FORECAST: a flexible software to forward model cosmological hydrodynamical simulations mimicking real observations

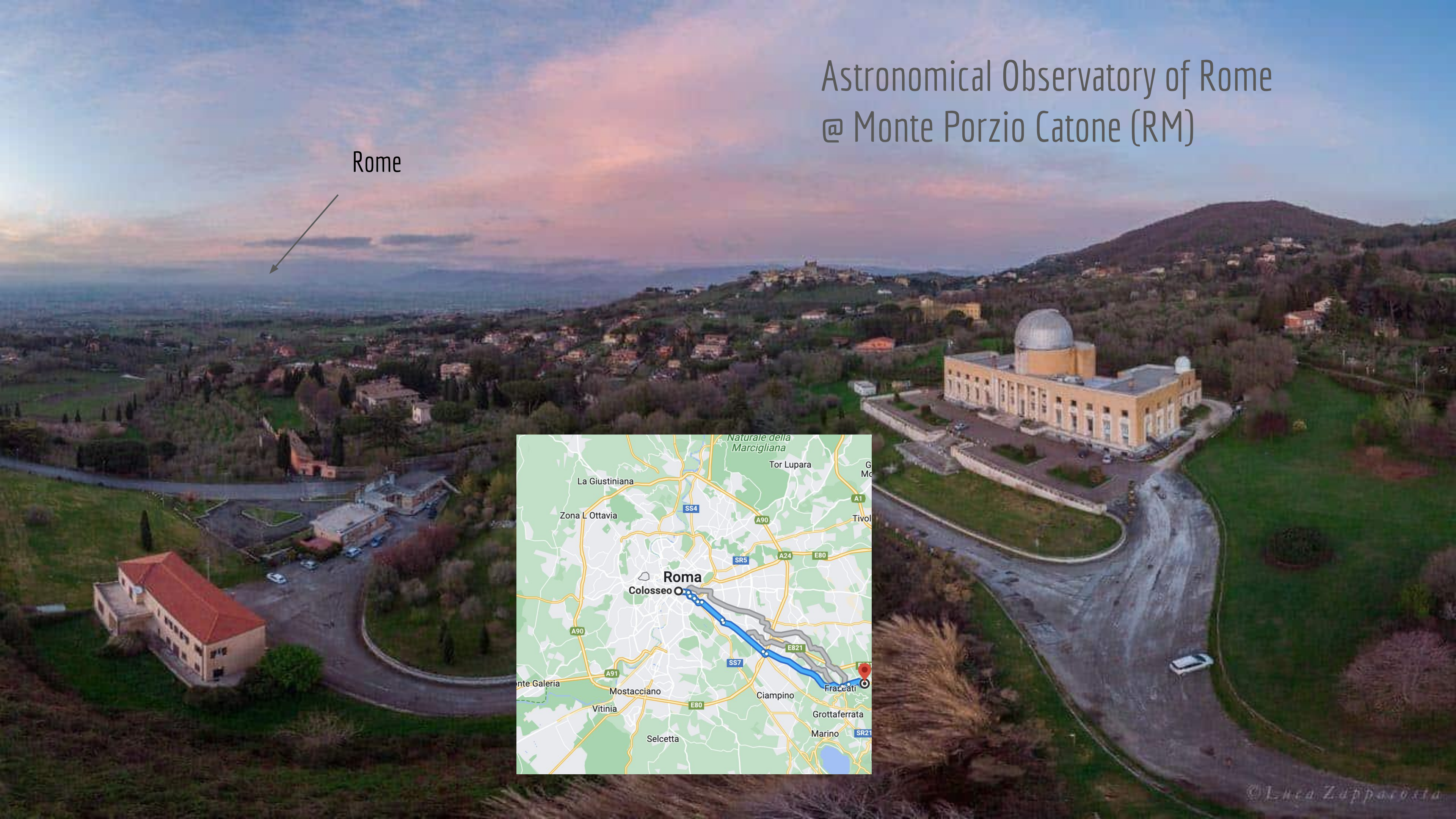
Flaminia Fortuni

she/her · flaminia.fortuni@inaf.it

post-doc at INAF - Osservatorio Astronomico di Roma

Perspectives on Large Scale Structure
27-30 June in Prague, Czech Republic

Astronomical Observatory of Rome
@ Monte Porzio Catone (RM)



extragalactic group

http://www.caesar-astro.it/

galaxy formation & evolution, high-z galaxies, image analysis


Emiliano Merlin


Adriano Fontana


Paola Santini


Laura Pentericci


Marco Castellano


Antonello Calabrò

other collaborators

Carlo Giocoli


Luca Graziani

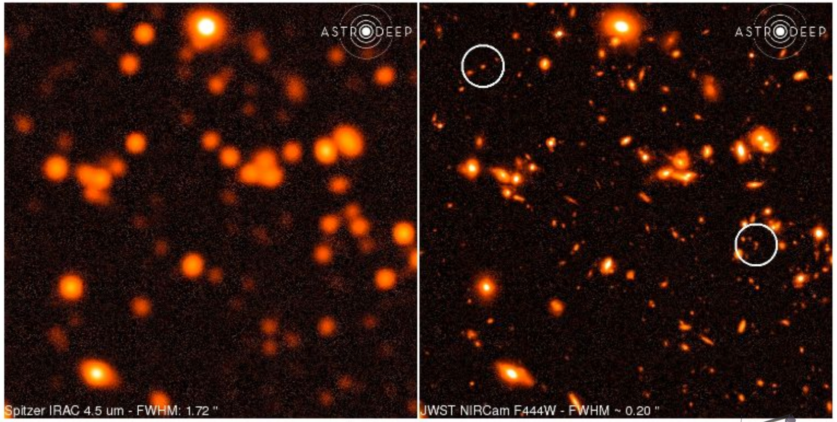

Erik Romelli


Stephane Charlot


The next-generation surveys

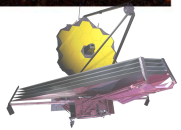
PROBING THE HIGH-z UNIVERSE TO AN UNEXPLORED EXTENT

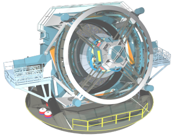
- improving data quality
- reducing photometric uncertainties
- vastly enhanced statistics (big-data)
- depth up to redshift $z>10$
- wide range of wavelengths



Credits: D.Paris, A.Fontana, E.Merlin

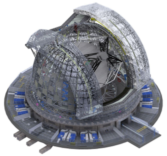
http://www.astrodeep.eu/movies-jwst1/ for the full video





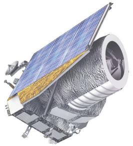
VERA RUBIN OBSERVATORY (LSST)

- **dark Universe**; small bodies in the Solar System to outer regions of the Milky Way and nearby **galaxies**; transient phenomena



EXTREMELY LARGE TELESCOPE (ELT)

- extra-solar planets; resolved stellar populations; **high-redshift galaxies**; fundamental physics and **Cosmology**



EUCLID

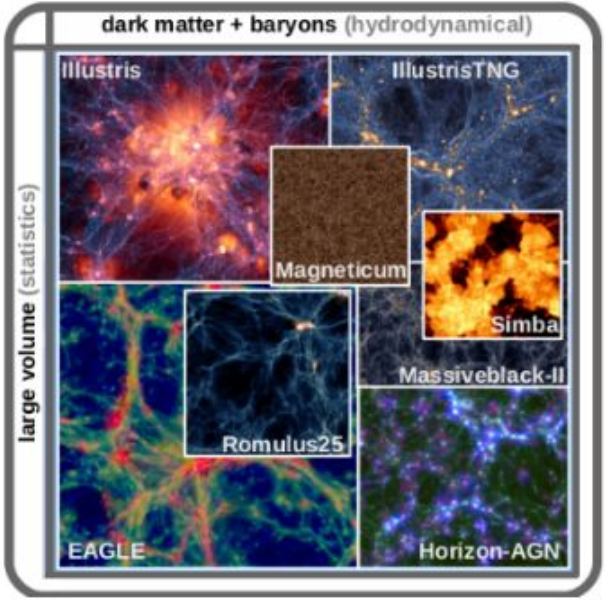
- geometry of the **dark Universe** with distance-redshift relationship and **evolution of cosmic structures**

NANCY GRACE ROMAN TELESCOPE (WFIRST)

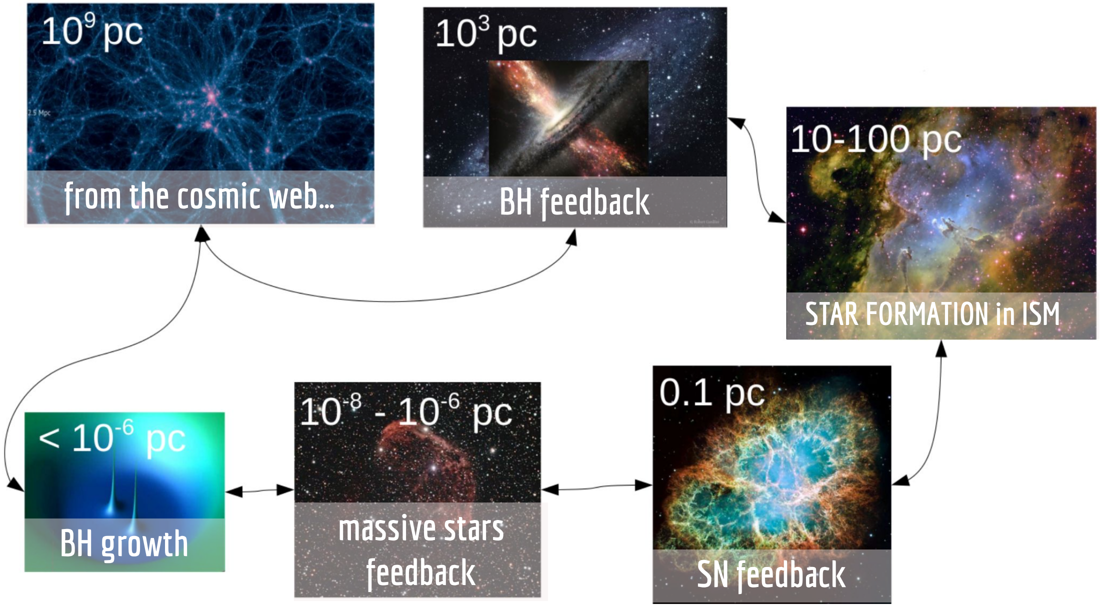
- **dark Universe**; exoplanets; first generations of **galaxies and quasars**

State-of-the-art hydrodynamical simulations

DARK MATTER & BARYONS DYNAMICS + **SUB-RESOLUTION PHYSICAL MODEL**

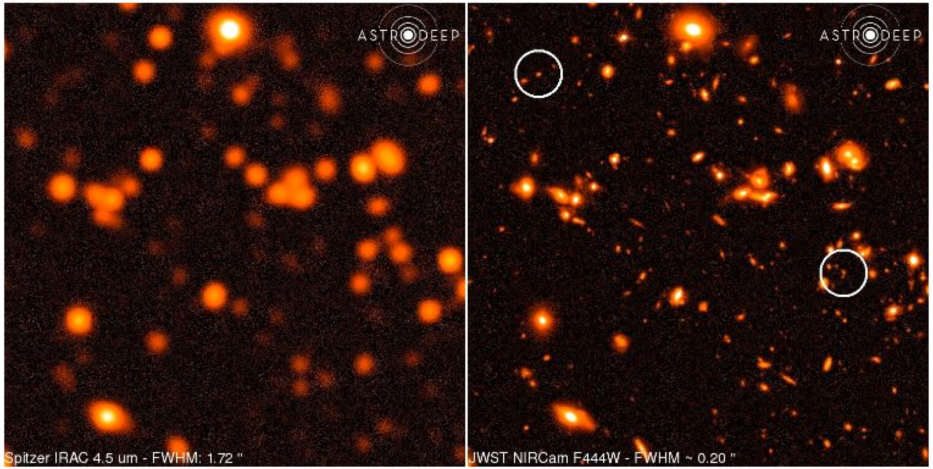


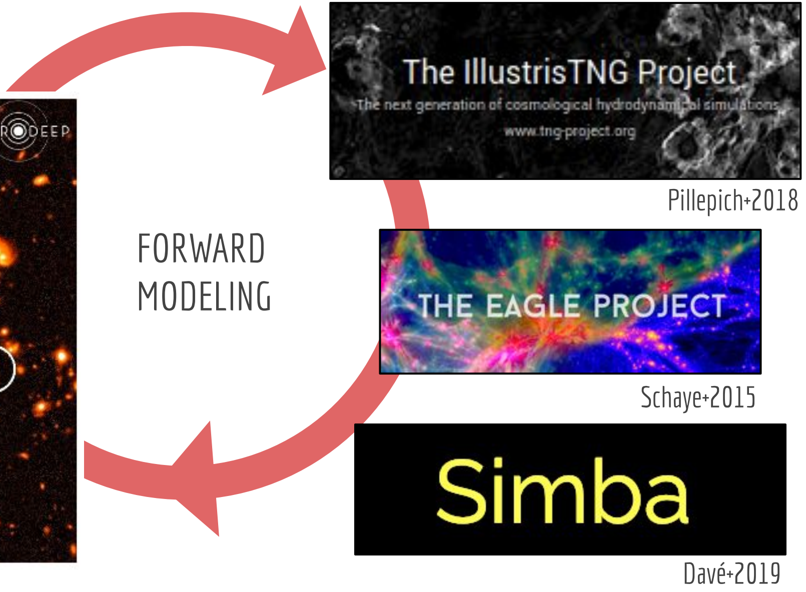
adapted from Vogelsberger+2020



adapted from D.Sijacki

The next-generation surveys in synergy with the next-generation simulations



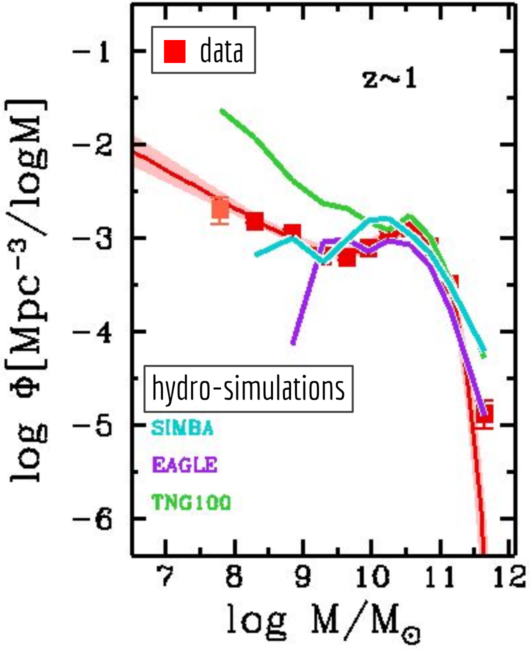


Credits: D.Paris, A.Fontana, E.Merlin

http://www.astrodeep.eu/movies-jwst1/ for the full video

Comparison between observations and simulations

PASSIVE GALAXY STELLAR MASS FUNCTION



Santini, FF +2022

INDIRECT APPROACH

in simulations: sum of the stellar masses of all the stellar resolution elements belonging to a subhalo (simulated galaxy)

in observations: inferred from observations: e.g. observed flux and simplified assumptions in SED-fitting

Forward modeling of hydrodynamical simulations

models are used to simulate observations

- the underlying physics is known
- hydro-sims reproduce real galaxy complexity

hot method in extragalactic astronomy!

2022ApJ...926..194D 2022/02 cited: 2
Deep Realistic Extragalactic Model (DREaM) Galaxy Catalogs: Predictions for a Roman Ultra-deep Field
Drakos, Nicole E.; Villasenor, Bruno; Robertson, Brant E. *and 9 more*

2020MNRAS.499.5702B 2020/12 cited: 25
The Universe at z > 10: predictions for JWST from the UNIVERSEMACHINE DR1
Behroozi, Peter; Conroy, Charlie; Wechsler, Risa H. *and 8 more*

2023MNRAS.518.6318S 2023/02 cited: 1
Mock galaxy surveys for HST and JWST from the IllustrisTNG simulations
Snyder, Gregory F.; Peña, Theodore; Yung, L. Y. Aaron *and 3 more*

2021MNRAS.502.4858S 2021/04 cited: 9
Mock light-cones and theory friendly catalogues for the CANDELS survey
Somerville, Rachel S.; Olsen, Charlotte; Yung, L. Y. Aaron *and 9 more*

2021MNRAS.501.1591P 2021/02 cited: 13
Realistic mock observations of the sizes and stellar mass surface densities of massive galaxies in FIRE-2 zoom-in simulations
Parsotan, T.; Cochrane, R. K.; Hayward, C. C. *and 5 more*

...and more

FORECAST

- FORECAST makes synthetic **astronomical images** between **rest-frame UV to NIR**
  flexible: *all* hydrodynamical simulations
  adaptable: *all* imaging observations, arbitrary FoV, filters and depth

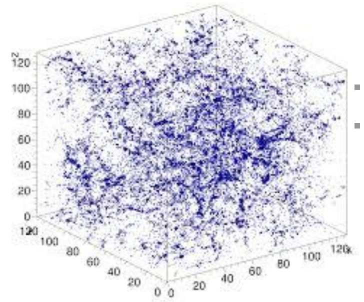
- making straightforward comparisons with present-day/next-generation surveys and improving theoretical models

- testing observational biases (selection of the survey, techniques of source extraction, deblending, physical models assumed in the SED-fitting stage)

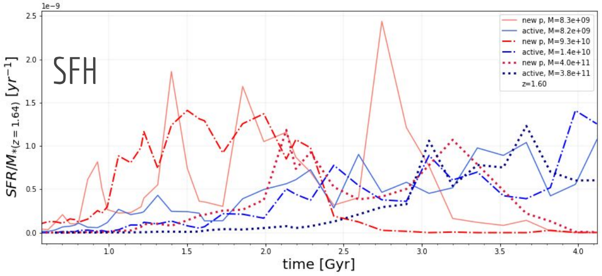
Fortuni et al. subm. AA/2023
arXiv 2305.19166
www.astrodeep.eu/FORECAST

FORECAST

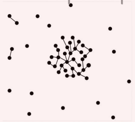
HYDRODYNAMICAL SIMULATION → particle-based FORWARD MODELING → MOCK OBSERVATORY



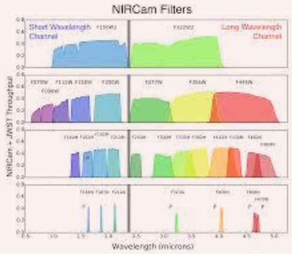
- large-volume
- full hydrodynamics

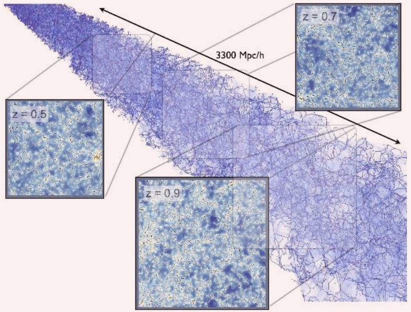


post-processing of the simulation:

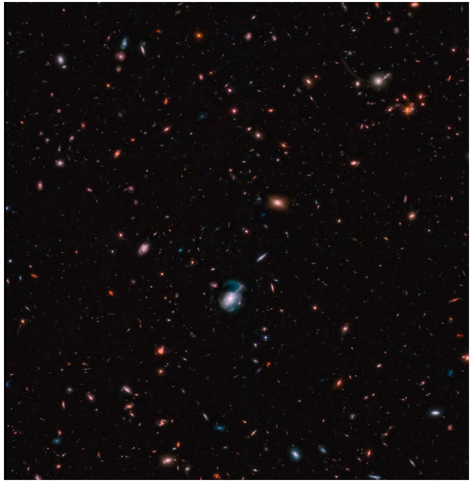


stellar particles as galaxy tracer
gas cells as dust tracer





imaging from rest-frame **ultraviolet** to **near-infrared** bands

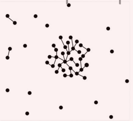


crop of *JWST*/CEERS
Credits: https://ceers.github.io/

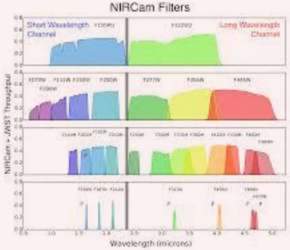
FORECAST

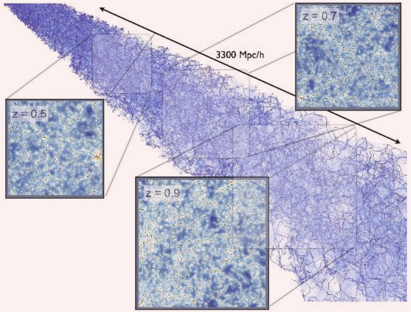
particle-based
FORWARD MODELING

post-processing of the simulation:



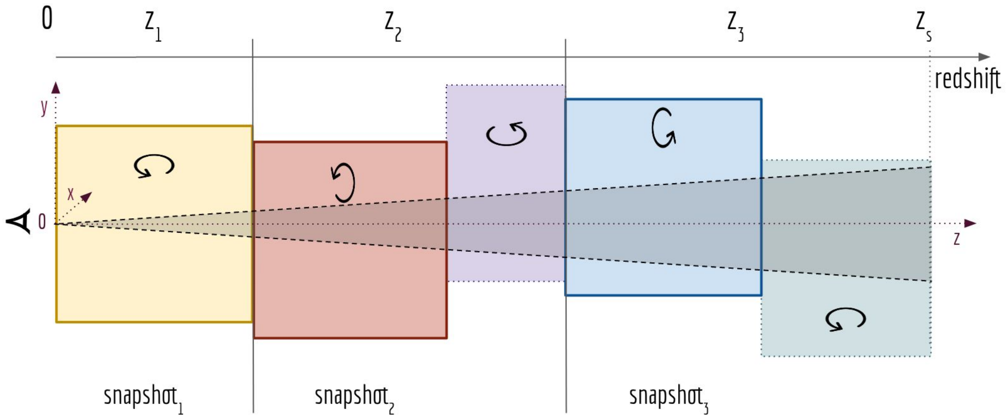
stellar particles as galaxy tracer
gas cells as dust tracer

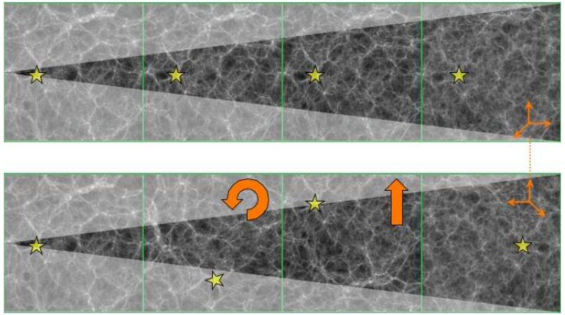




Light-cone construction

Giocoli+2015

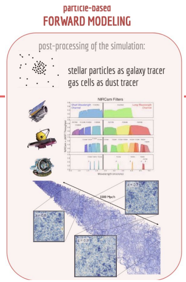




Blaizot+05

- **random tiling** to avoid replication effects
- **selection in FoV** with the line-of-sight in the center of the box

Let there be light...



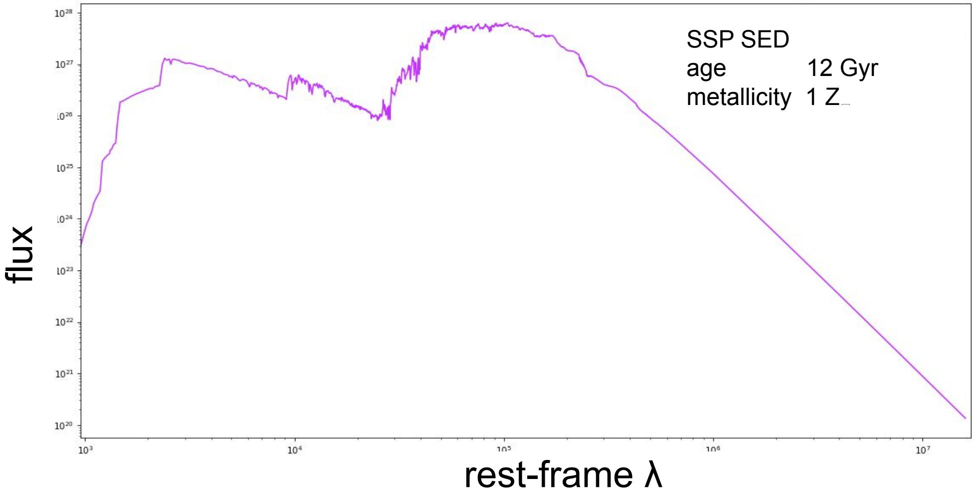
stellar particles = Single Stellar Populations at this resolution (1e6-1e7 M☉)

rest-frame SSPs SEDs

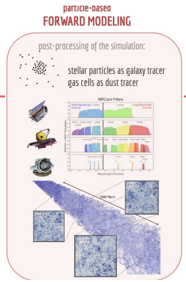
(Bruzual & Charlot 2003; Gutkin+2016)

k-correction → observer-frame fluxes

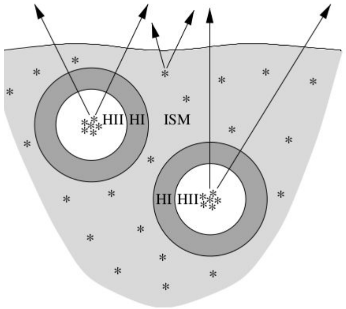
IGM absorption (Inoue+2014)



...and dust



gas cells/particles



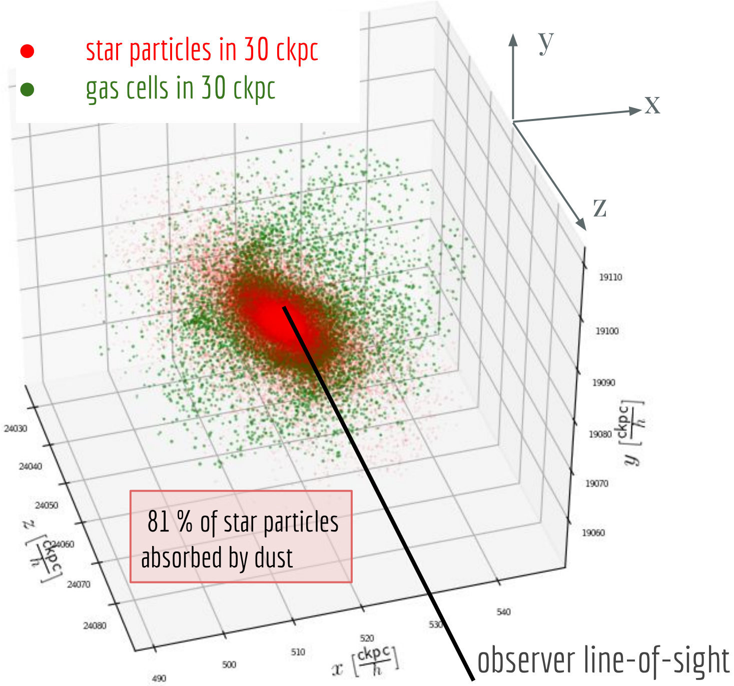
“spatially unresolved” dust attenuation

(Charlot & Fall 2000)

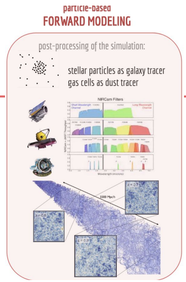
“spatially resolved” dust attenuation

(Devriendt & Guiderdoni 2000, Nelson+2019)

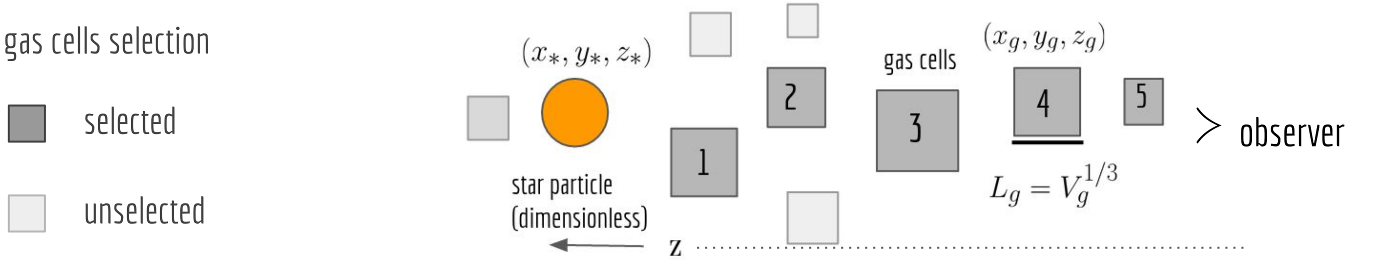
$$\tau^a_\lambda = \left(\frac{A_\lambda}{A_V}\right)_\odot (1+z)^{-0.5} \left(\frac{Z_g}{Z_\odot}\right)^\gamma \left(\frac{N_{HI}}{N_{HI,0}}\right)$$



“Resolved” dust attenuation



(i) mapping gas distribution (N_{HI} and Z_{gas}) around each galaxy, along the line-of-sight of each SSP



(ii) computing mean dust attenuation (due to large computational times)

$$< Z_g >_{gal} = \frac{\sum_i Z_{i,g} \cdot M_{i,HI}}{\sum_i M_{i,HI}}$$

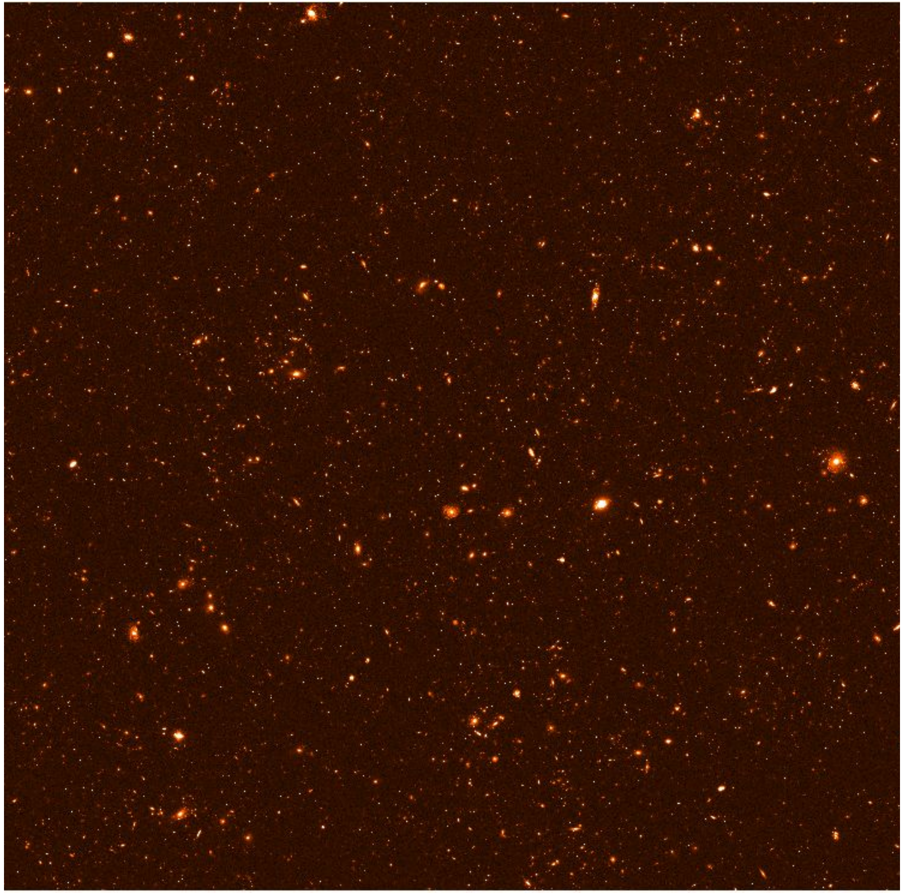
$$< N_{HI} >_{gal} = \frac{\sum_i N_{i,HI} \cdot M_{i,HI}}{\sum_i M_{i,HI}}$$

$$\tau_\lambda^a = (\frac{A_\lambda}{A_\odot})(1+z)^\beta (Z_g/Z_\odot)^\gamma (N_{HI}/N_{HI,0})$$

applied to dust-free SED

$$L_{obs,SSP}(\lambda) = L_{em,SSP}(\lambda) e^{-\tau_\lambda^a}$$

Testing FORECAST emulating the CANDELS GOODS-South field



Final simulated light-cone in H160 band (PSF and noise added in post-processing)

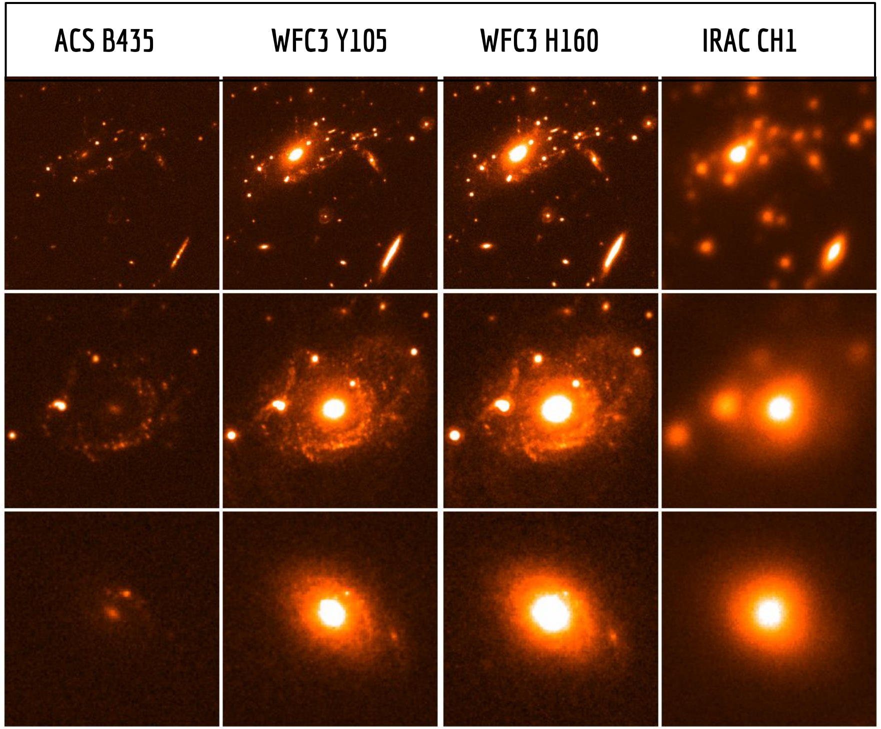
light-cone between $0.1 \leq z \leq 7.0$
field of view 200 $arcmin^2$, with pixel scale 0.06 arcsec

ASTRODEEP catalogue from **CANDELS GOODS-South field** (Merlin+2021) exploiting **IllustrisTNG100** (Weinberg+2017, Pillepich+2018)

post-processing:

- instrumental PSF + bkg gaussian noise + shot noise
- RMS map

Testing FORECAST emulating the CANDELS GOODS-South field



Example of small areas containing a group and single objects, in 4 simulated bands (in μJy; light-cone with PSF and noise added in post-processing)

simulated sources have realistic morphologies!

GLOSSARY

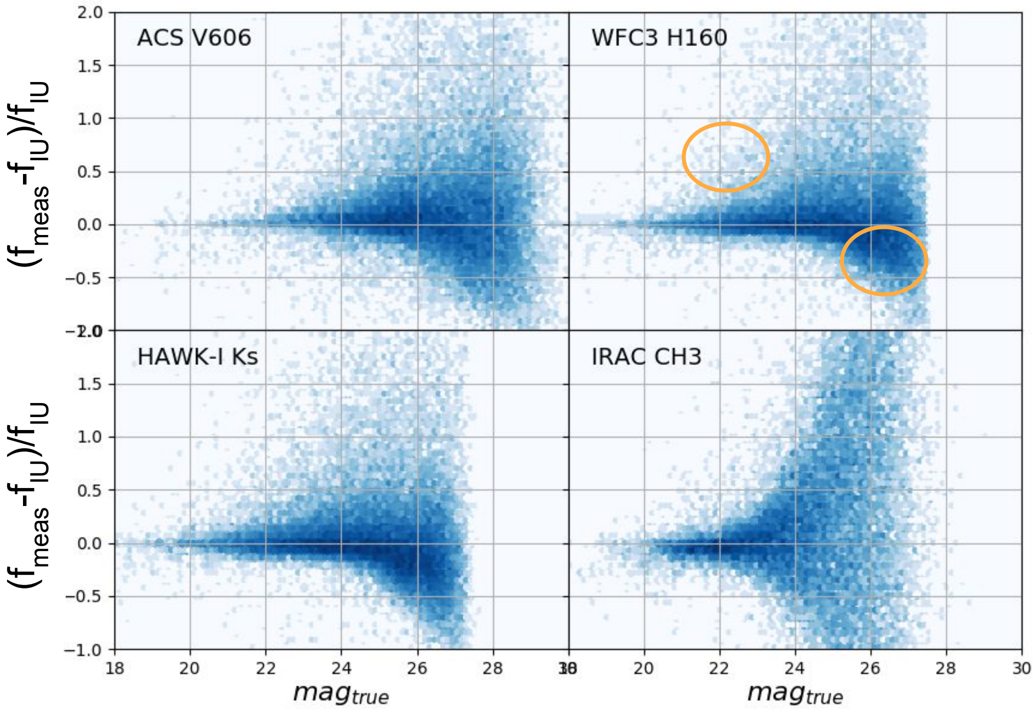
Input Universe or IU: properties of simulated galaxies in the Input Universe

measured: properties of simulated galaxies *measured* from the mock images with photometric techniques

Test1: photometry

A-PHOT, Merlin+2019
T-PHOT, Merlin+2015,16b

df/f in 13 bands on sources matched with IU



$$df/f = (f_{meas} - f_{IU})/f_{IU}$$

- Average flux well recovered

- Bright outliers

contamination with neighbourhood sources

- Declining trend at faint mag

measurements on detection band (*H*160)

Estimation of galaxy properties from astronomical images: SED-fitting

FILTERS

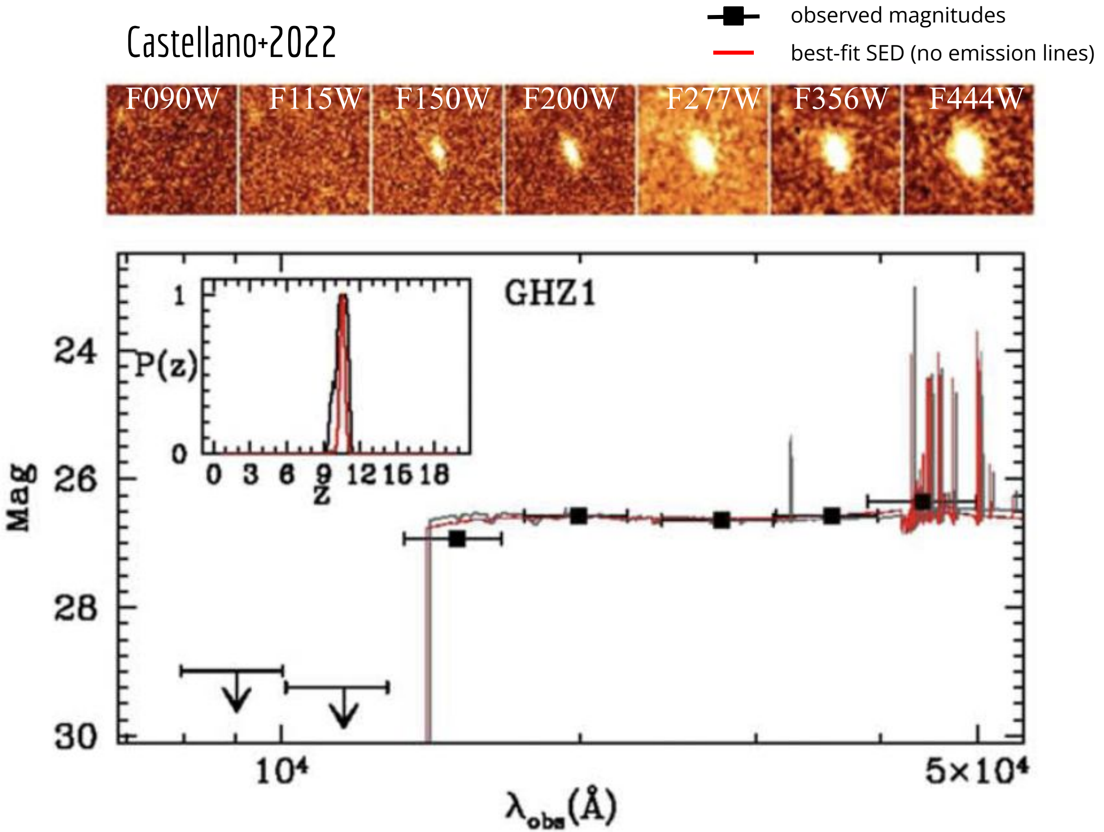
PSF

PHOTOMETRY EXTRACTION

LIBRARY OF TEMPLATES:

- stellar evolution models
- IMF
- star formation history
- metallicity enrichment
- IGM absorption
- dust attenuation

Castellano+2022

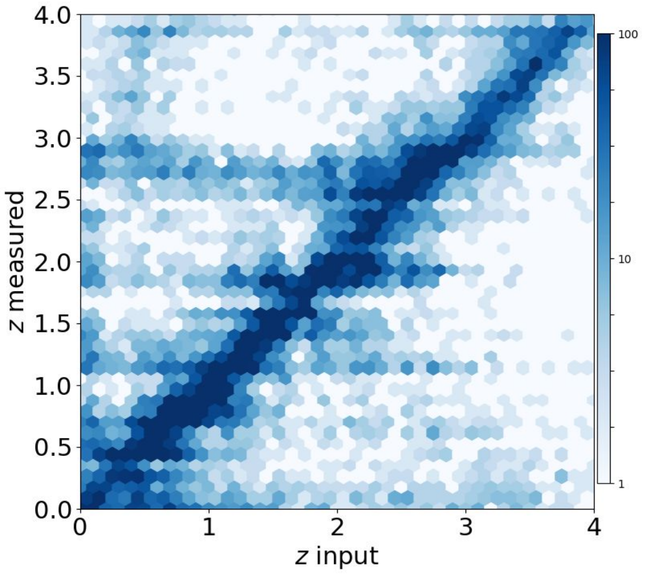


Test2: photometric redshifts

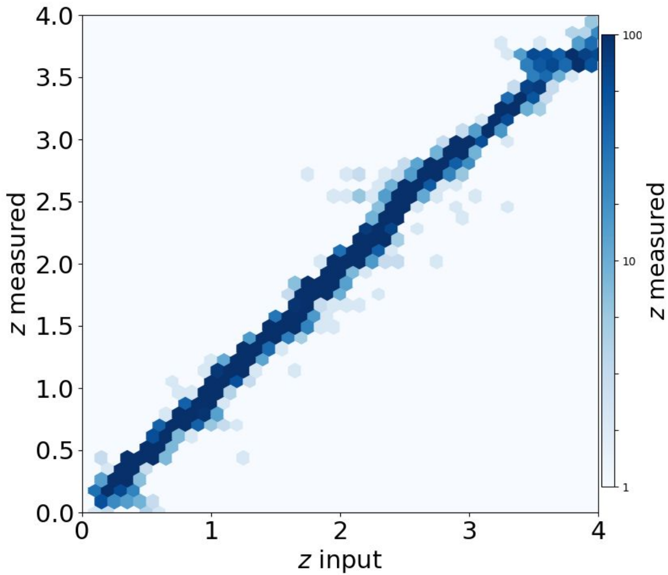
dataset	dz	outliers (%)
ASTRODEEP 13 bands	-0.003 ± 0.055	21.47
z-phot TU fluxes	-0.012 ± 0.021	0.21
z-phot measured fluxes	-0.011 ± 0.055	25.2

zphot, Fontana+2000

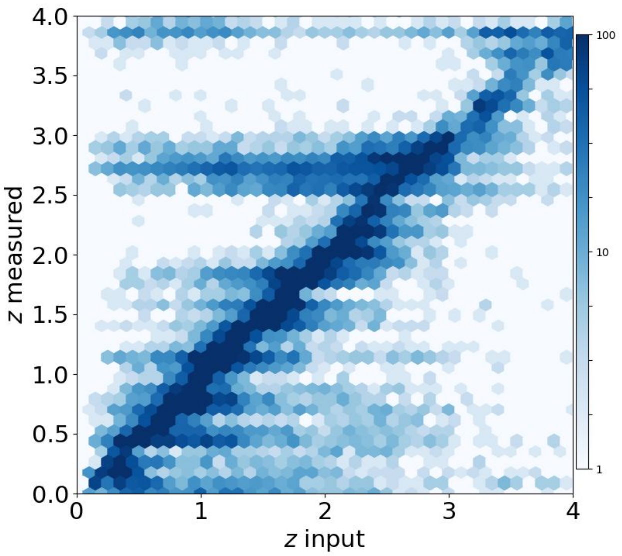
ASTRODEEP data in 8 bands



z (at fixed error, no noise)

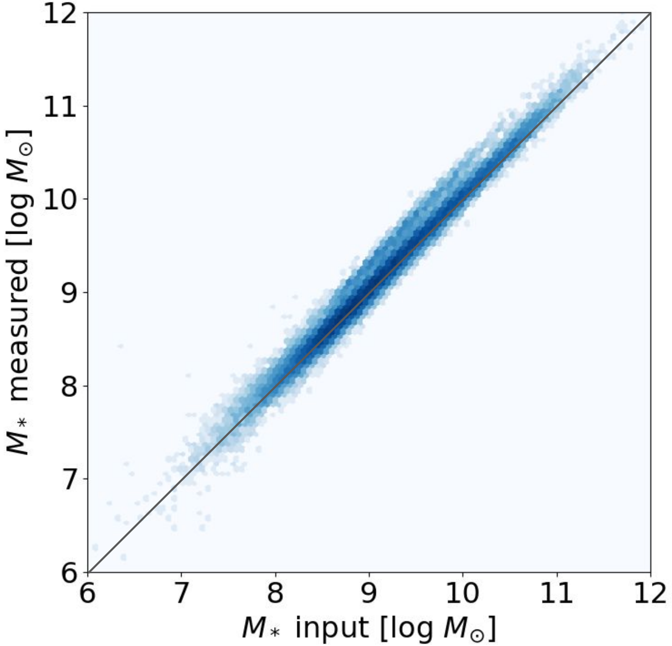


z (measured fluxes)

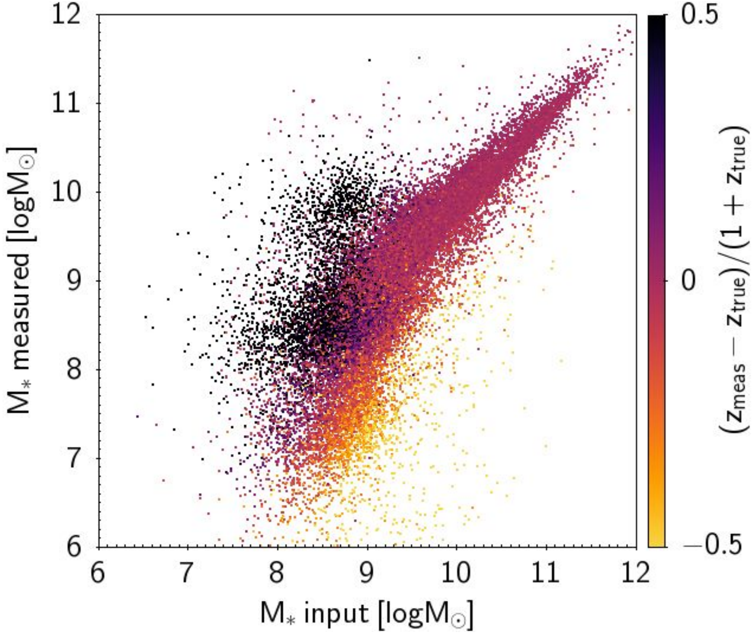


Test n-th: stellar mass

MASS (at correct redshift, no noise)

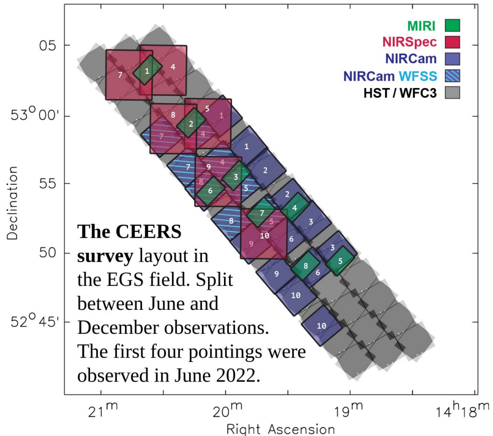


MASS (noise+photo-z)



JWST mock dataset

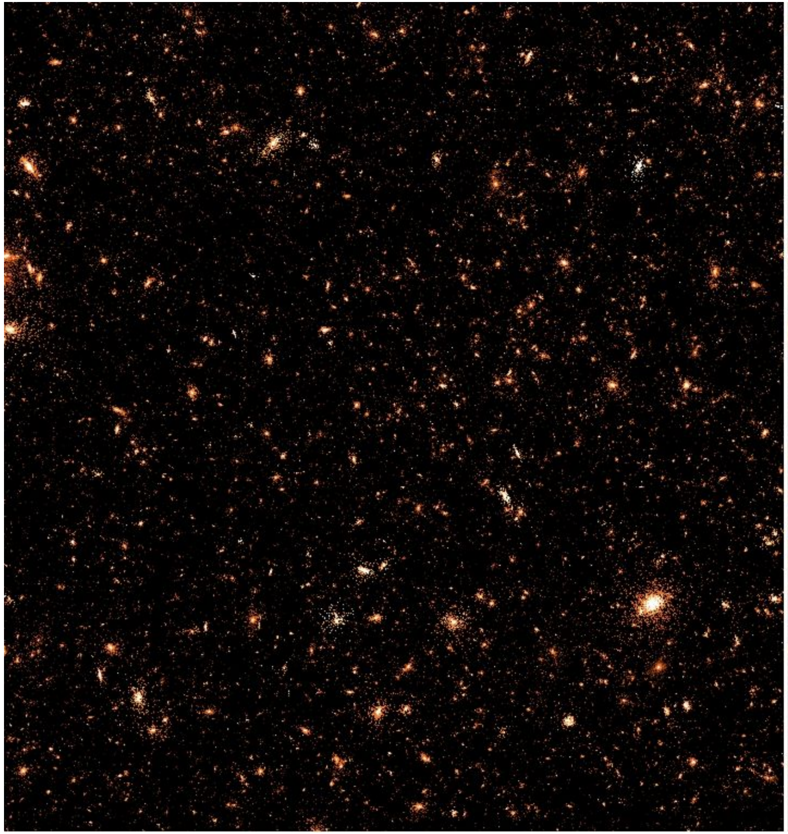
- light-cone at z=0-20 with IllustrisTNG
- modeled with SEDs from Gutkin+2016 (with nebular lines)
- 8 NIRCam filters (f090w, f115w, f150w, f200w, f277w, f356w, f410m, f444w)
  2 MIRI filters (f560w, f770w)
- 5σ depth from *JWST* CEERS/*NIRCam* (Finkelstein+2022)
  JWST CEERS/*MIRI* (Papovich+2022)
  JWST GLASS/*NIRCam* (Merlin+2022)



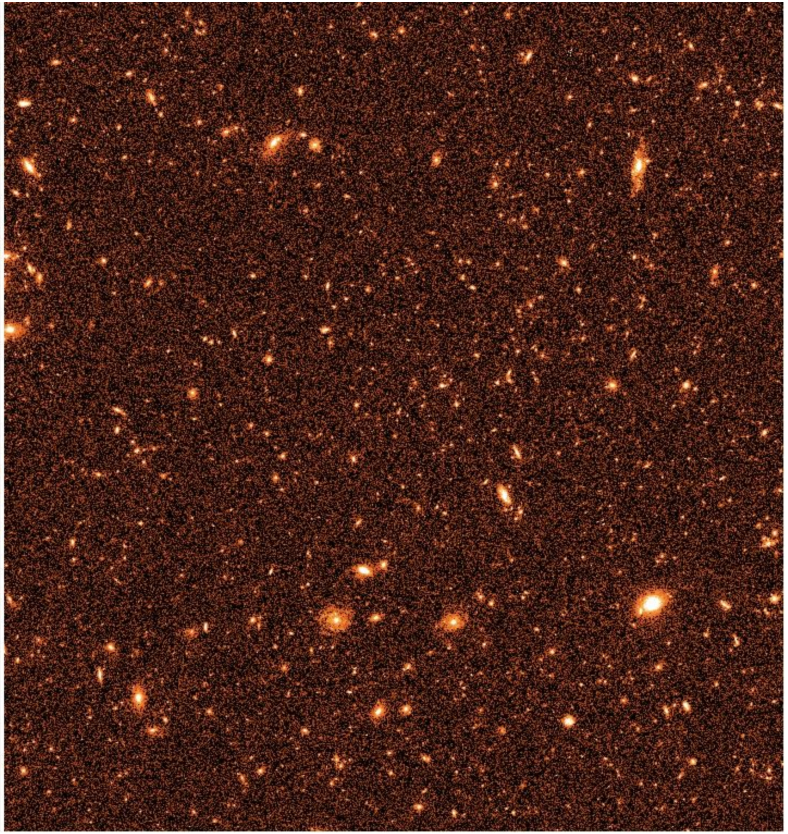
The CEERS survey layout in the EGS field. Split between June and December observations. The first four pointings were observed in June 2022.

JWST mock dataset

Input Universe f444w

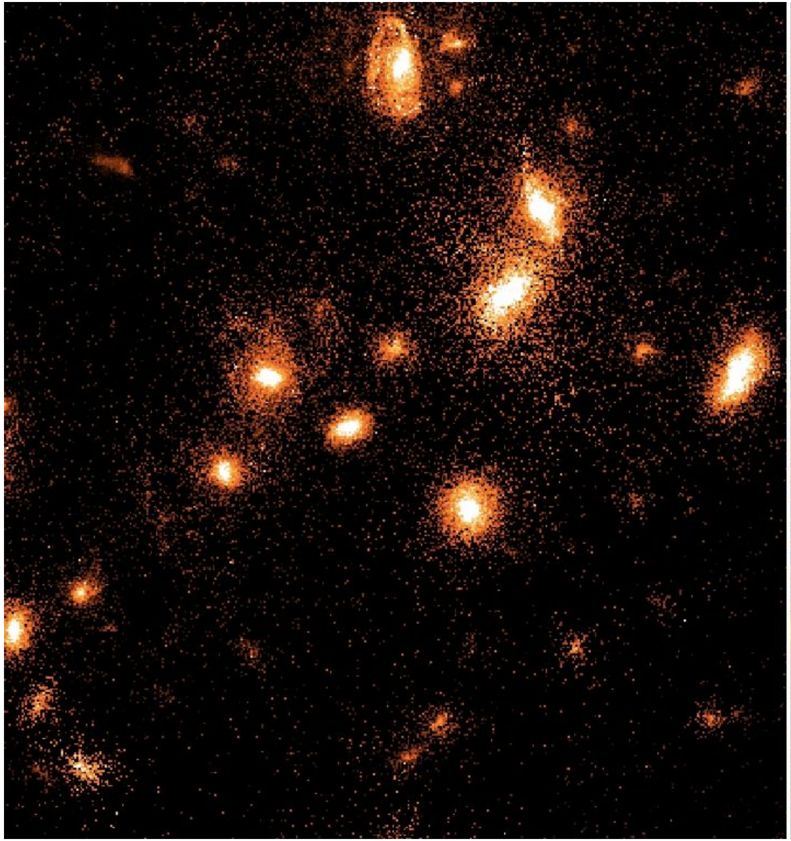


NIRCam f444w



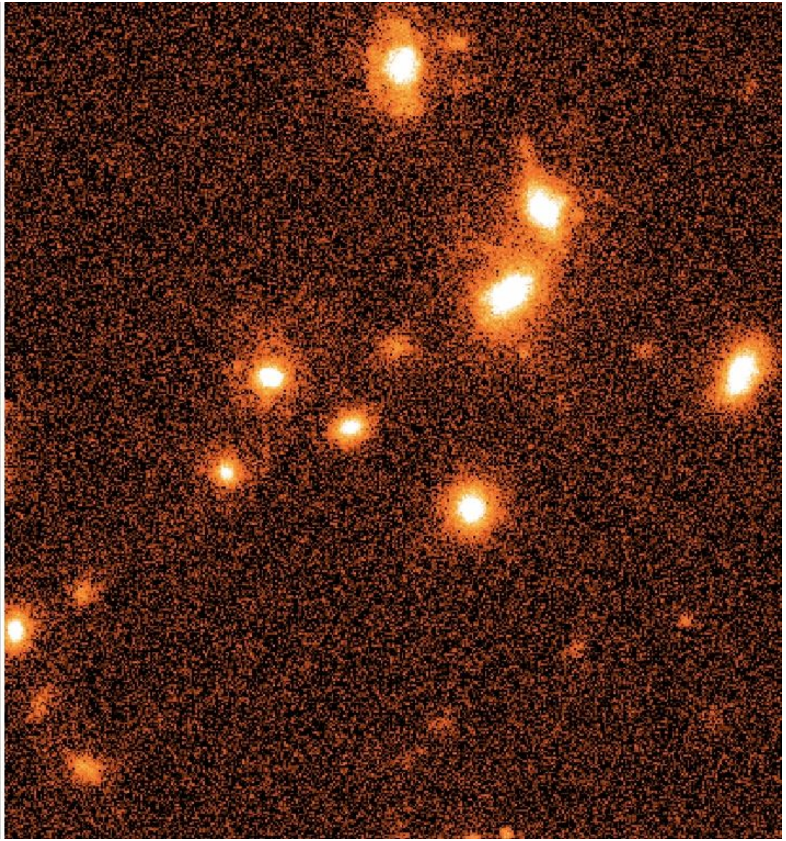
JWST mock dataset

Input Universe f444w



NIRCam f444w

30” x 30”



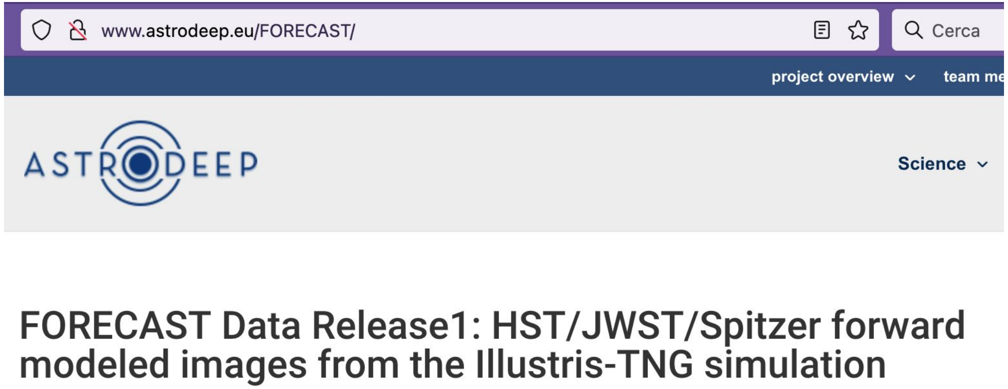
Summary

- FORECAST is a **flexible and adaptable** tool that forward models cosmological hydro-sims into synthetic astronomical images

- the tool extracts high-resolution light-cones from the output snapshots of any cosmological hydrodynamical simulation, derives observed fluxes (accounting for dust attenuation) and creates scientific images

- we release @ www.astrodeep.eu/FORECAST
  - **CANDELS GS** dataset in 13 filters
  - **JWST** dataset in 10 filters

forward-modeled with IllustrisTNG100



What's next?

- FORECAST code & datasets are available to the community on our website, www.astrodeep.eu/FORECAST
  - CANDELS GS and the JWST-like survey can be emulated with different depths using the 200 sq.arcmin cone
  - other datasets (different sims, FoV size, filters, depths) can be 'quickly' done (10 to 15 days of running)

- improving the forward-modeled physics
  - the treatment of dust: we need to account for dust emission
  - modeling AGN emission from the properties of BH particles

- emulating spectroscopic data, ..., test gravitational lensing, ..., suggestions?
- what happens if we forward model sims in different Cosmologies?

some references

Blaizot J., Wadadekar Y., Guiderdoni B., Colombi S. T., Bertin E., Bouchet F. R., Devriendt J. E. G., Hatton S., 2005, , 360, 159

Bravo M., Lagos C. d. P., Robotham A. S. G., Bellstedt S., Obreschkow D., 2020, , 497, 3026

Bruzual G., Charlot S., 2003, , 344, 1000

Henriques B. M. B., White S. D. M., Thomas P. A., Angulo R., Guo Q., Lemson G., Springel V., Overzier R., 2015, Monthly Notices of the Royal Astronomical Society, 451, 2663–2680

Jain B., Seljak U., White S., 2000, The Astrophysical Journal, 530, 547–577

Kitzbichler M. G., White S. D. M., 2007, , 376, 2

Laigle C., et al., 2019, Monthly Notices of the Royal Astronomical Society, 486, 5104–5123

Marinacci F., et al., 2018, Monthly Notices of the Royal Astronomical Society

Merlin E., et al., 2019, , 490, 3309

Merritt A., Pillepich A., van Dokkum P., Nelson D., Hernquist L., Marinacci F., Vogelsberger M., 2020, , 495, 4570

Merson A. I., et al., 2013, , 429, 556

Mobasher B., et al., 2015, , 808, 101

Nelson D., et al., 2018, The IllustrisTNG Simulations: Public Data Release (arXiv:1812.05609)

Pichon C., Thiébaut E., Prunet S., Benabed K., Colombi S., Sousbie T., Teyssier R., 2010, , 401, 705

Schaye J., et al., 2015, , 446, 521

Schreiber C., et al., 2018, , 618, A85

Snyder G. F., Lotz J. M., Rodriguez-Gomez V., Guimarães R. d. S., Torrey P., Hernquist L., 2017, , 468, 207

Springel V., 2010, Monthly Notices of the Royal Astronomical Society, 401, 791–851

Tessore N., Winther H. A., Metcalf R. B., Ferreira P. G., Giocoli C., 2015, , 2015, 036

Thomas D., Maraston C., Schawinski K., Sarzi M., Silk J., 2010, , 404, 1775

Weinberger R., et al., 2016, Monthly Notices of the Royal Astronomical Society, 465, 3291–3308

Backup Slides

FORECAST input requirements

STELLAR PARTICLES/CELLS (1e6-1e7 $M_\odot$)

- 3D coordinates (x,y,z)
- initial stellar mass
- stellar mass
- stellar metallicity
- age
- redshift

GAS CELLS/PARTICLES

- 3D coordinates (x,y,z)
- mass
- volume
- gas metallicity
- neutral hydrogen column density
- redshift

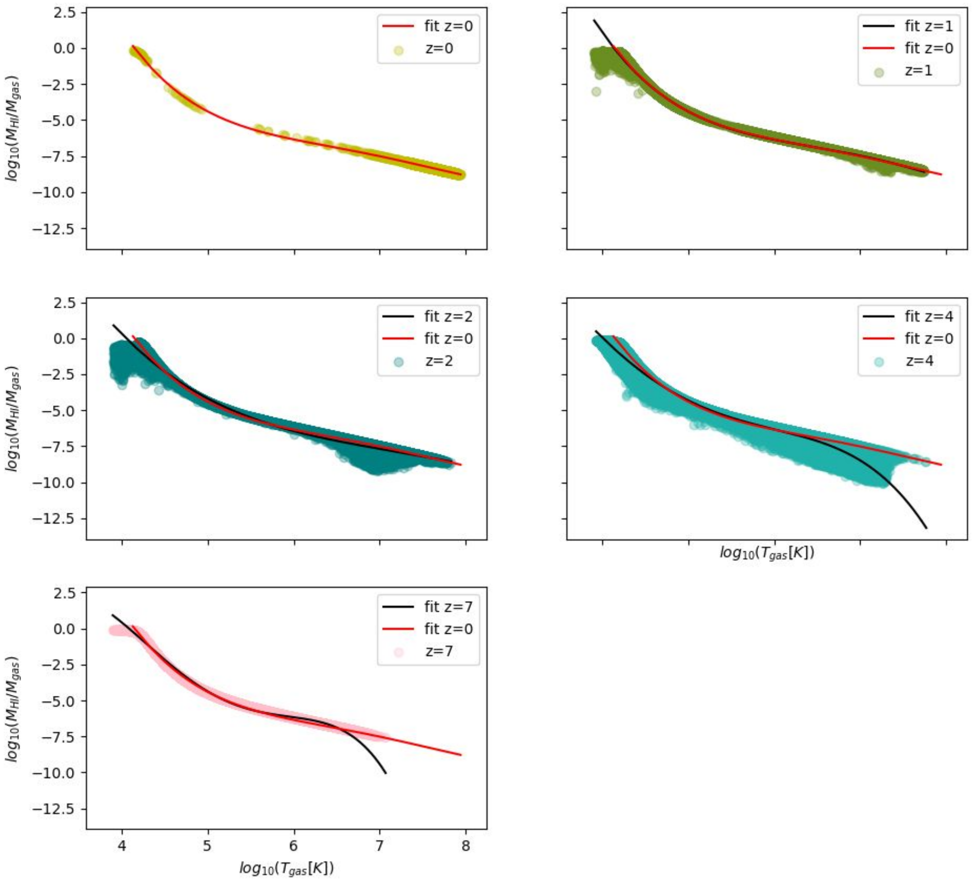
$$\tau_\lambda^a = \left(\frac{A_\lambda}{A_V}\right)_\odot (1+z)^{-0.5} \left(\frac{Z_g}{Z_\odot}\right)^\gamma \left(\frac{N_{HI}}{N_{HI,0}}\right)$$

dust: relation between gas and dust in IllustrisTNG

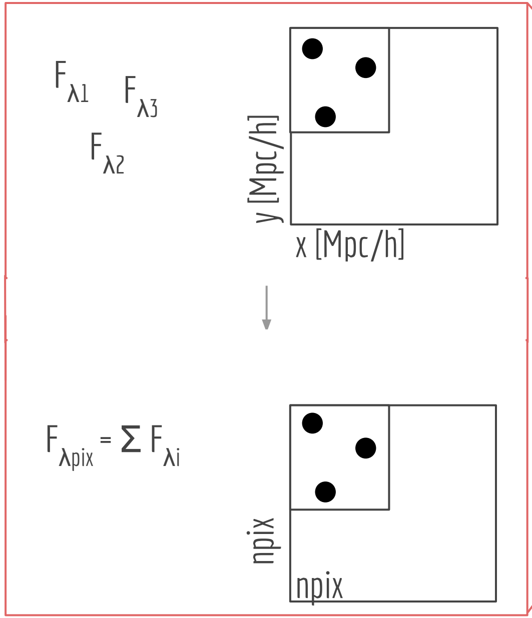
N_{HI} available only at few redshift ("big snapshots")

⟶ inferred from z=0 relation

MHI/Mg vs Tg



Mapping and “pixelating” fluxes



PIXELATING

- translating comoving projected positions in pixel coordinates
- summing the flux of each particle within a single pixel

NOISE AND PSF

Adding noise and PSF on top

Computational performances

TNG PROJECT DESCRIPTION PEOPLE + CONTACT RESULTS IMAGES + VIDEOS DATA ACCESS EXPLORE

Snapshot List:

TNG100-1 has [100] snapshots available, with 73.0 TB total data volume.

Over all snapshots, have **609,688,208** FoF groups, **469,094,112** Subfind groups, and **1,193,030,999,901** particles.

Snapshot [#]	Redshift	Age [Gyr]	Lookback [Gyr]	Download Snapshot	Download FoF & Subfind
0	20.05	0.179	13.624	[Snapshot] (494.1 GB)	[Groupcat] (12.5 MB)
1	14.99	0.271	13.532	[Snapshot] (494.1 GB)	[Groupcat] (78.4 MB)
2	11.98	0.370	13.433	[Snapshot] (1.7 TB)	[Groupcat] (456.5 MB)
3	10.98	0.418	13.385	[Snapshot] (1.7 TB)	[Groupcat] (788.5 MB)
4	10.00	0.475	13.328	[Snapshot] (1.7 TB)	[Groupcat] (1.2 GB)
5	9.39	0.517	13.286	[Snapshot] (493.3 GB)	[Groupcat] (1.6 GB)
6	9.00	0.547	13.256	[Snapshot] (1.7 TB)	[Groupcat] (1.9 GB)
7	8.45	0.596	13.207	[Snapshot] (492.6 GB)	[Groupcat] (2.4 GB)
8	8.01	0.640	13.163	[Snapshot] (1.7 TB)	[Groupcat] (2.8 GB)
9	7.60	0.687	13.116	[Snapshot] (491.5 GB)	[Groupcat] (3.2 GB)
10	7.24	0.732	13.071	[Snapshot] (490.9 GB)	[Groupcat] (3.5 GB)
11	7.01	0.764	13.039	[Snapshot] (1.7 TB)	[Groupcat] (3.8 GB)
12	6.49	0.844	12.959	[Snapshot] (489.4 GB)	[Groupcat] (4.3 GB)

https://www.tng-project.org/data/

disk storage, input ~ a lot! (e.g., 73 TB for TNG)

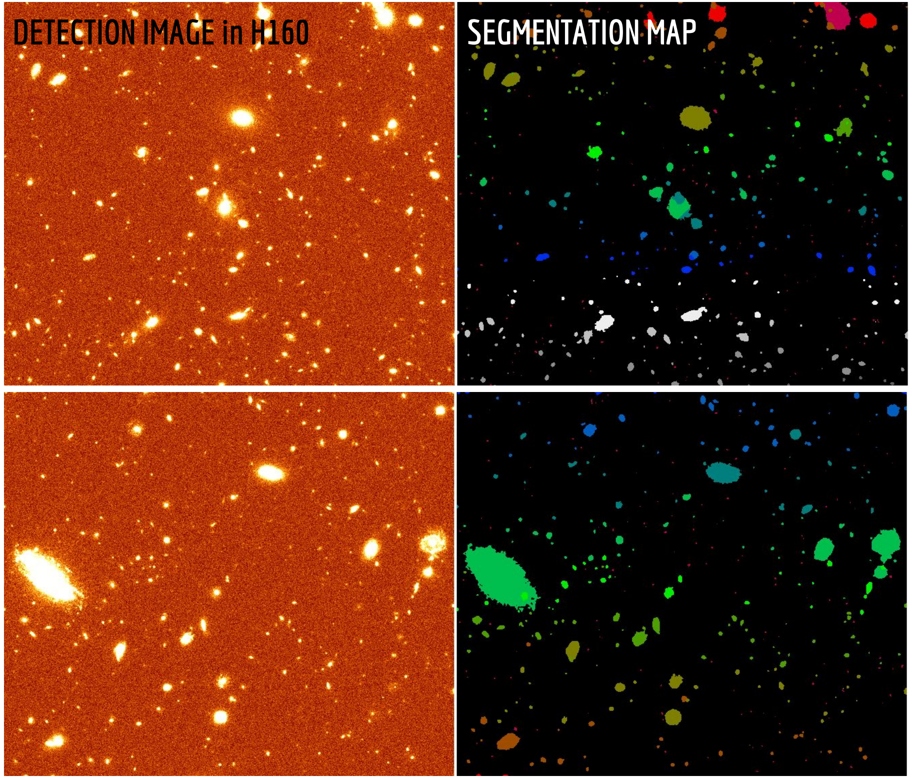
disk storage, output ~ 1.5 TB (images + catalogue)

memory ~ 40 to 180 GB to RAM, mostly due to convolution+integration of SEDs of billion of stellar particles + dust processing

running time: 10 - 20 days with min 48 to 240 cores available

the code is serial, but it can work on multiple snapshots simultaneously

Photometric analysis of the simulated images: detection



Examples of small areas of the detection image H160 and segmentation map

detection and deblending of sources with SExtractor (Bertin+96) software on the **H160** image

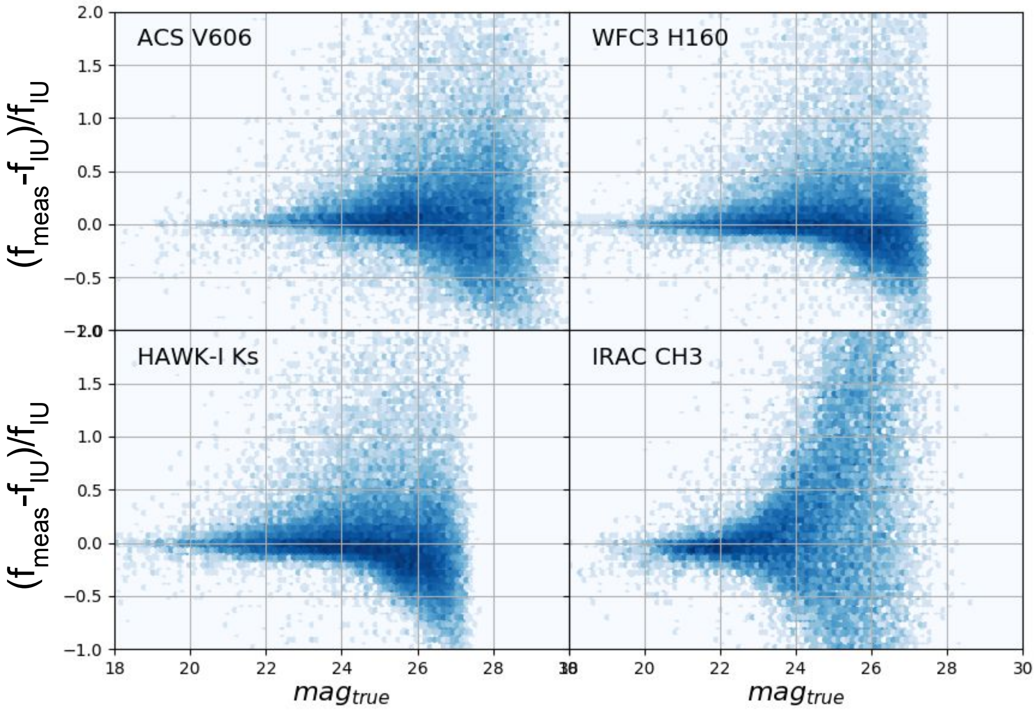
"Hot+Cold" detection procedure adopted by CANDELS team:

smaller/fainter + larger/brighter, removing double entries

catalogue + segmentation map with sources' IDs

Test1: photometry

df/f in 13 bands on sources matched with IU



FLUX MEASUREMENTS

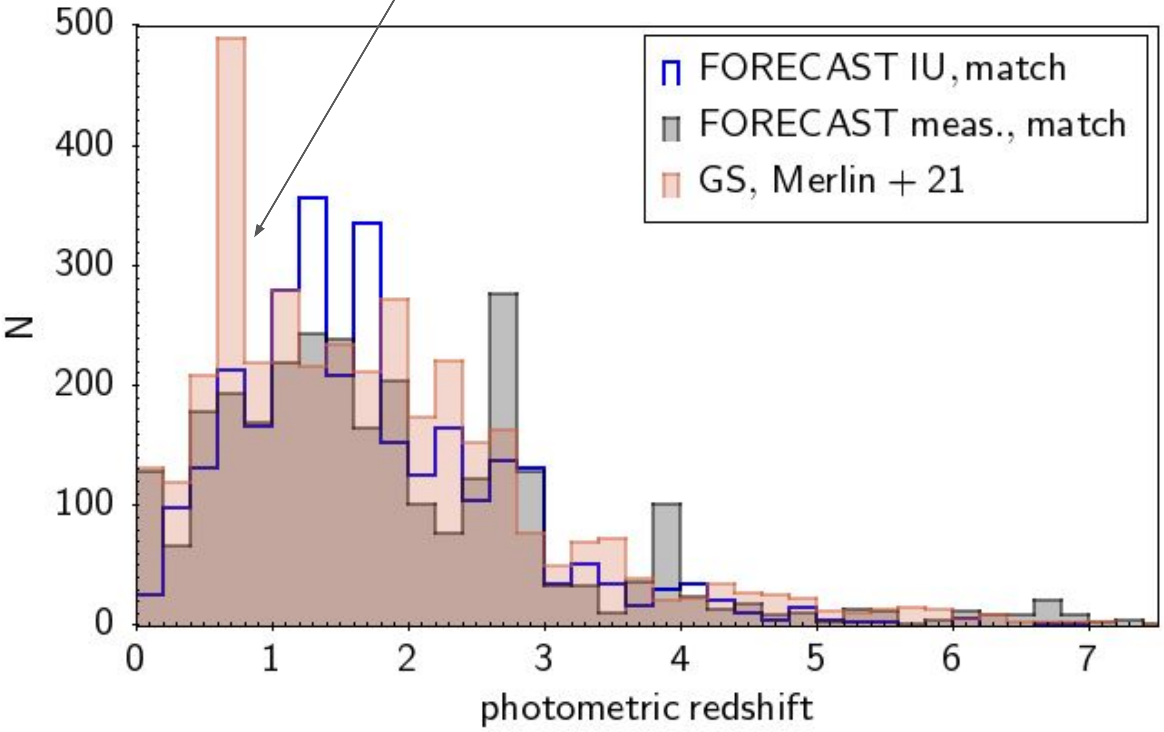
$$f_{band,tot} = f_{H,tot} \times (f_{band,segm}/f_{H,segm})$$

PSF-matched aperture photometry with A-PHOT (Merlin+19) on *HST* bands

template-fitting with T-PHOT (Merlin+15,16) on HAWK-I K_s and IRAC CH1,2,3,4 bands

Test2: photometric redshifts

local overdensities in GS (see Castellano+2007, Salimbeni+2009, Kang & Im 2009)

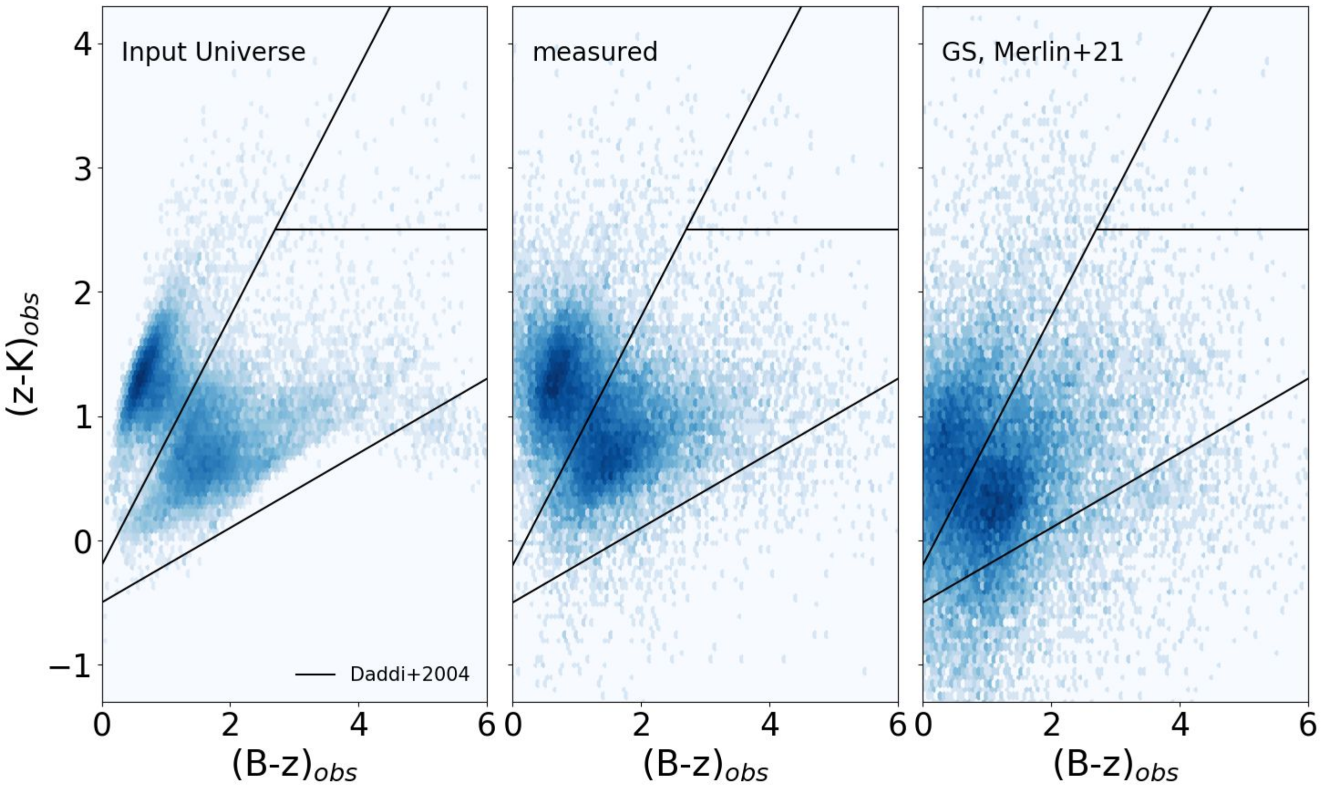


redshift (and physical properties) estimated with SED-fitting procedure (ZPHOT software, Fontana+00) :

- grid of parameters (ages, metallicities, τ-models for Star Formation History)
- Calzetti+00 dust extinction law
- Chabrier+03 IMF

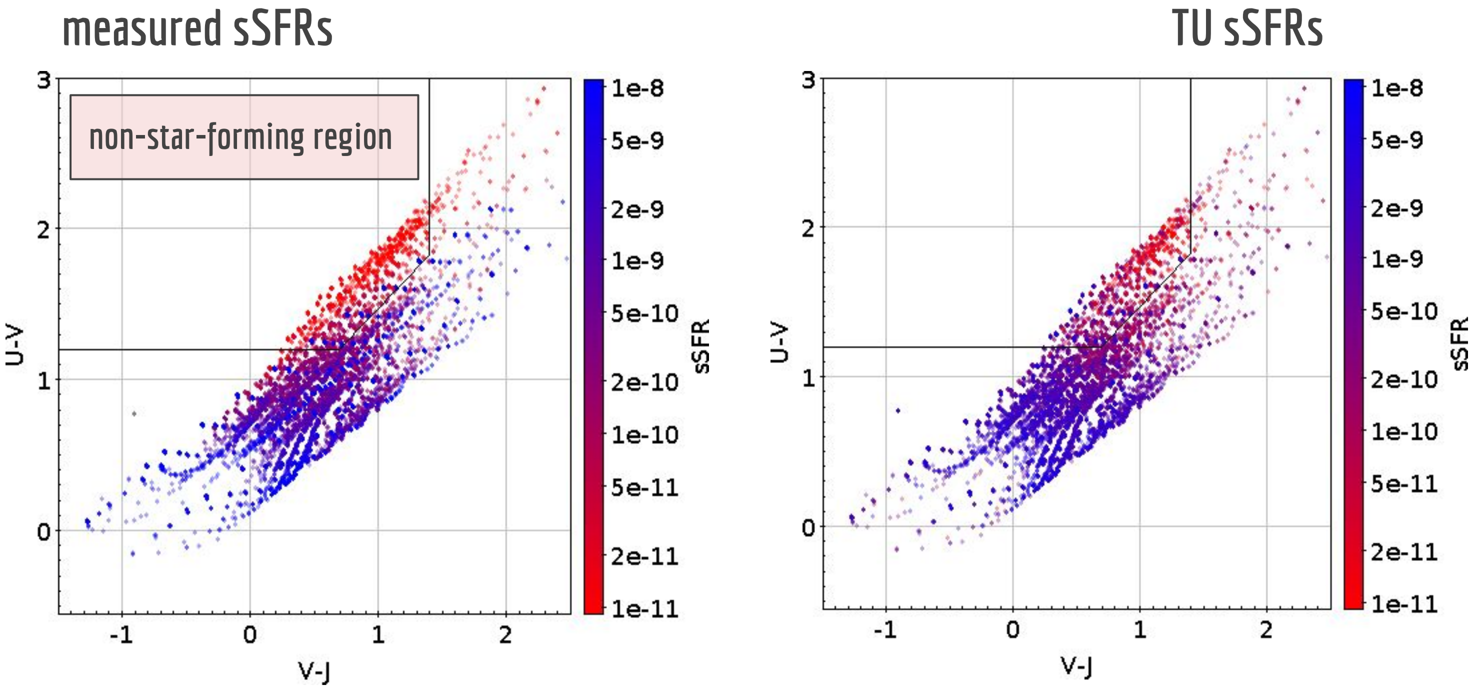
Test4: *BzK* color-color diagram

BzK, Daddi+04

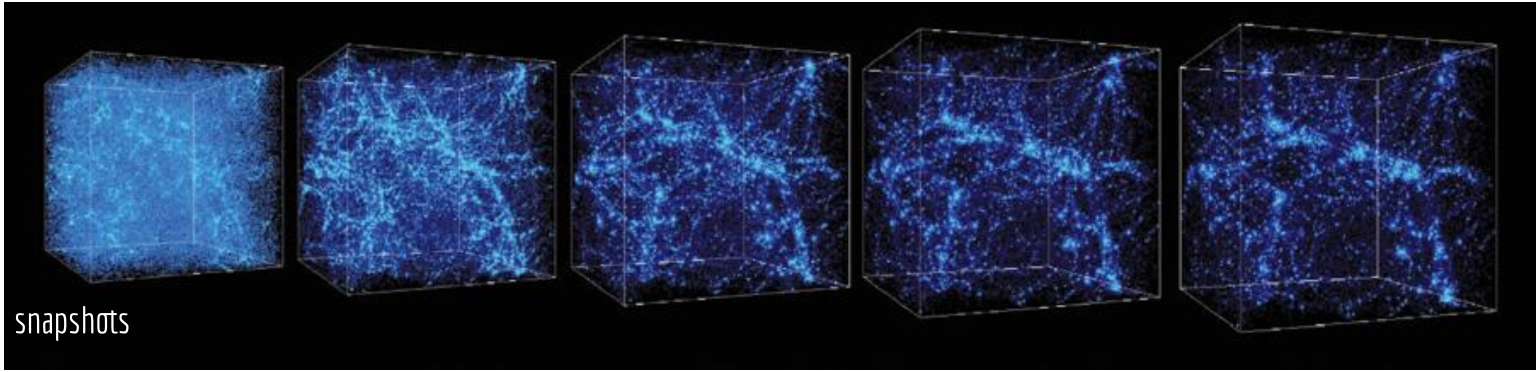


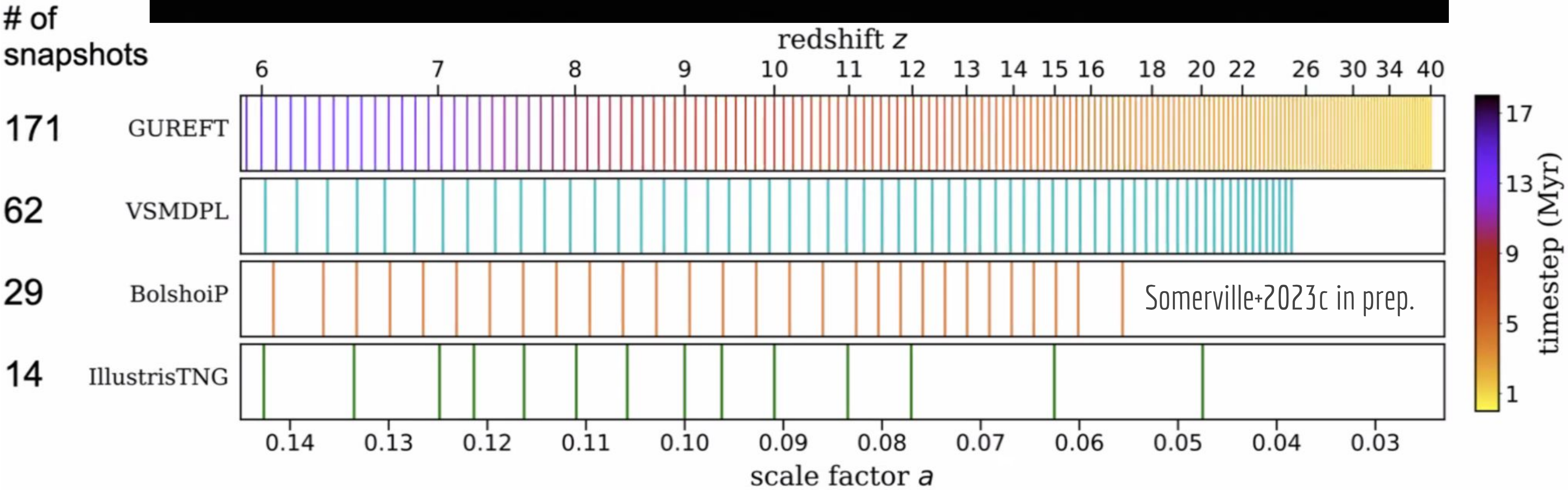
Test5: *UVJ* color-color diagram

detected sources with TU match:



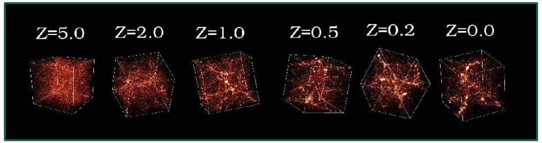
Cosmological simulations: redshift sampling





Limitations of the method

How does replication process affect our predictions?



We didn't check quantitatively, but we expect (see also Blaizot+2005):

- loss of clustering information due to random tiling of simulation box, affecting 2-point correlation functions

- finite volume effect: larger volume needed to describe density fluctuations on larger scales

- the average properties of the overall population conserved despite finite time-steps between simulation outputs (typically 100 Myr)