IR are **zodiacal light** and **straylight** from Milky Way stars.
- Foreground galaxies and stars contaminate through **continuum emission**.
- Target galaxies at z>1 typically won't have a continuum detection.





Geometric contamination model

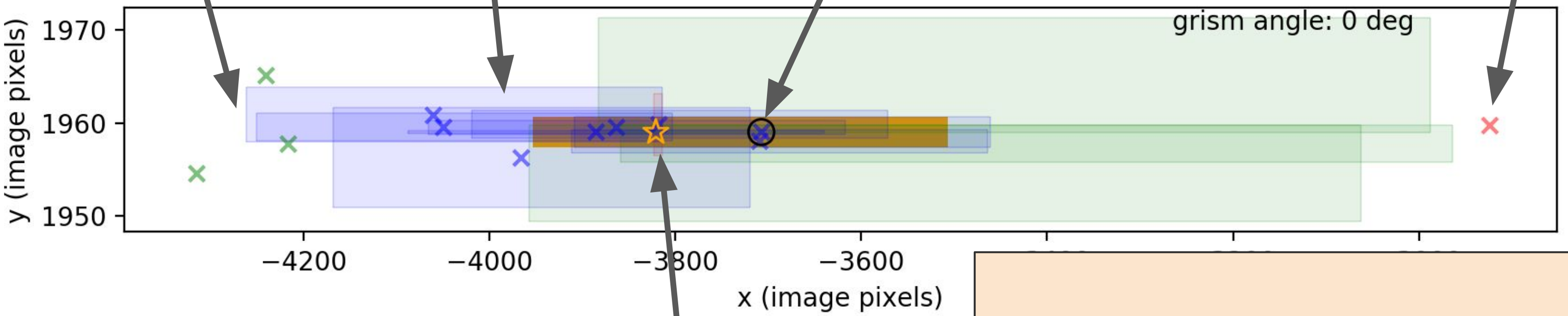
[Based on PROFESS by S. de la Torre]

Sources contributing 2nd order image

Sources contributing 1st order image

Target galaxy position

Source contributing 0th order image



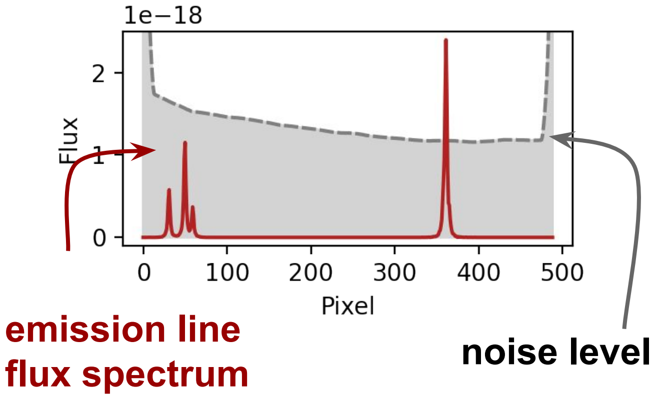
Target Ha line

Note aspect ratio is 10:1

Rectangle width is galaxy half-light radius.

Pypelid SNR definition

- Pypelid **simulates the 1D emission line spectrum** using the line fluxes, morphology, number of exposures, and noise sources.
- **Backgrounds:** zodiacal emission, outfield straylight and MW extinction.
  - *Not simulated*: persistence, cosmic rays, decontamination residuals, …
- **The SNR is computed from the sum in quadrature of S/N over the 1D spectrum.**

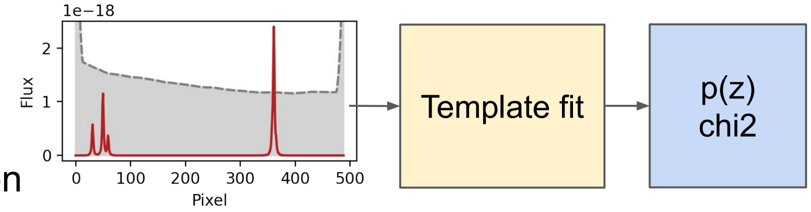


$$SNR = \sqrt{\sum_i \frac{f_{\mathrm{pix},i}^2}{\sigma_i^2}}$$

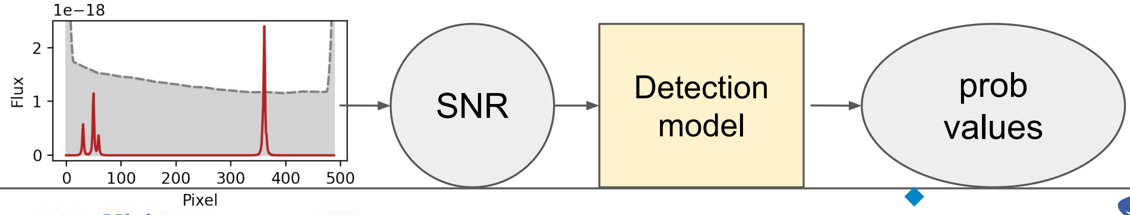
Pypelid detection model

Pypelid has two modes to simulate redshift measurements:

1. **Template fit:** Generate a realization of the noisy spectrum and run a template fit over redshift grid.
   a. Gives p(z) distribution and minimum chi2 solution (template fit amplitude).



2. **Probabilistic:** Apply a calibrated detection model that maps SNR to measurement success probability (and wrong z, interloper probability).



Finanziato dall'Unione europea
NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA
INAF
ISTITUTO NAZIONALE DI ASTROFISICA
Euclid consortium

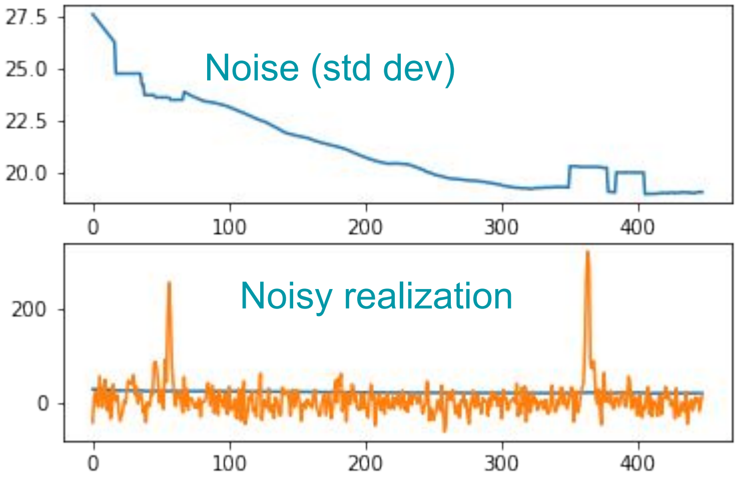
1. *z* measurements with template fits

Windows with pypelid background estimate

- *Pypelid* implements a simple template fit code to measure the redshift on simulated spectra (emission lines only).

- The output includes the *p(z)* distribution and template fit amplitude (*chi2*). The amplitude can be used to quantify the quality of fit and control the spurious detection rate.

- The implementation is made to run fast and has not been compared with more sophisticated redshift fitting codes (AMAZE, EAZY…) with full sample selection criteria.
  Don't expect it to reproduce completeness and interloper rates!



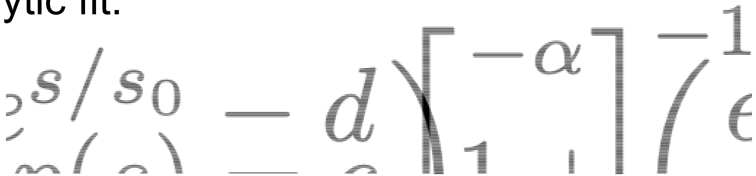
z~1.7

$$a = \frac{D\sigma^{-2}T}{T\sigma^{-2}T}$$

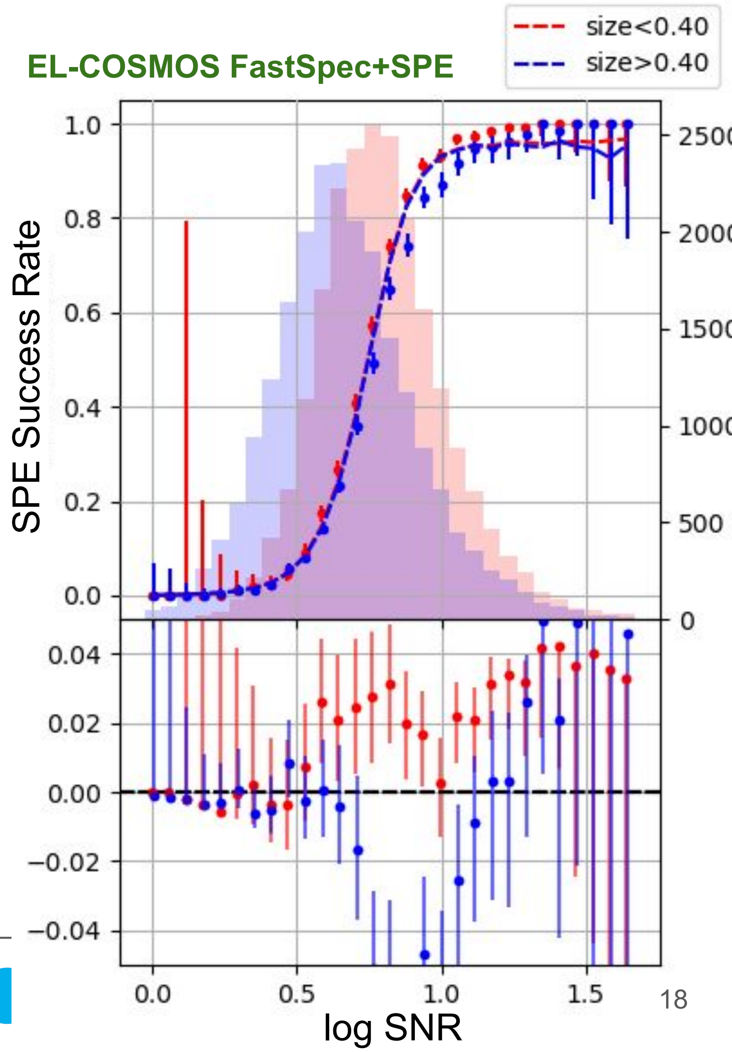
Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
INAF ISTITUTO NAZIONALE DI ASTROFISICA
euclid consortium

2. Probabilistic detection model

- We calibrate a detection model using simulations run through the *Euclid* pipeline (SPE redshift measurement only in this case).
- The detection model maps SNR to ***SPE success probability*** for a given sample selection *(SPE_PROB>0.95)*
- Analytic fit:



- **This allows us to reproduce the purity and completeness of the *Euclid* pipeline for particular galaxy samples.**



Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

The spherical cow meets a real cow

Bypass sims:

- Signal and noise are perfectly separated.
- Well-behaved noise can be summed in quadrature.
- The selection function is parametrized by one
- parameter: total SNR of emission lines.

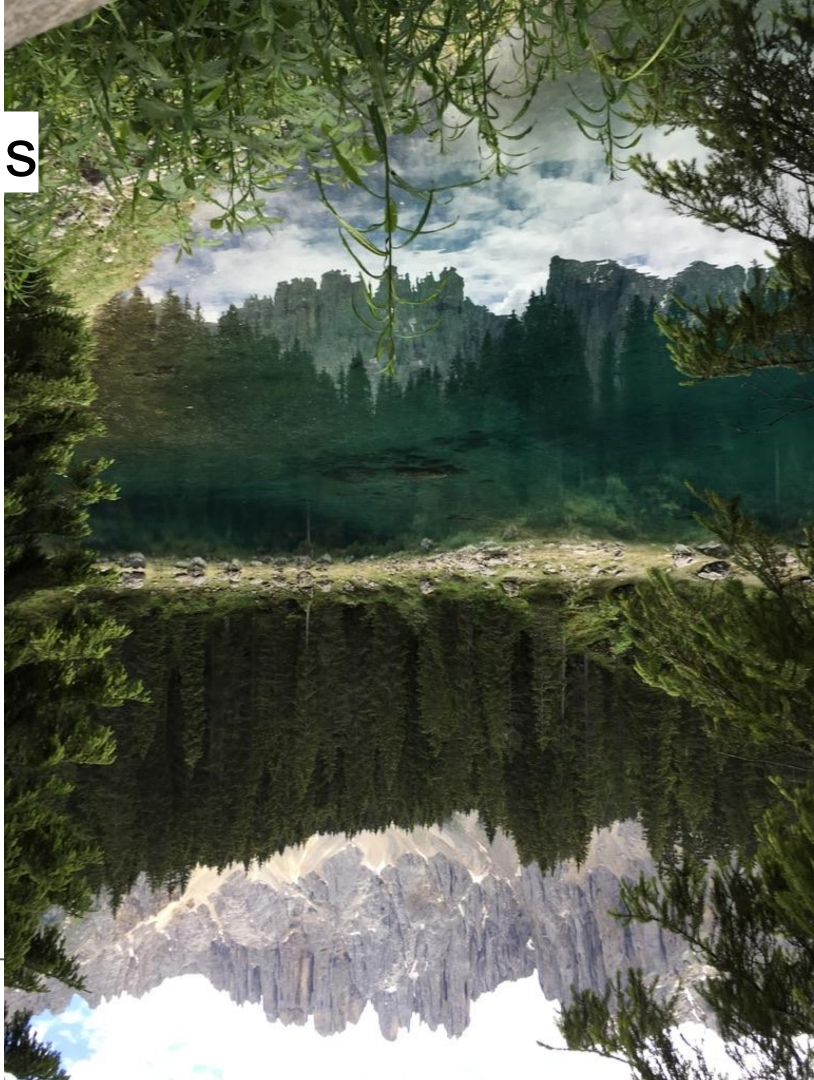
Real data:

- Systematics corrupt the signal and noise estimates.
- Necessitates the use of median combinations and sigma clipping to extract the signal.
- Redshift measurement depends on a complex fit with poorly characterized noise.
- Results in a complex selection function, difficult to assess.

Franco

Make real data look like simulations

- Can we use simplifying assumptions of E2E simulations on real data?
    - Focus on emission lines. Discard continuum signal to approximate perfect decontamination.
    - Use matched filter to identify emission lines in 1D or 2D continuum-subtracted images.
- Alternative methods provide validation of measurements and informs simulations.



Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA