Introducing SkyPy

Will Hartley

SkyPy in a nutshell

"A package for modelling the Universe"

http://skypyproject.org/

- Open-source project, inspired by astropy.
- Intended to be community driven, by science needs
  - → open to contributions from the community
  - → open to new members (contact me if interested in joining)
- Light structure:
  - **Board:** Adam Amara (Portsmouth), Sarah Bridle (Manchester), Brian Nord (Fermilab)
  - **Co-ordinators** (current): Ian Harrison (Cardiff), Will Hartley (Geneva)
  - **Overall, around 20+ members**, most are inactive.
    - A few key contributors: Lucia Fonseca de la Bella, Richard Rollins, Philipp Sudek, Nicolas Tressore + others from before my time.
  - Contributor guidelines, code of conduct

SkyPy in a nutshell

Project aims, structure:

- Perform fast simulations of astrophysical sources – mostly at catalogue level.
  - → to produce mock surveys
  - → for simulation-based inference
- Validated, unit-tested code to perform common tasks in constructing mock surveys.
- Modular framework, to enable interoperability with other software.
- Driven by yaml config files and a simple pipeline set-up to minimise learning curve.

Current status:

- Basic galaxies module is fairly complete, and under-going tuning, validation (this talk). Based on existing methods (e.g. Herbel et al. 2017, Tortorelli et al. 2021).
- Multiwavelength, time-domain, survey realism to come.

Simulation – analytic / catalogue level

Can mean many things….

Toy models

Idealised

Realistic

Ideal / complete

Simulation – analytic / catalogue level

Toy models

Idealised

Realistic

Ideal / complete

For demonstrating principles, prototyping etc.

→ Highly contrived population / SEDs etc.

Simulation – analytic / catalogue level

Toy models

Idealised

Realistic

Ideal / complete

For software, pipeline development, testing; simple mock surveys, where detail is unimportant.

→ simple well-understood and controlled truth values; limited sampling of param space.

Simulation – analytic / catalogue level

Toy models

Idealised

Realistic

Ideal / complete

For rigorous tests of completed pipelines; mock surveys with full diversity, but simplifications where aspects don’t matter.

→ Attempt to achieve similar complexity to real data via prescriptions, avoiding the cost and messiness of full instrument simulation.

Simulation – analytic / catalogue level

Toy models

Idealised

Realistic

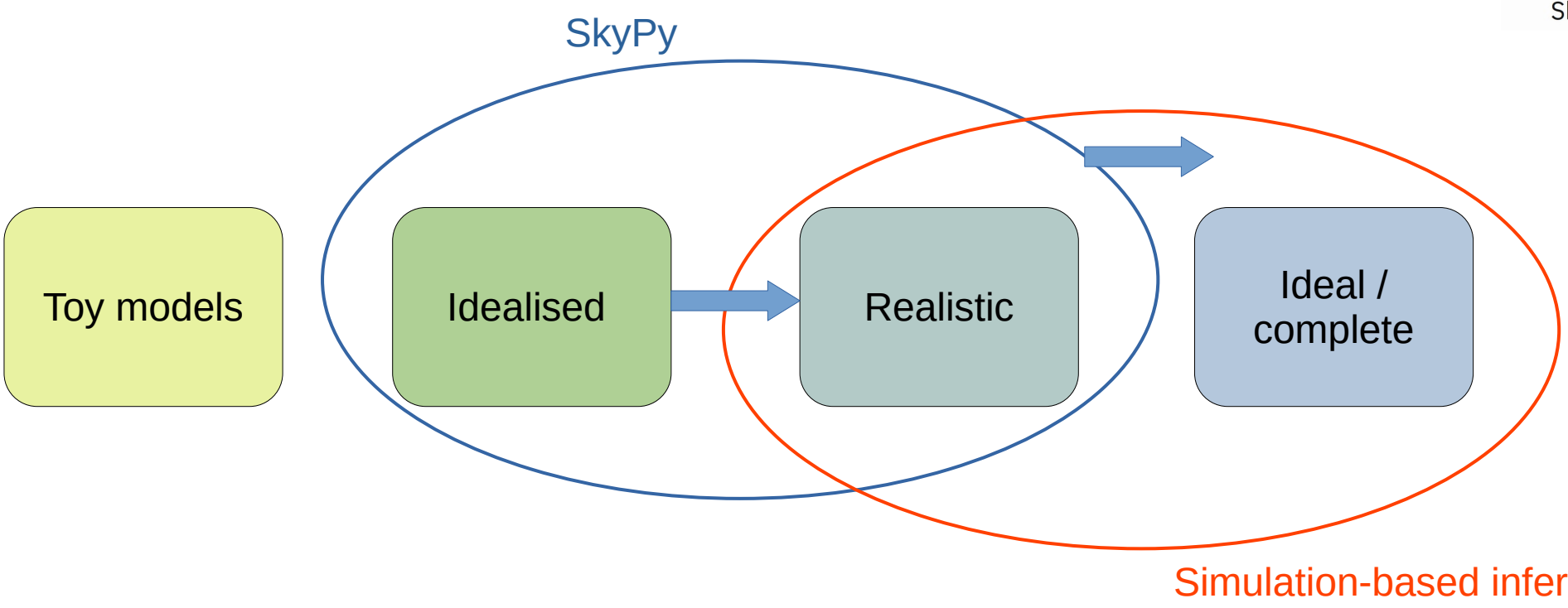
Ideal / complete

For end-to-end analysis of experiments, or cases where measurement biases are non-trivial and important.

→ Image simulations + re-extraction, with well-tuned population parametrisation.

Simulation – analytic / catalogue level

SkyPy



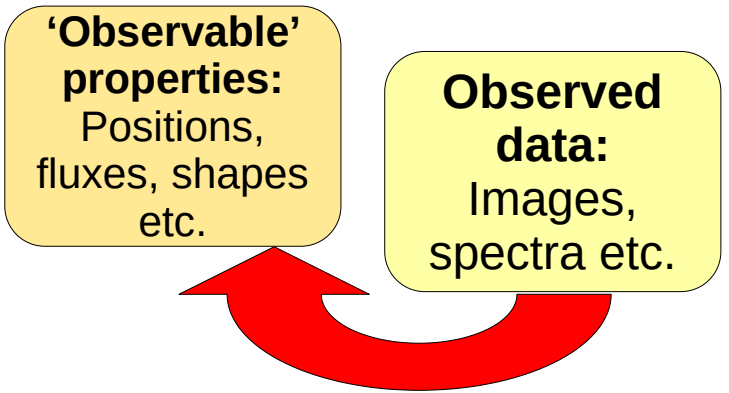
Simulation-Based Inference (SBI)

A.K.A. Forward modelling / likelihood-free inference

A traditional analysis:

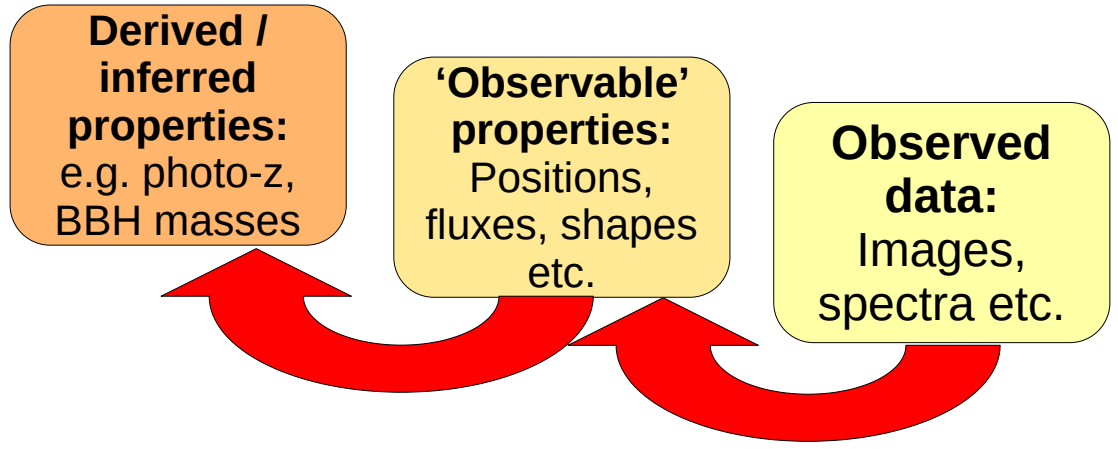
Observed data: Images, spectra etc.

Simulation-Based Inference (SBI)



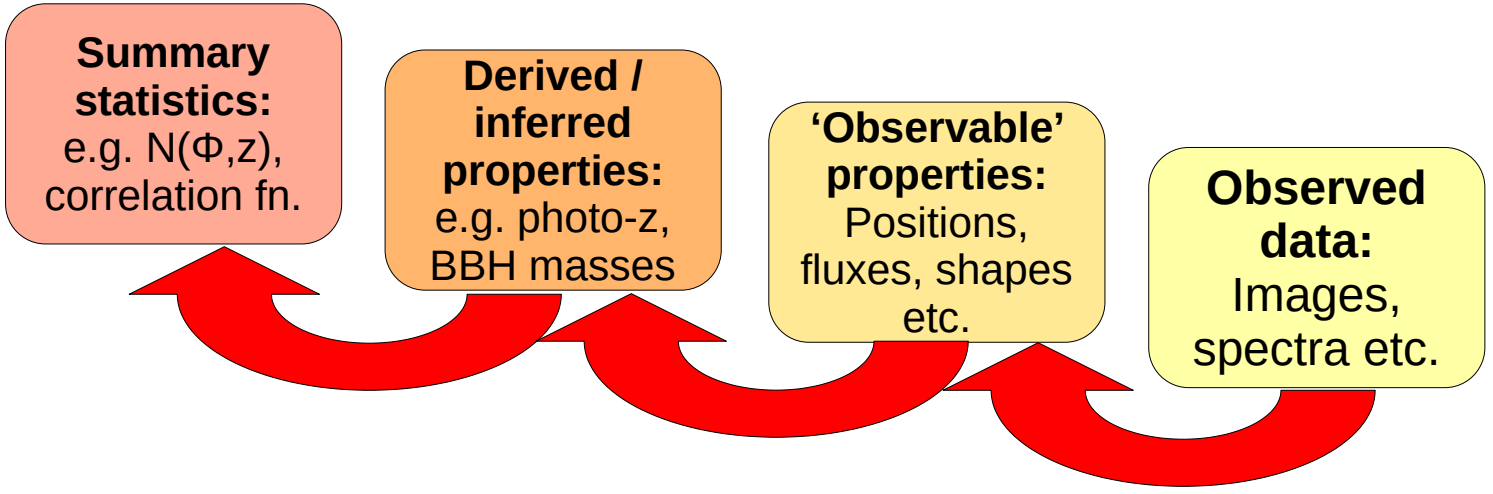
Simulation-Based Inference (SBI)

SkyPy



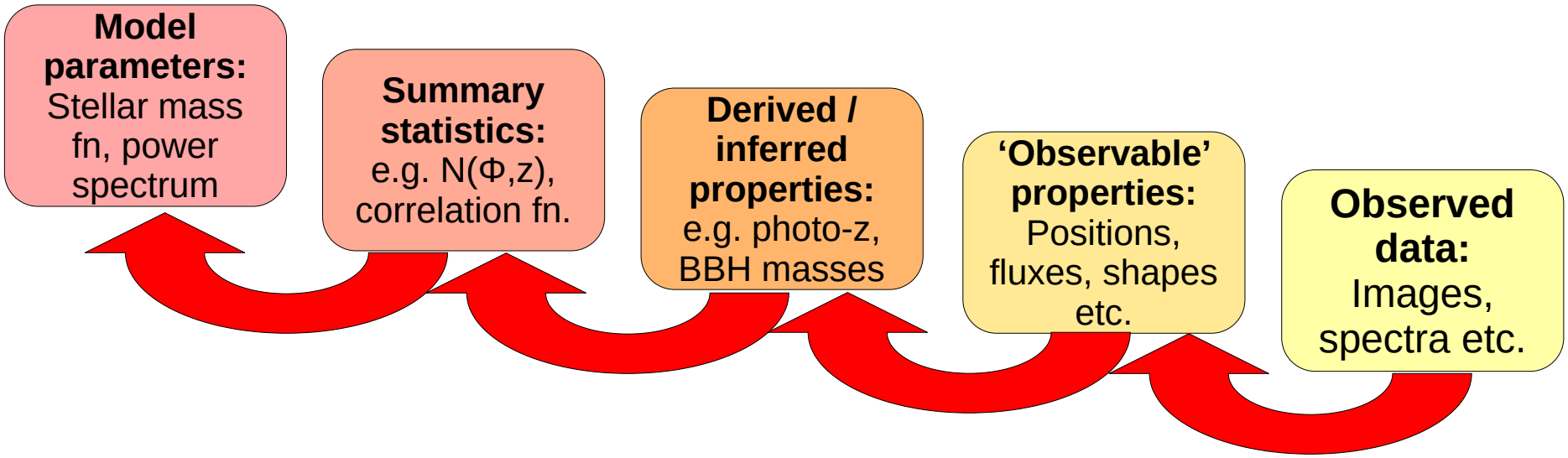
Simulation-Based Inference (SBI)

SkyPy



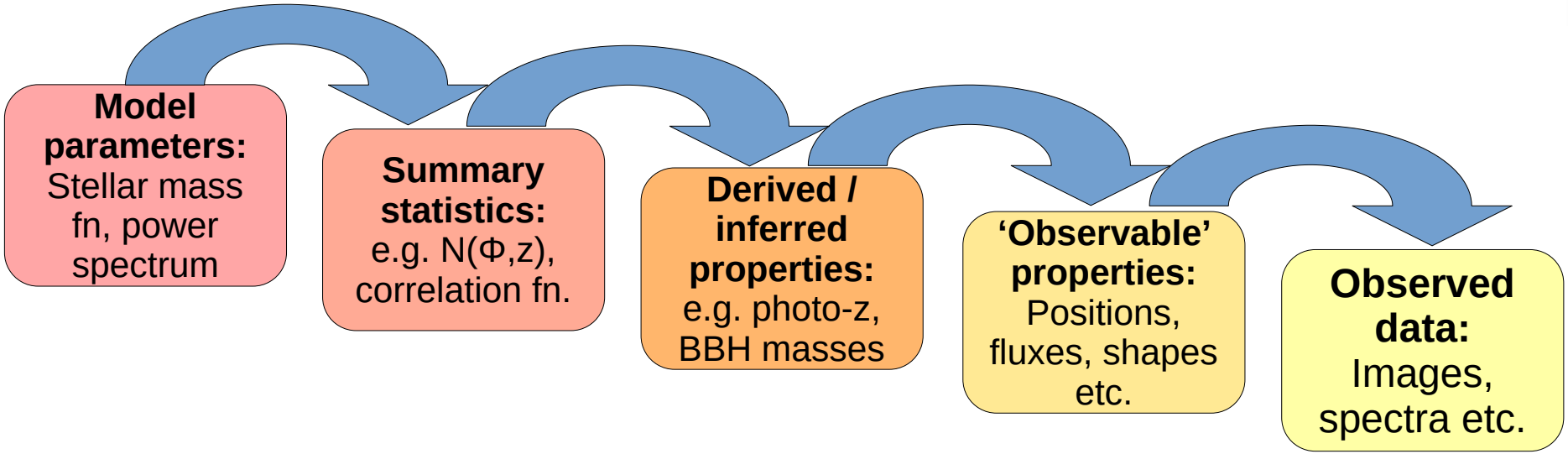
Simulation-Based Inference (SBI)

SkyPy



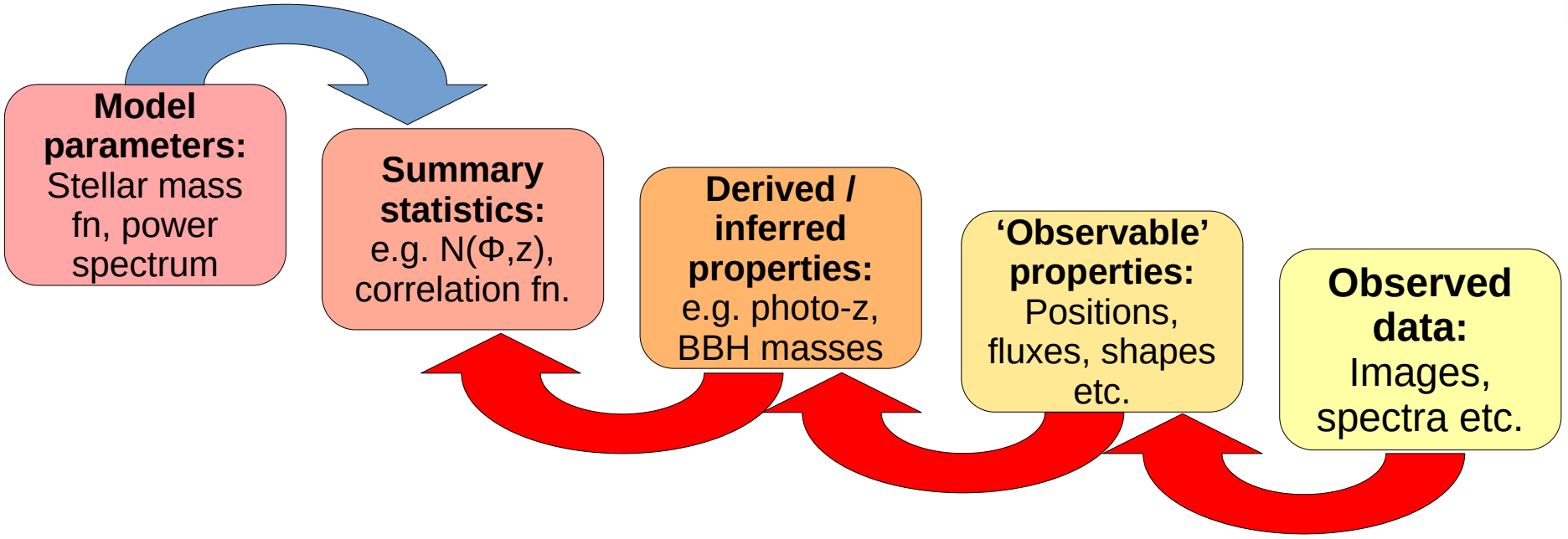
Simulation-Based Inference (SBI)

SkyPy



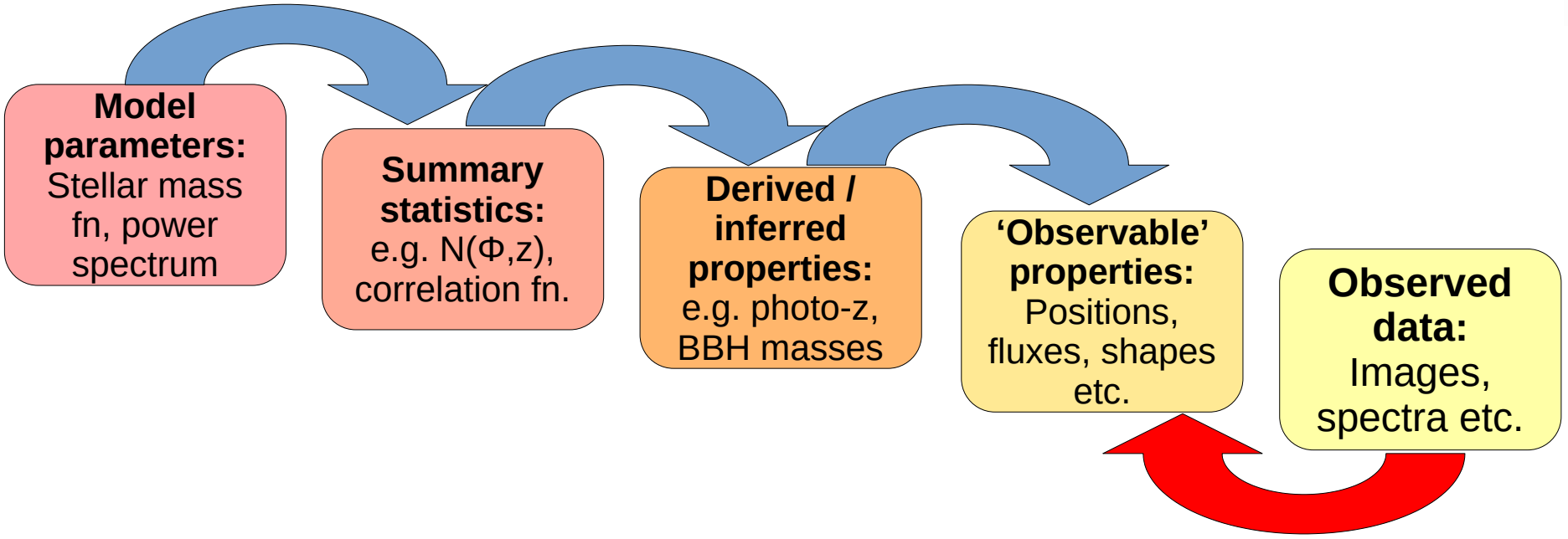
Simulation-Based Inference (SBI)

SkyPy



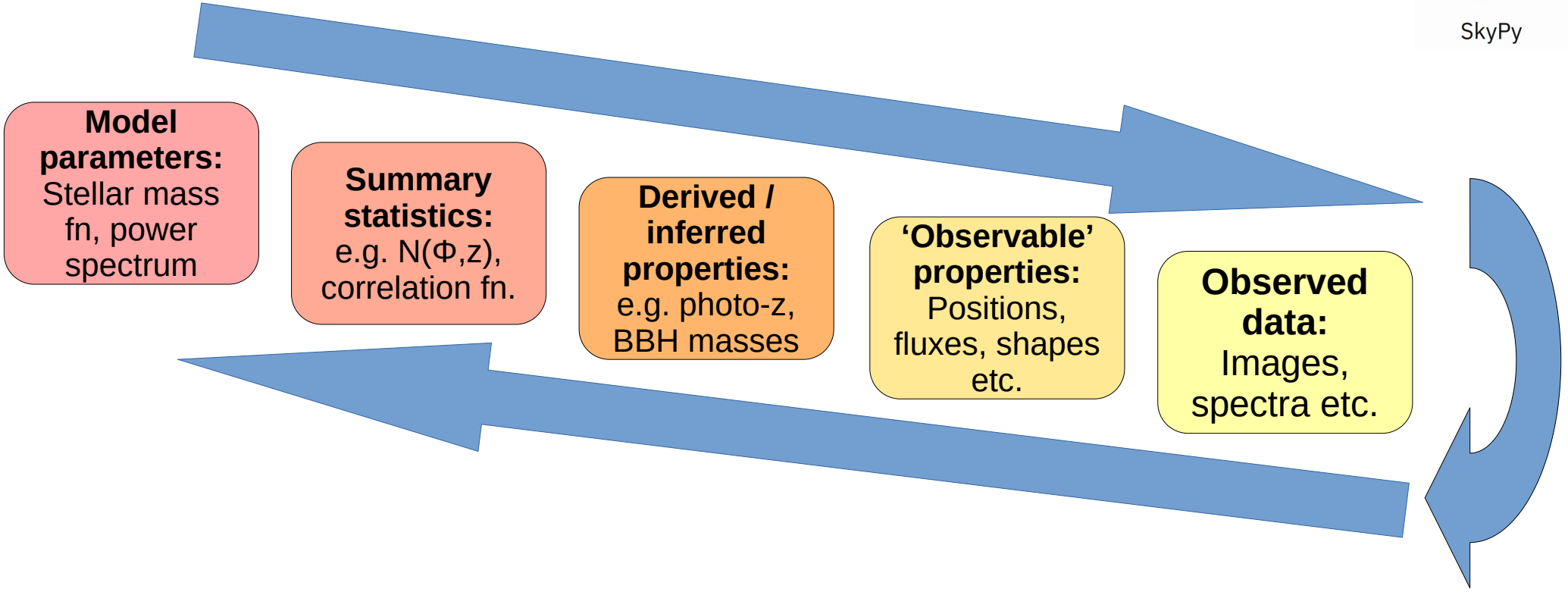
Simulation-Based Inference (SBI)

SkyPy



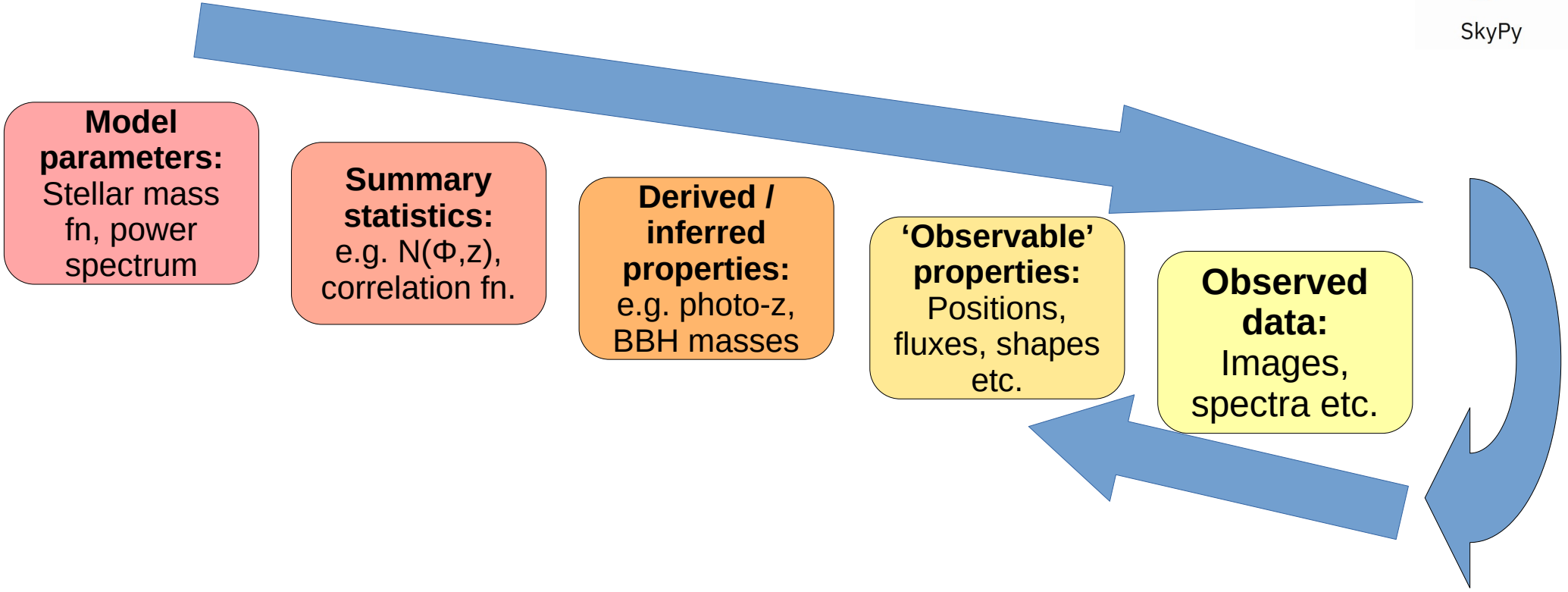
Simulation-Based Inference (SBI)

SkyPy



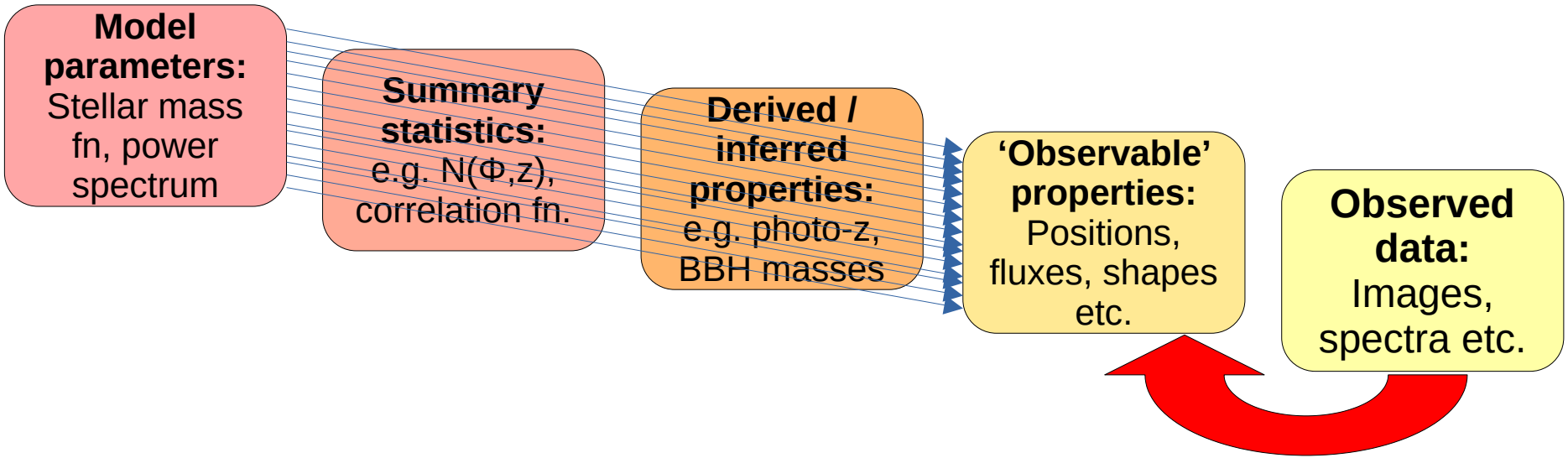
Simulation-Based Inference (SBI)

SkyPy

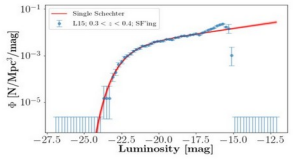


Simulation-Based Inference (SBI)

SkyPy



SkyPy Galaxies module



Luminosity function parameters
Φ*, M*, α

Samples drawn from LFn:
z, L

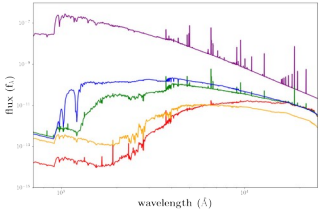
with survey area

Cosmology

Sampled SEDs (R-F):
z, L, coeffs

Observable properties:
fluxes

Catalogue products:
Noisy fluxes



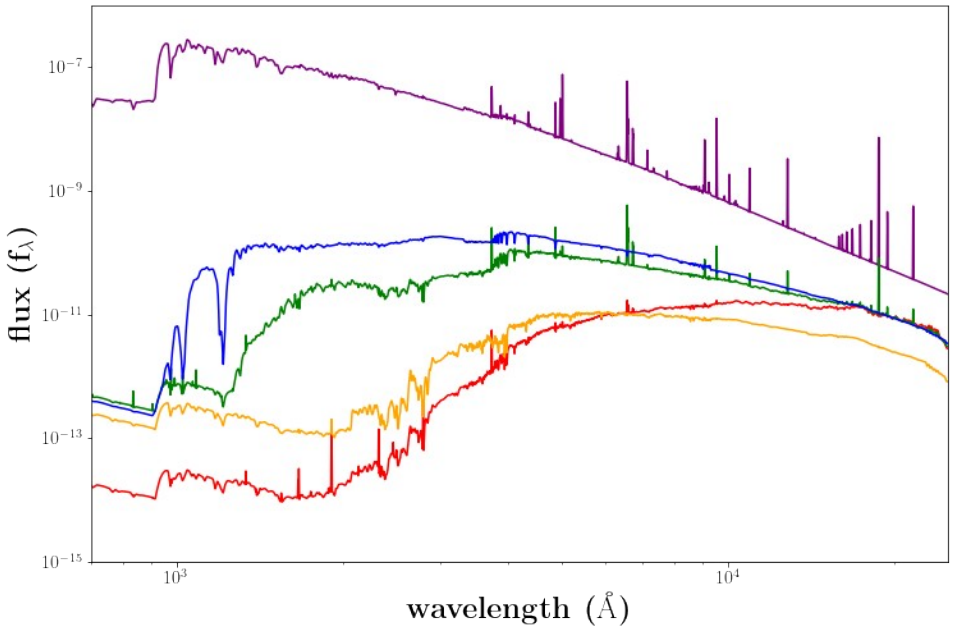
Rest-frame SEDs + coefficients

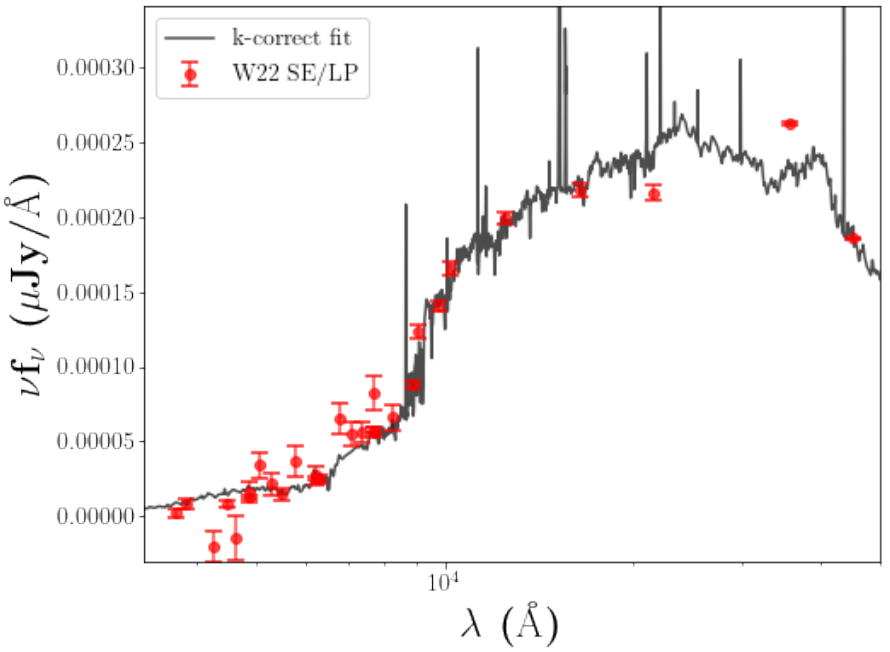
Survey error model

Calibration + Testing data

SkyPy

- Based on the COSMOS2020 catalogue (Weaver et al. 2022).
- Photo-z from 25 photometric bands → kept fixed.
- Re-fit SEDs, stellar masses using EAzY with k-correct templates.





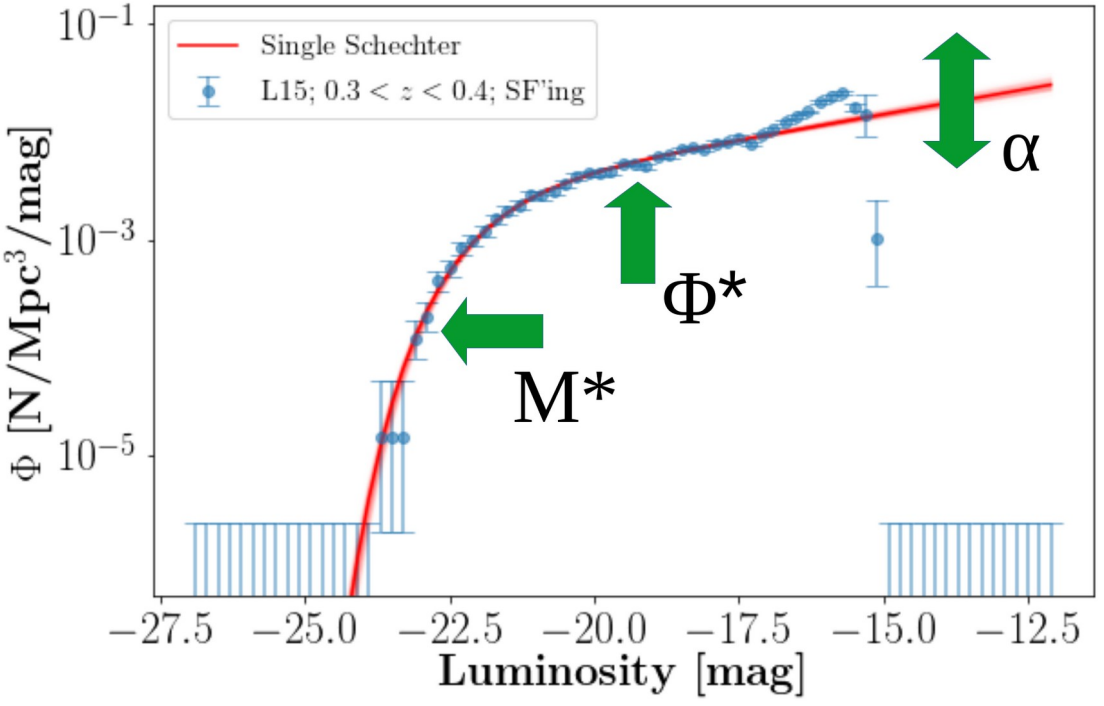
Component I: Galaxy luminosity function

SkyPy

$$n(M)\,\mathrm{d}M = 0.4\,\ln 10\,\phi^*\,[10^{0.4(M^*-M)}]^{\alpha+1}\,\exp[-10^{0.4(M^*-M)}]\,\mathrm{d}M.$$

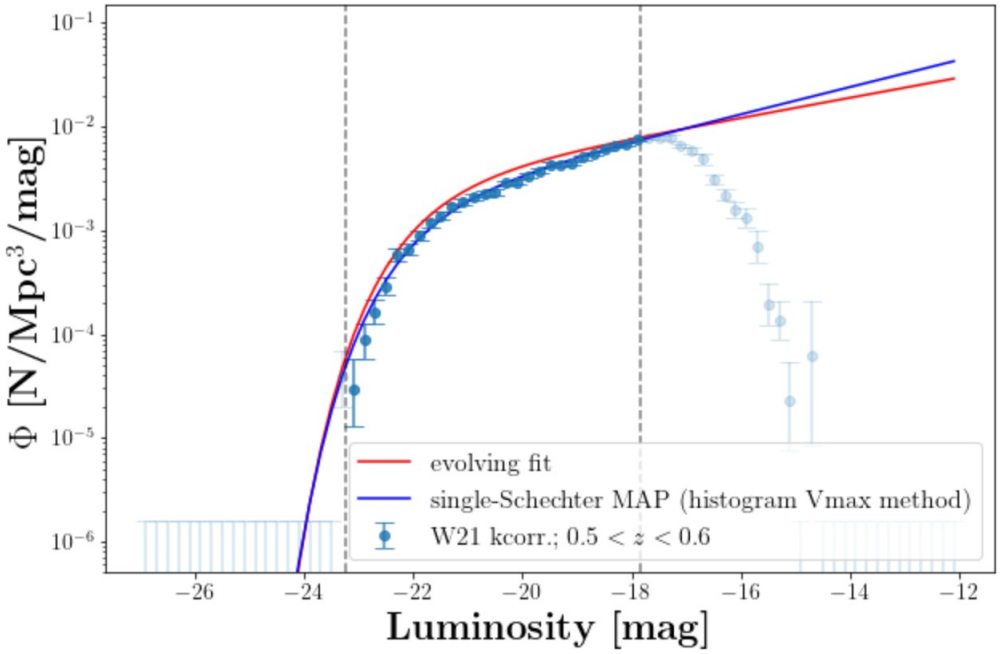
Fitting methods:

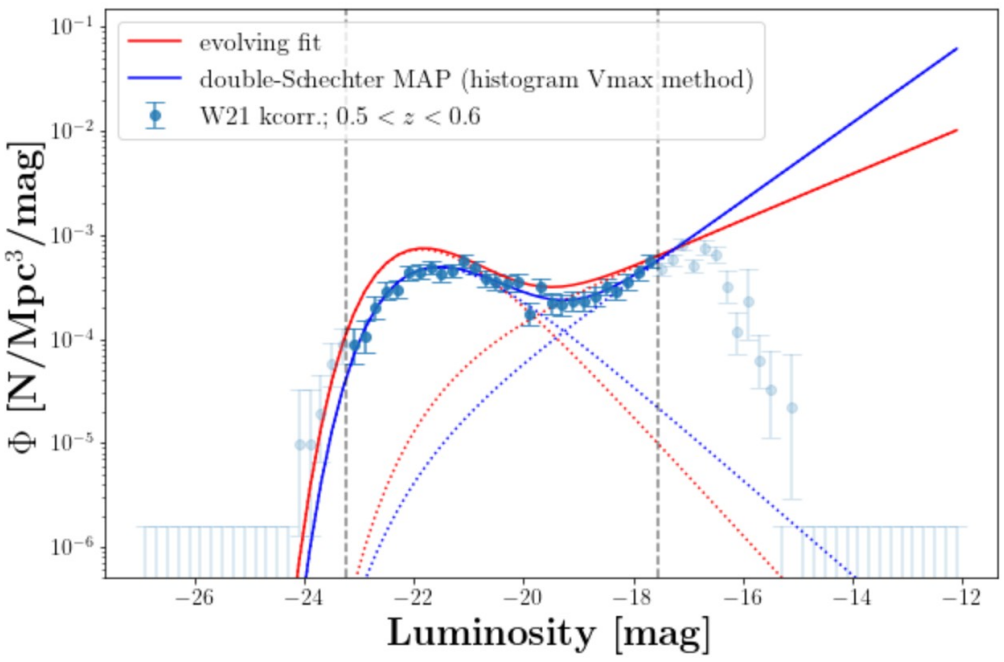
- V / V_{max}
  - Binned quantities → loss of information.
- "STY" (object-by-object likelihood)
  - Amplitude constrained separately.
  - Slow.
- Aird et al. (X-ray LF)
  - Includes amplitude.
  - Straight-forward to include redshift evolution.



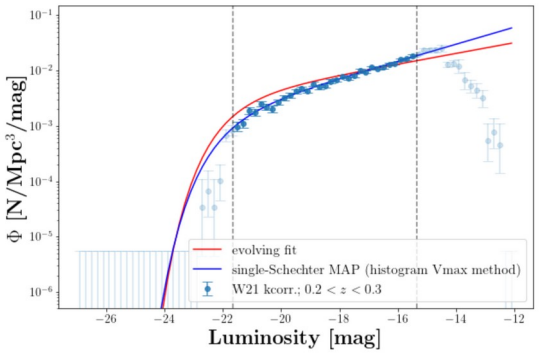
Component I: Galaxy luminosity function

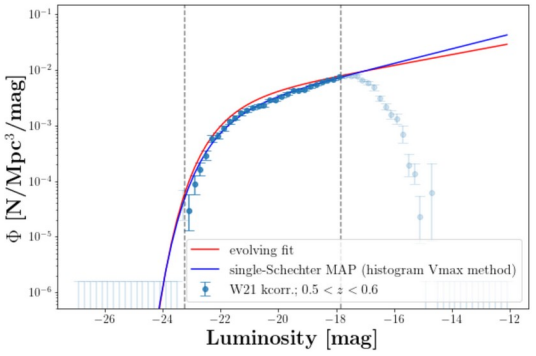
SkyPy

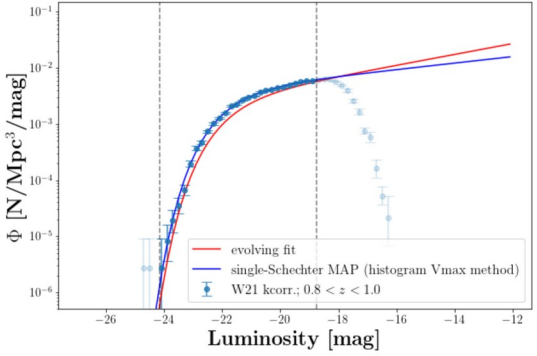


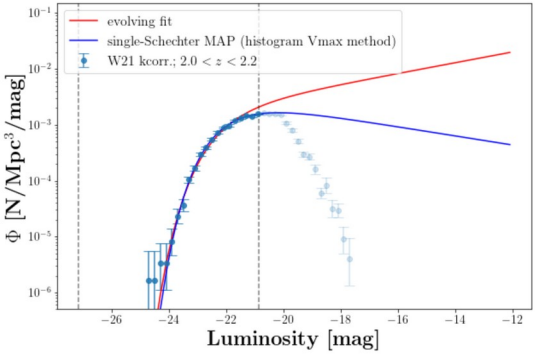


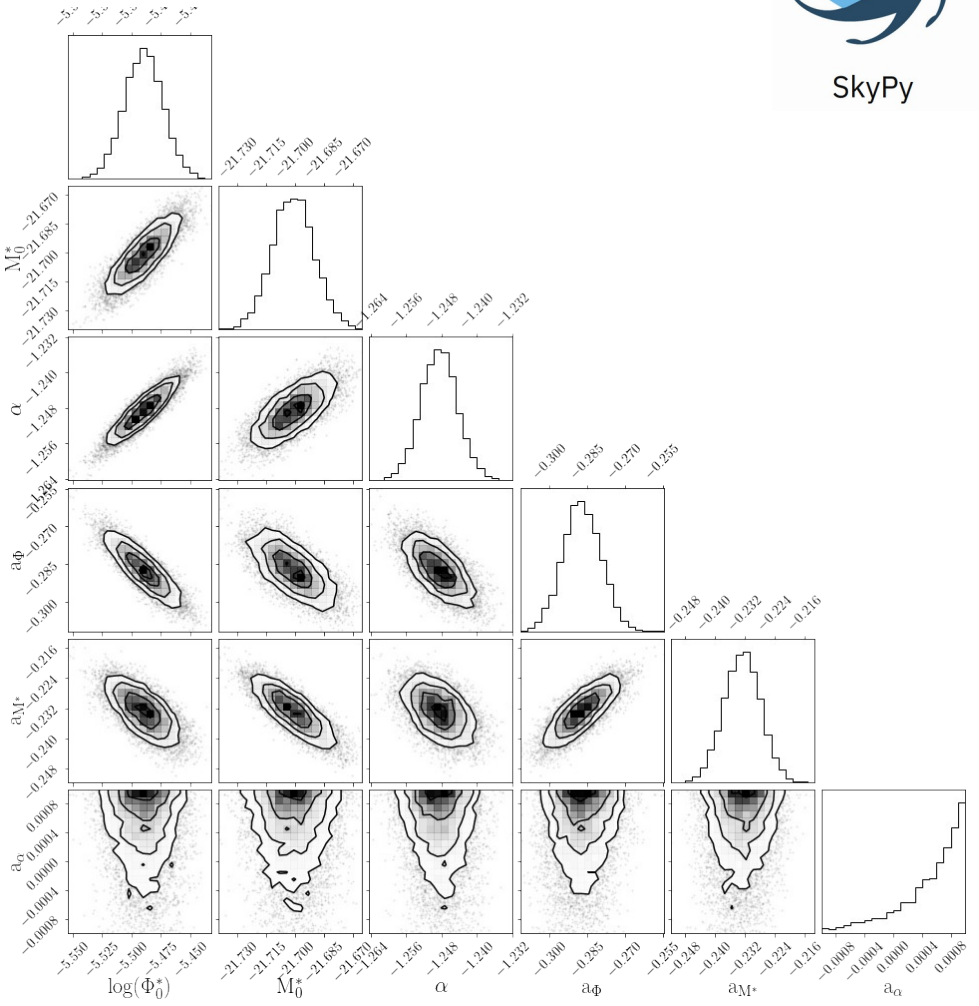
Component I: Galaxy luminosity function







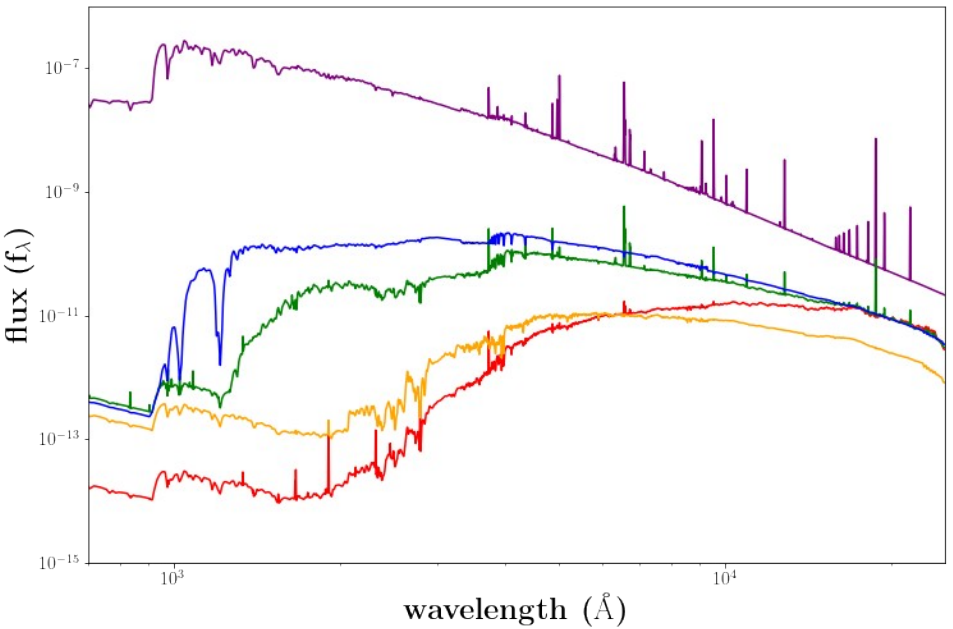


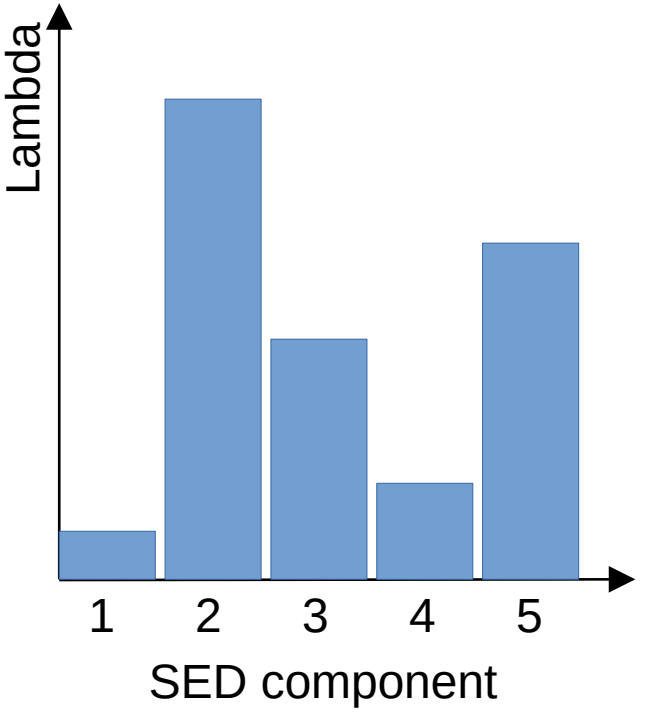


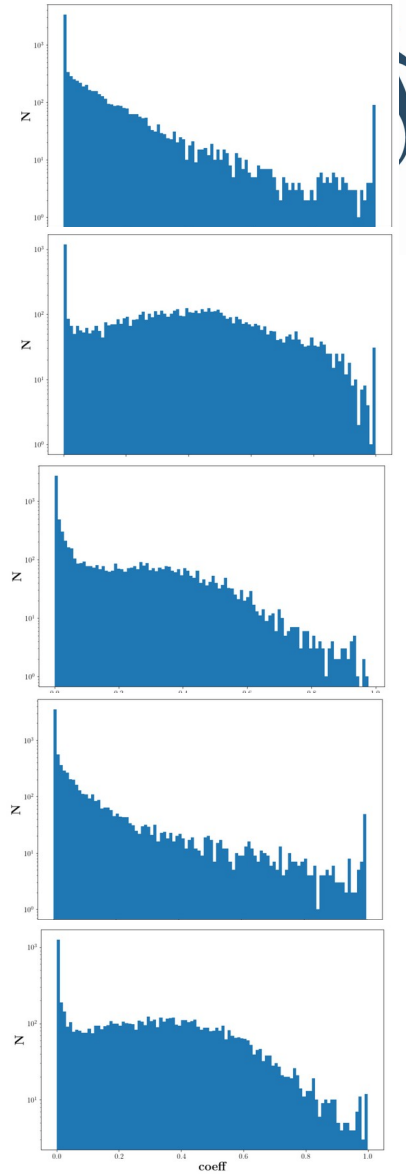
SkyPy

Component II: Rest-frame SED model

- Coefficients for linearly-combined SED compenents are sampled via a Dirichlet process.





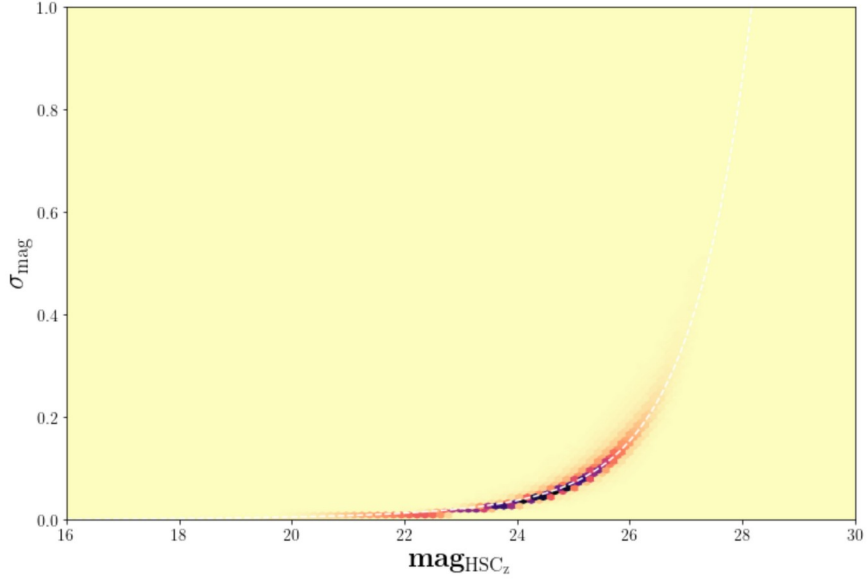


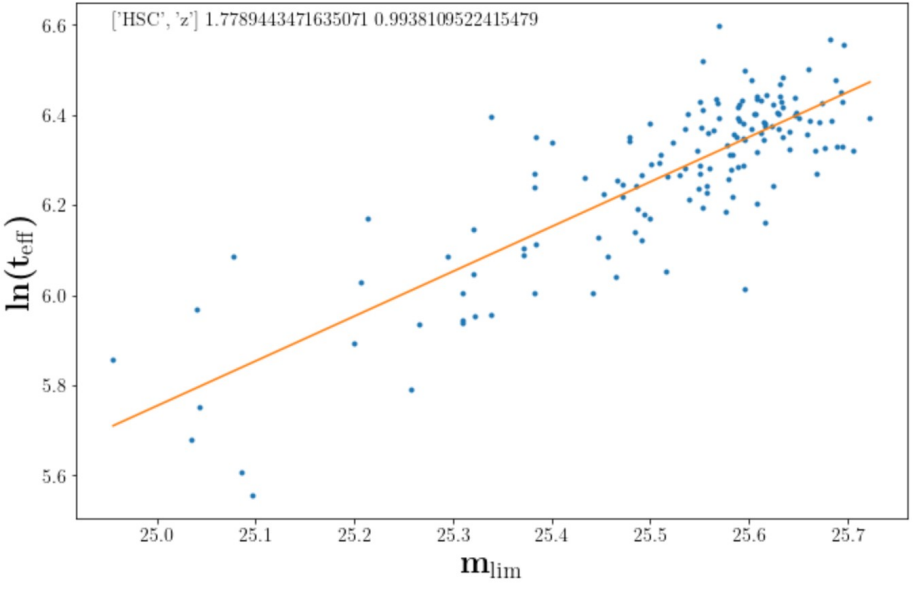
Component III: Survey error model

- Photometric errors based on Rykoff model (newly implemented, Philipp Sudek).

SkyPy

$$\sigma_m(F|F_{\rm noise}, t_{\rm eff}) = \frac{2.5}{ln\ 10}\left[\frac{1}{Fkt_{\rm eff}}\left(1+\frac{F_{\rm noise}}{F}\right)\right]^{1/2} \qquad F_{\rm noise} = \frac{F_{\rm lim}^2 kt_{\rm eff}}{10^2} - F_{\rm lim} \qquad ln\ t_{\rm eff} = a + b(m_{\rm lim} - 21)$$



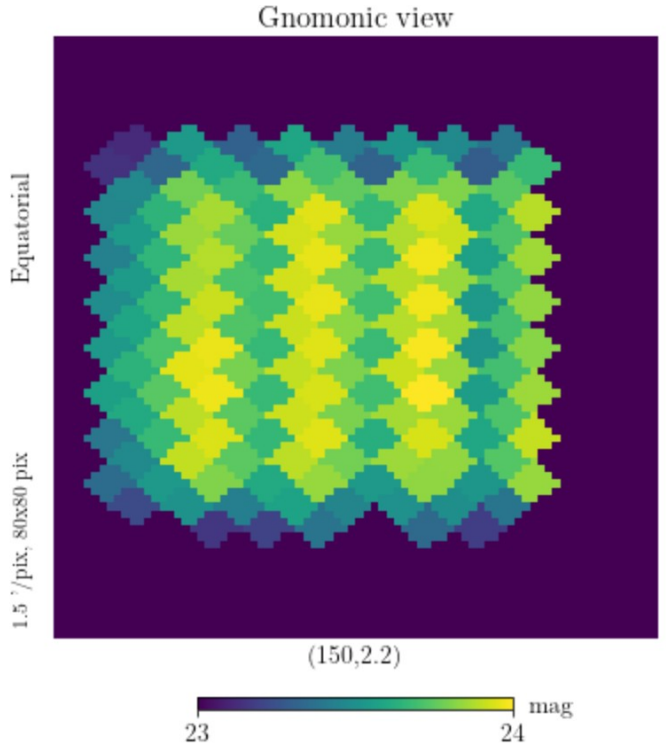


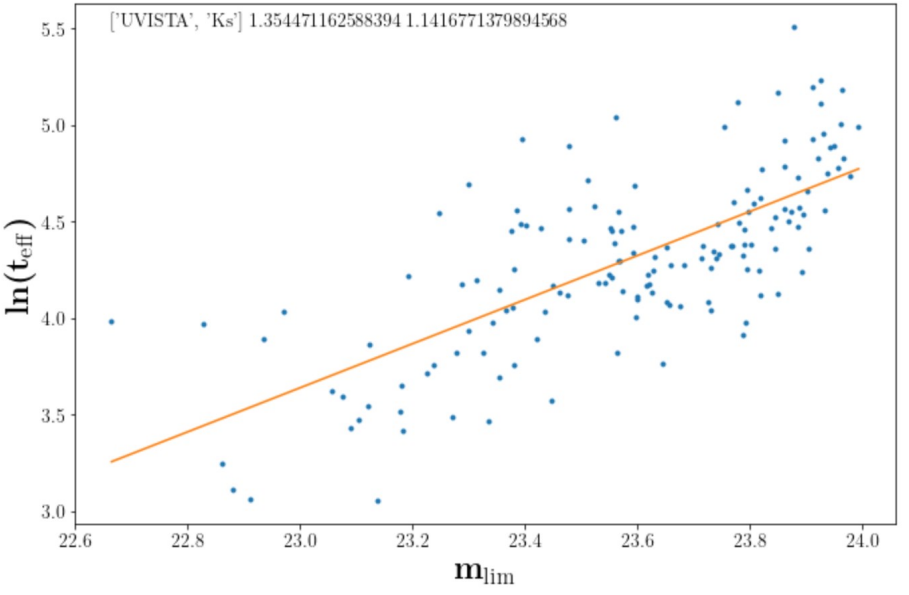
Component III: Survey error model

- Photometric errors based on Rykoff model (newly implemented, Philipp Sudek).

SkyPy

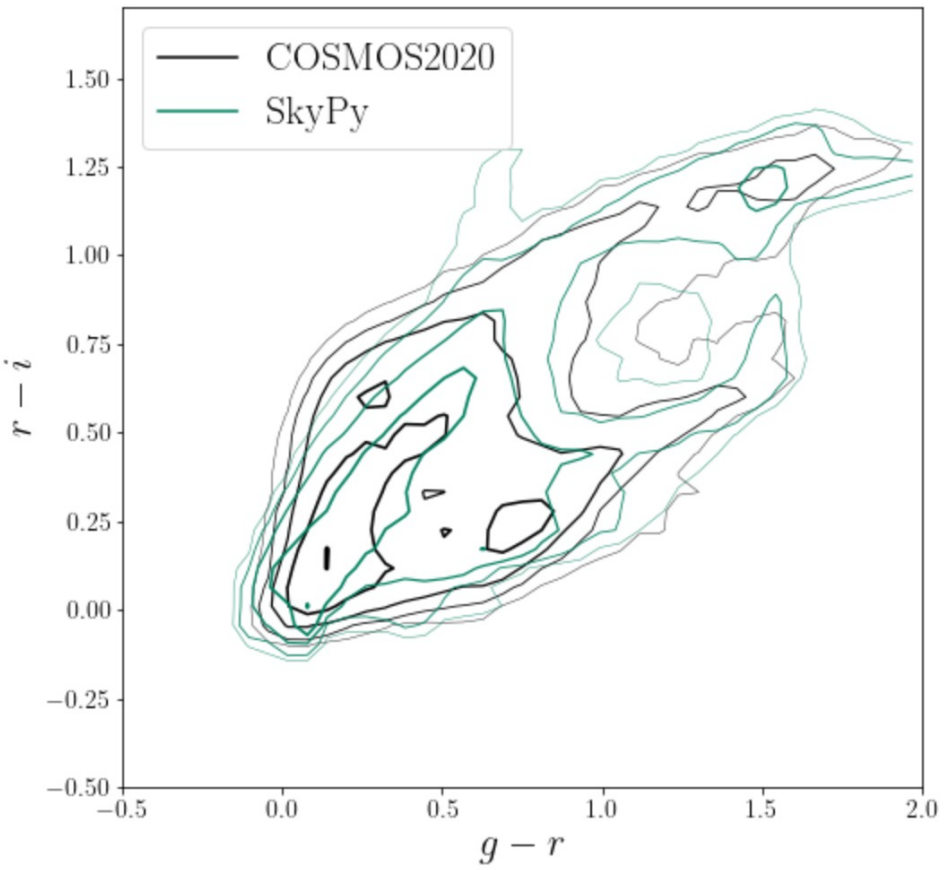
$$\sigma_m(F|F_{\rm noise}, t_{\rm eff}) = \frac{2.5}{ln\ 10}\left[\frac{1}{Fkt_{\rm eff}}\left(1+\frac{F_{\rm noise}}{F}\right)\right]^{1/2} \qquad F_{\rm noise} = \frac{F_{\rm lim}^2 kt_{\rm eff}}{10^2} - F_{\rm lim} \qquad ln\ t_{\rm eff} = a + b(m_{\rm lim} - 21)$$

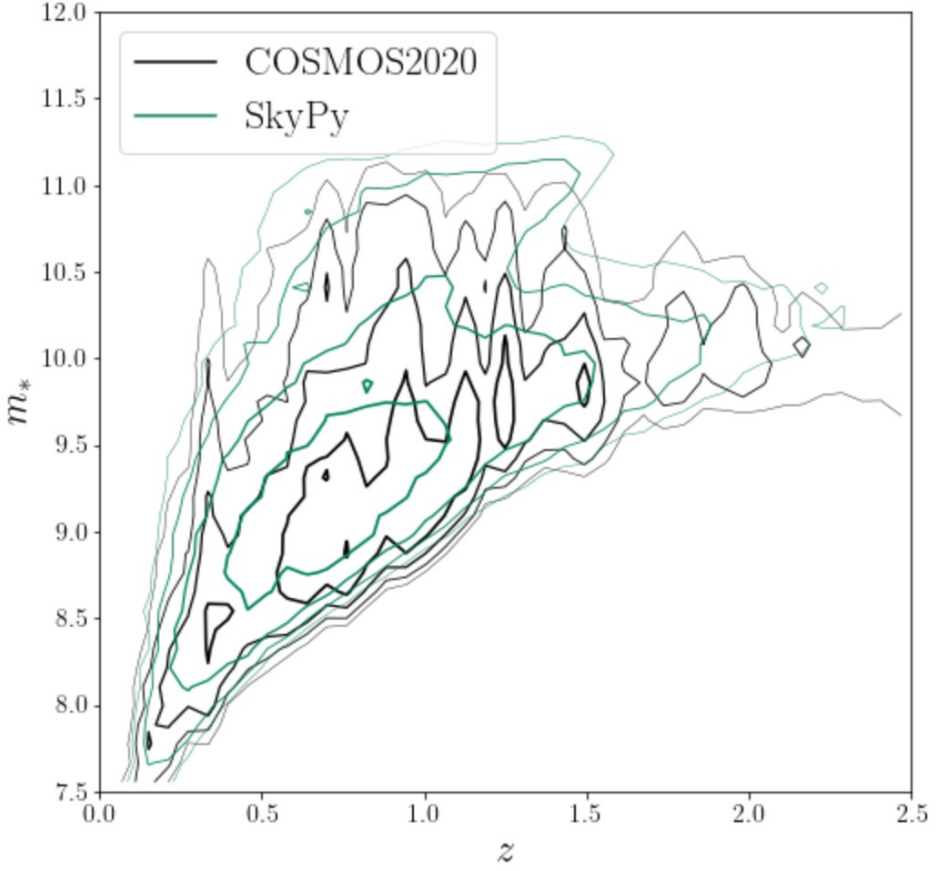




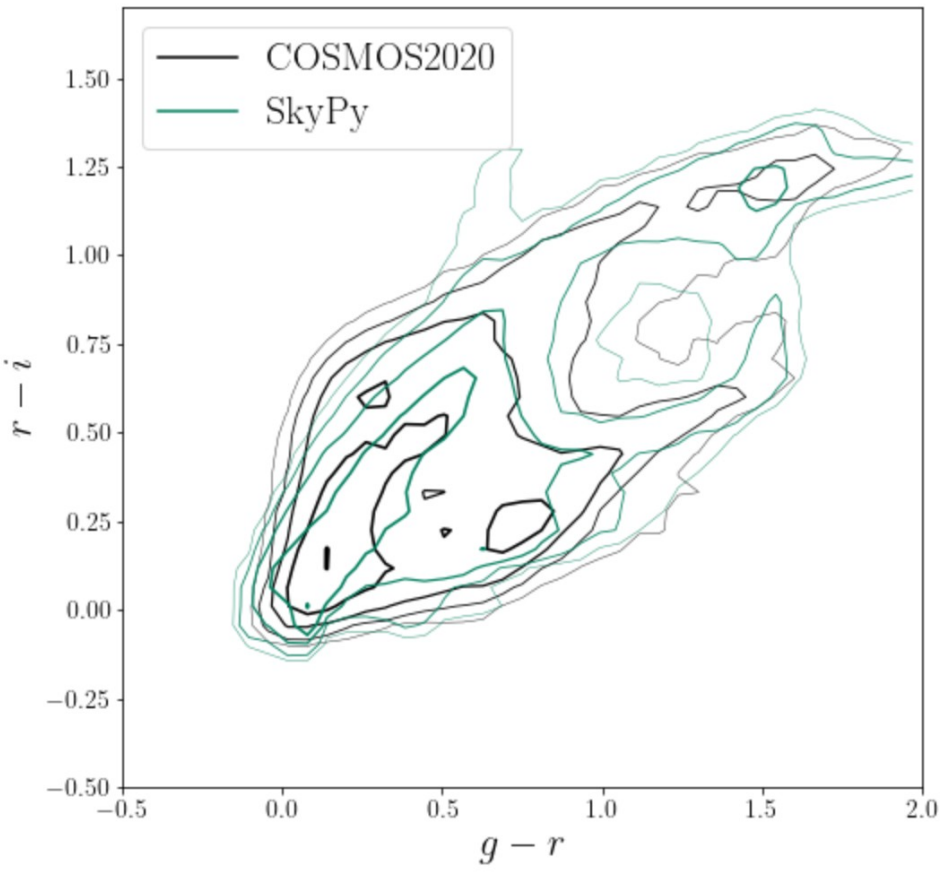
Current performance

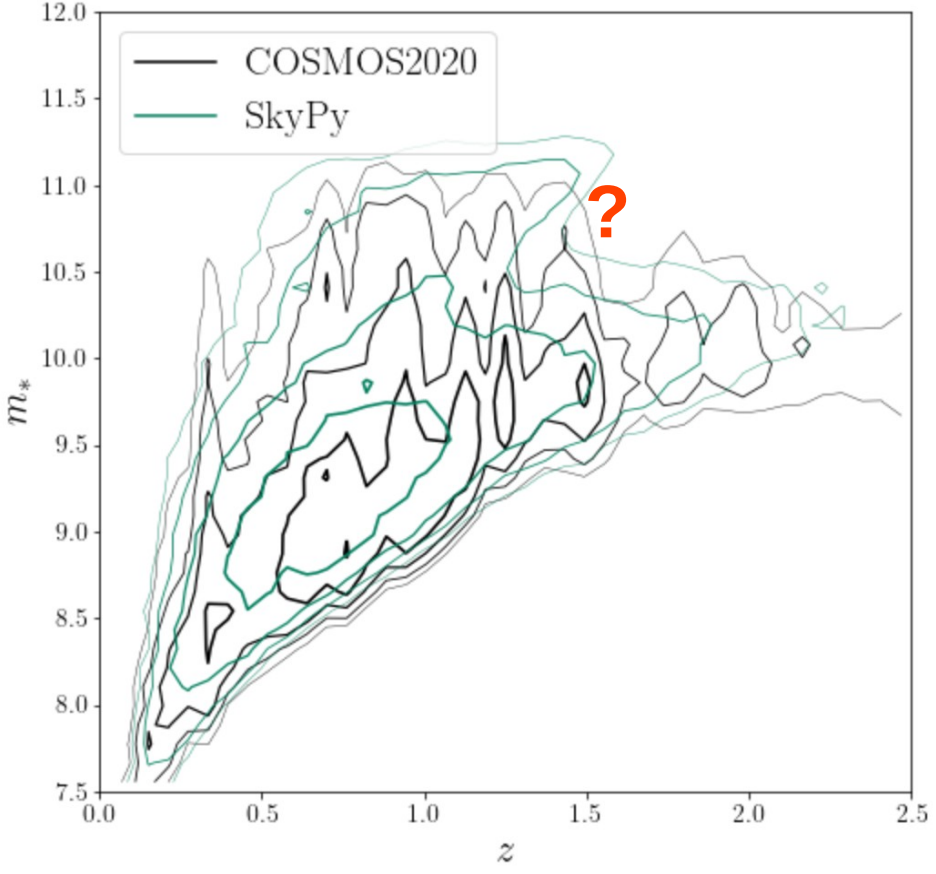
SkyPy





Current performance



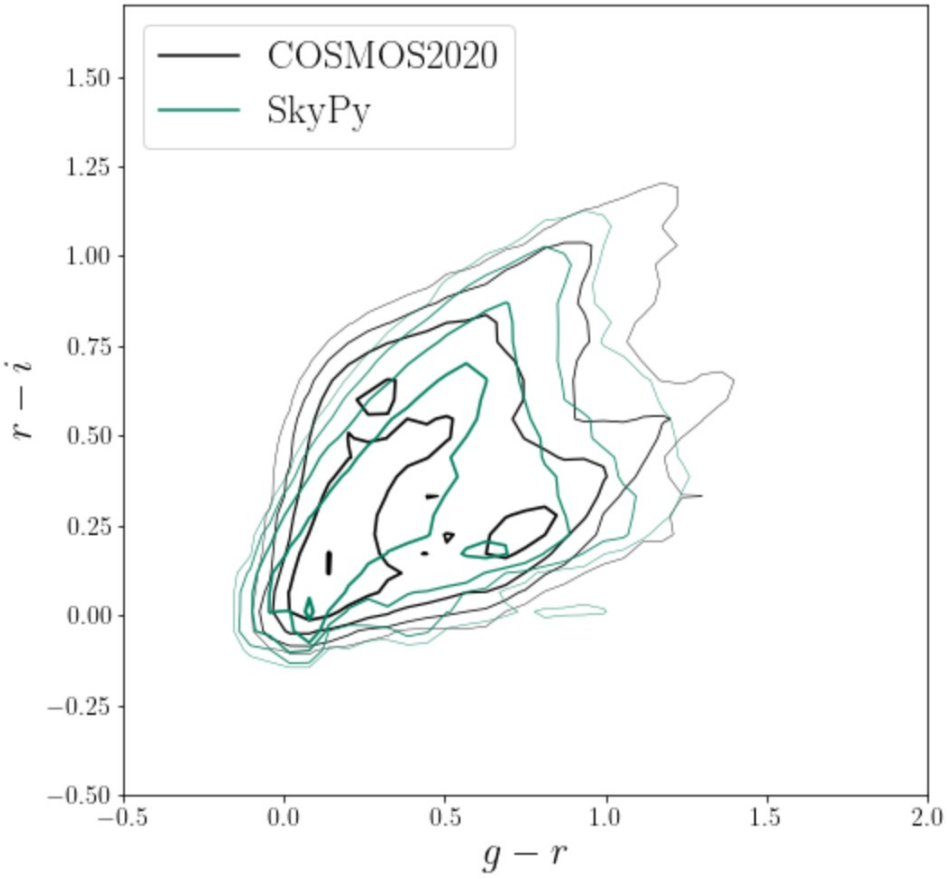


SkyPy

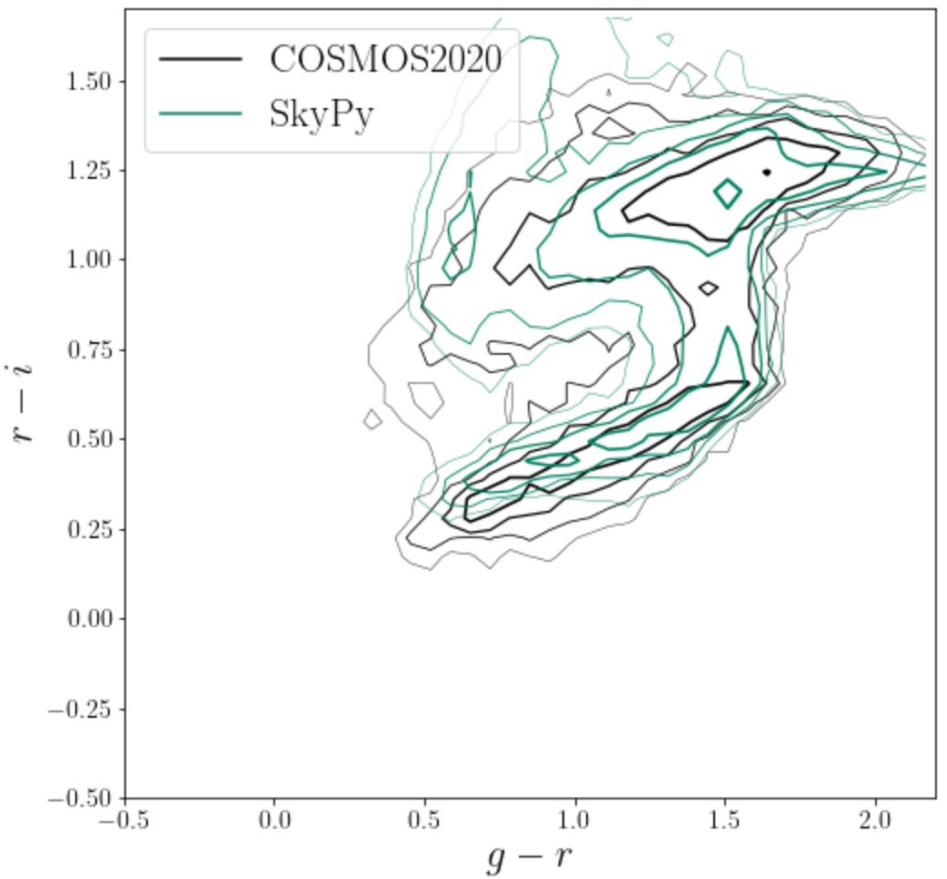
Current performance

SkyPy

Star-forming / blue



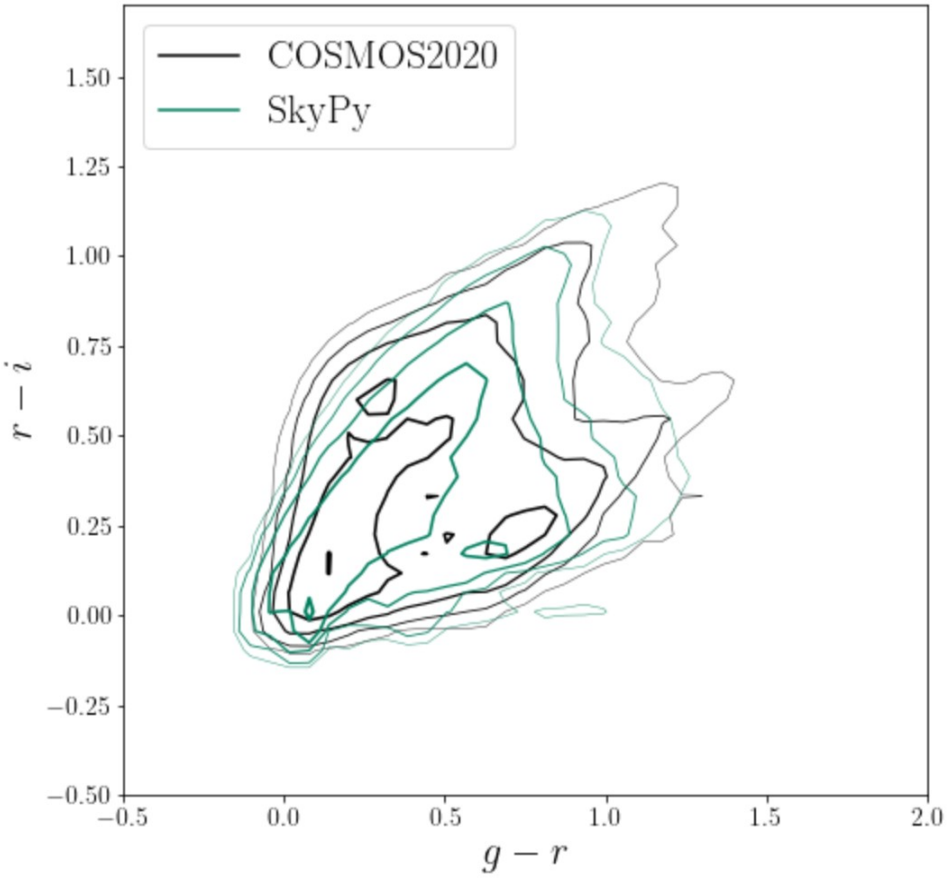
Passive / red



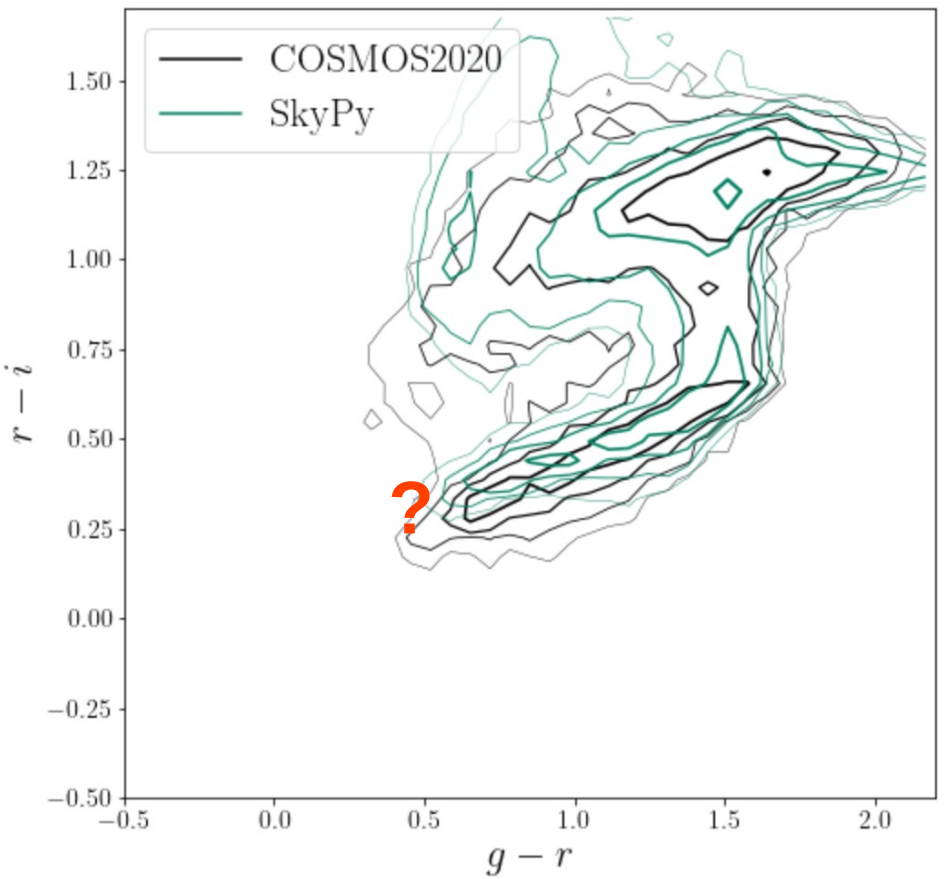
Current performance

SkyPy

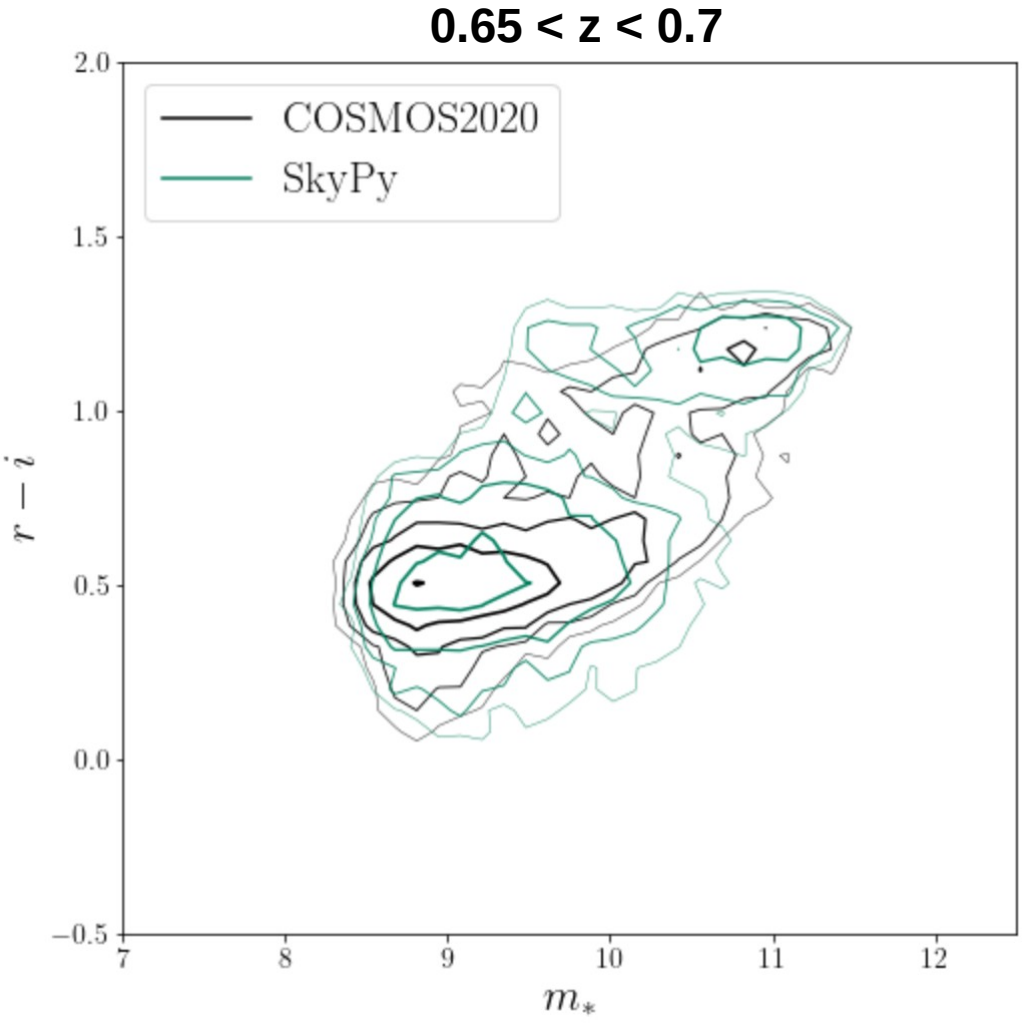
Star-forming / blue



Passive / red



Current performance



SkyPy

Remaining issues + Future direction

Issues for current work:

- Resolve the mass / luminosity dependence issue
- Change the error to be computed in flux space (mostly done).

Future:

- Develop modules → power spectrum (haloes), positions
- Add breadth (multi-wavelength, time domain, multi-messenger)
- Use the software for science! → Marine Leyvraz's project
- Develop and update website...