

From Photoionisation Models to Machine Learning : Interpreting Star Formation and AGN in High-Redshift Galaxies

Adele Plat

With Michaela Hirschmann

Laboratory for galaxy evolution and spectral modelling, EPFL, Switzerland

April 1, 2026



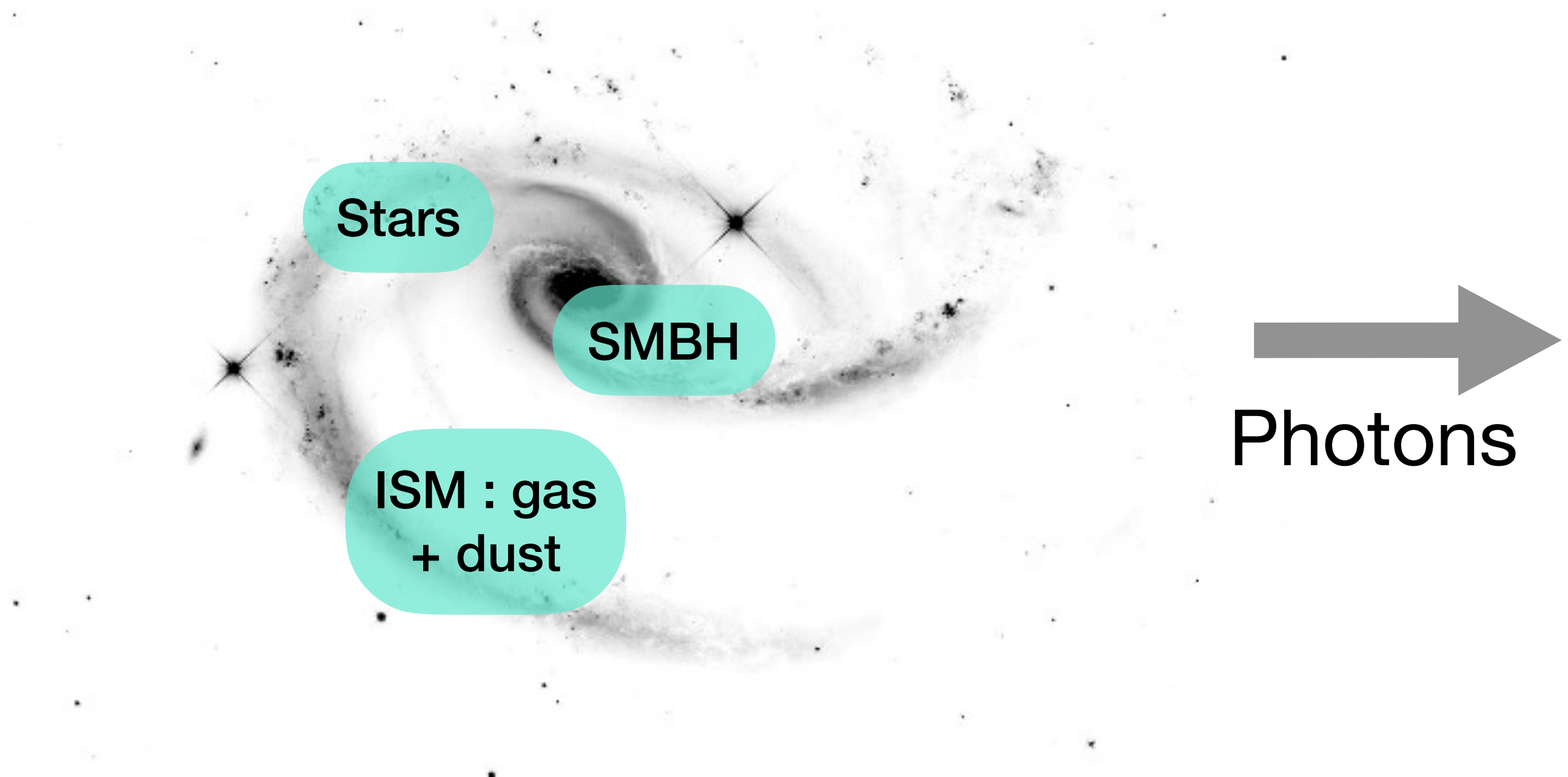
GALSPEC

EPFL

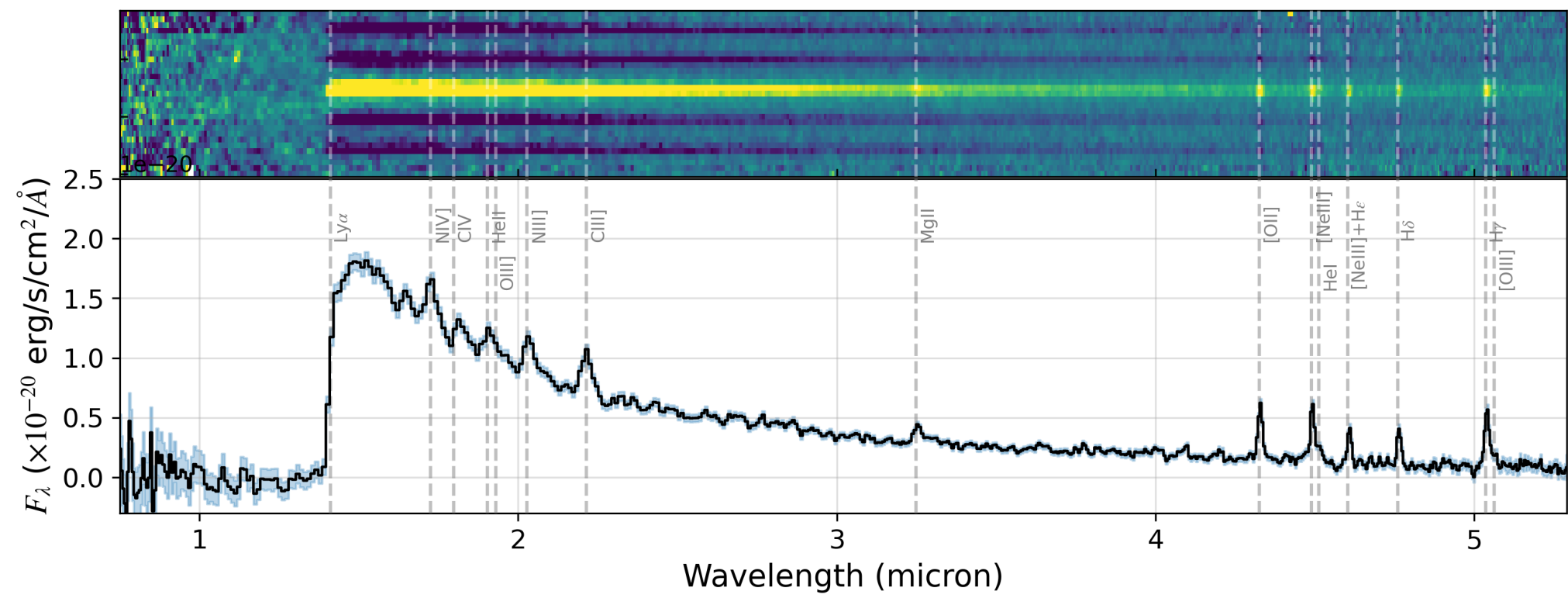


**Swiss National
Science Foundation**

Large samples of observed galaxy spectra over time



Observed galaxy spectra

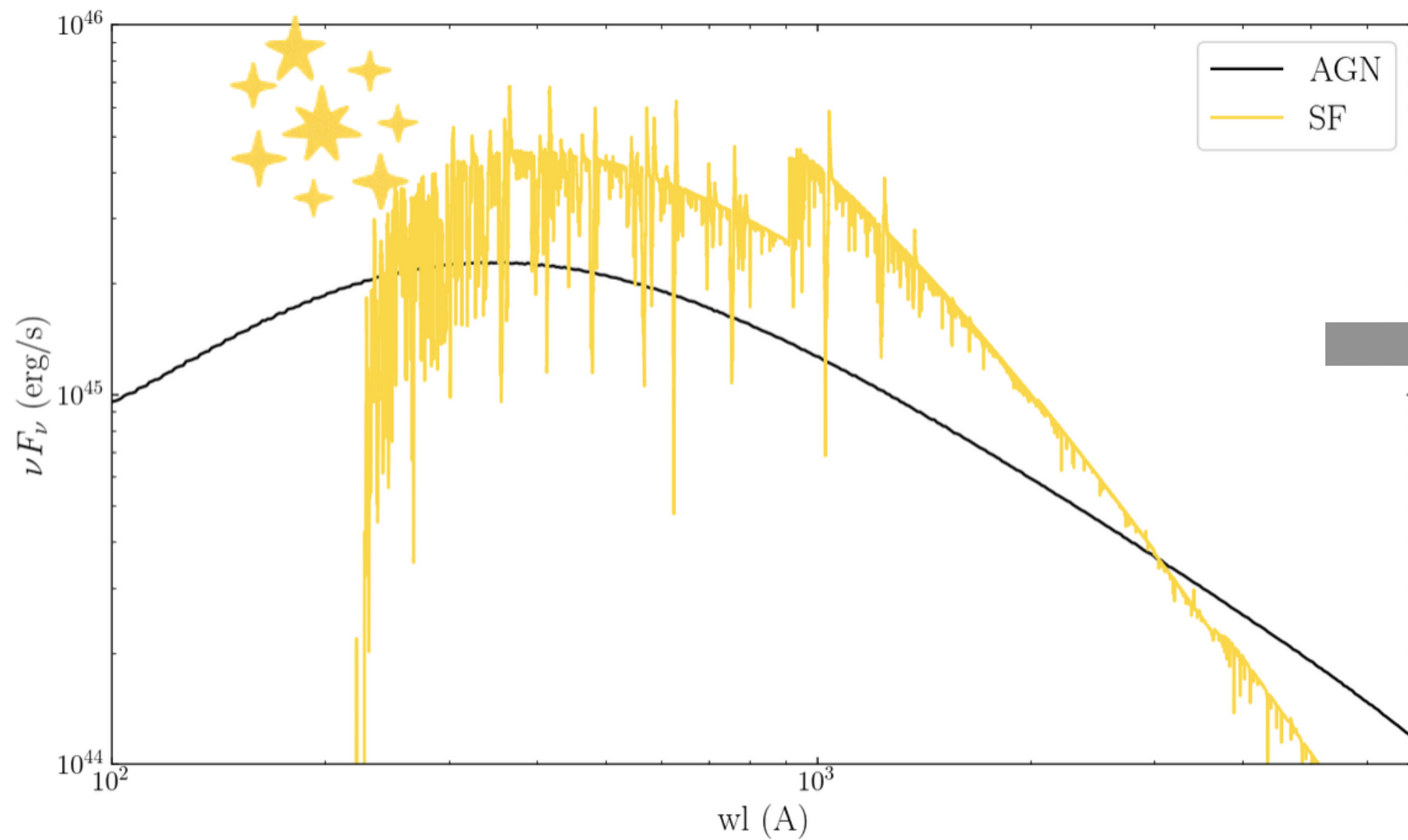


JWST Spectrum of a high-z galaxy

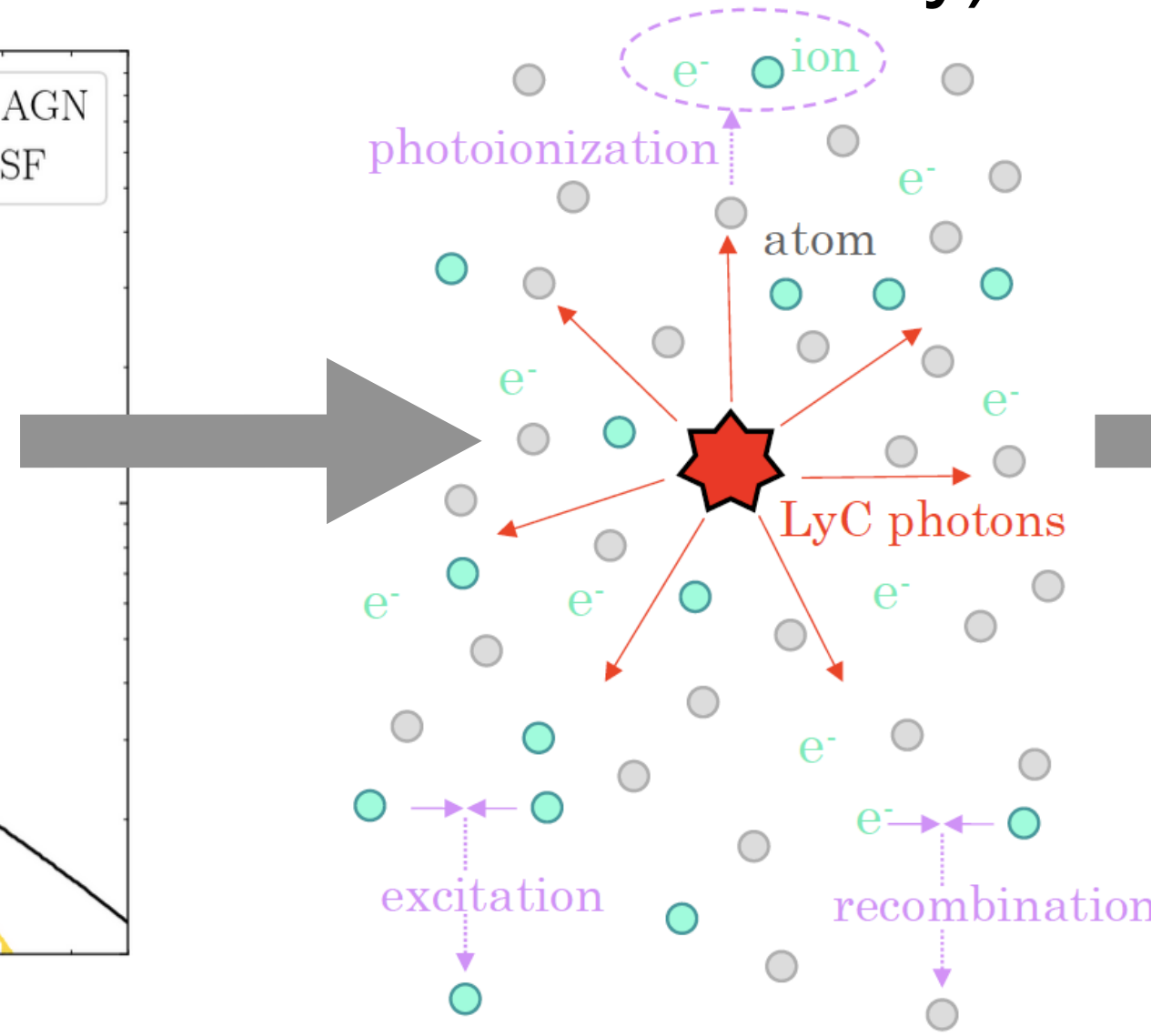
➡ How can we make sense of what we see, i.e. how can we infer redshift and physical galaxy properties from spectra?

Model galaxy spectra with emission lines

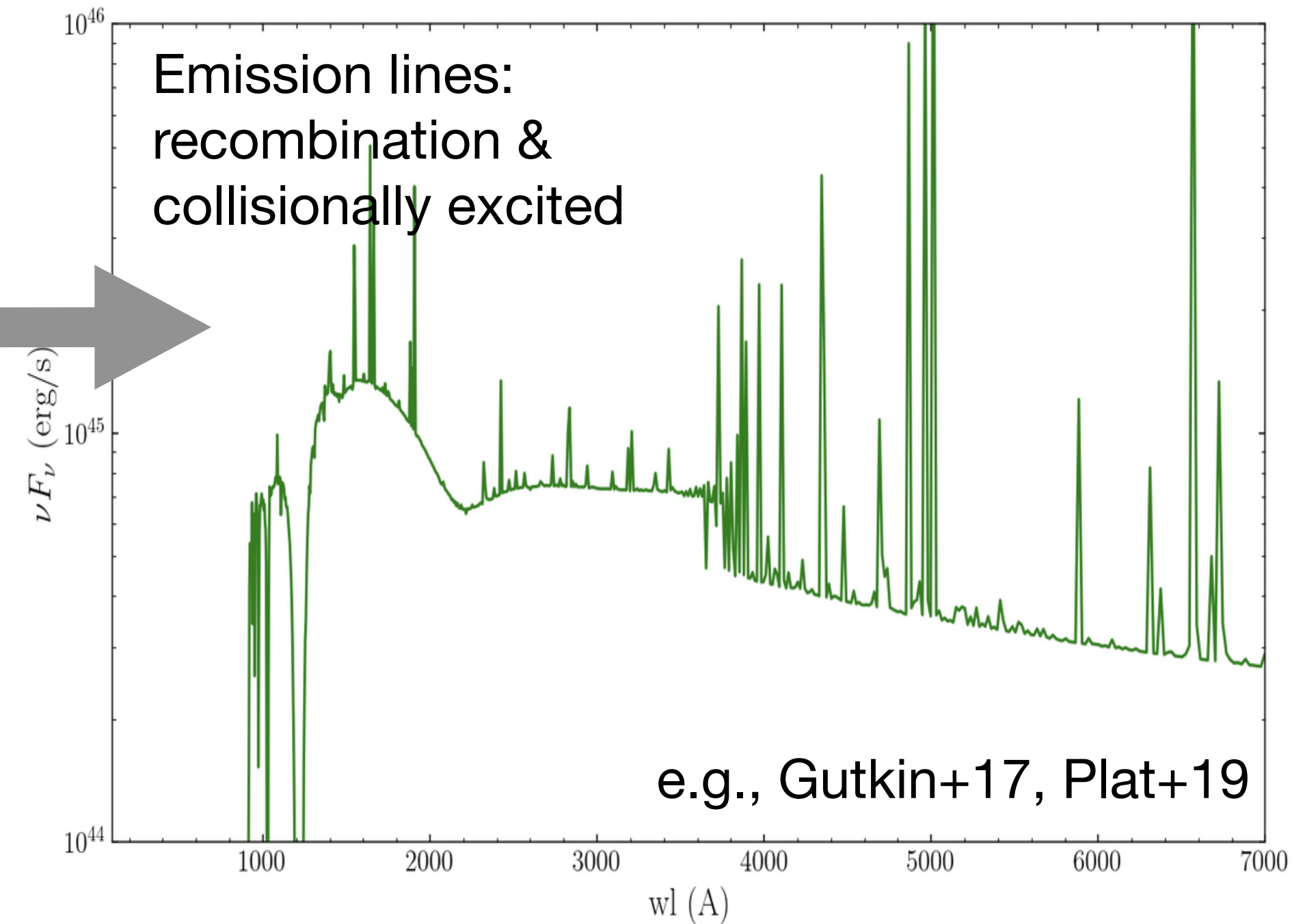
Source of ionising radiation (e.g., young massive stars, accreting BH)



Transmission through gas cloud with photo-ionisation code (e.g., Cloudy)



Transmitted spectrum with emission lines from HII regions

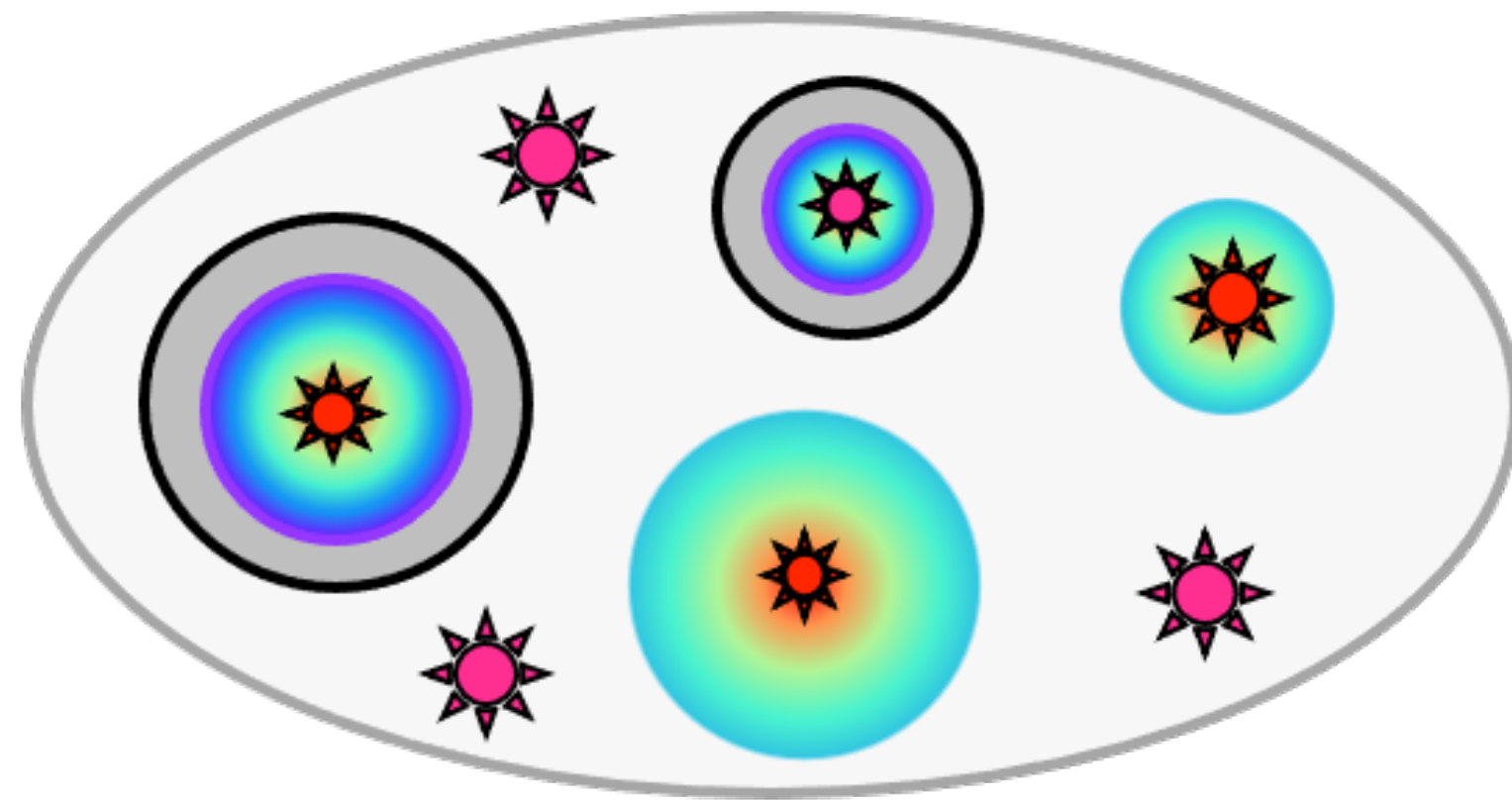


Emission-lines can be used to characterise the ionizing source and the gas properties :

- **HII regions:** information on SFR, gas metallicity, gas density, gas temperature, dust, type of ionising source
- **Ionised regions around BHs:** information on BH mass, gas accretion rate, gas metallicity, gas density, gas temperature, type of ionising source

Components of spectra: stellar radiation & HII regions

Star-forming galaxy models computed following the approach of Gutkin+16.



‘isochrone-synthesis’ technique:
star-formation history expanded
into a series of simple stellar
populations (SSP)

$$L_{\lambda}(t) = \int_0^t dt' \psi(t-t') S_{\lambda}(t', Z) T_{\lambda}(t, t')$$

Galaxy luminosity
at time t

Star-formation rate

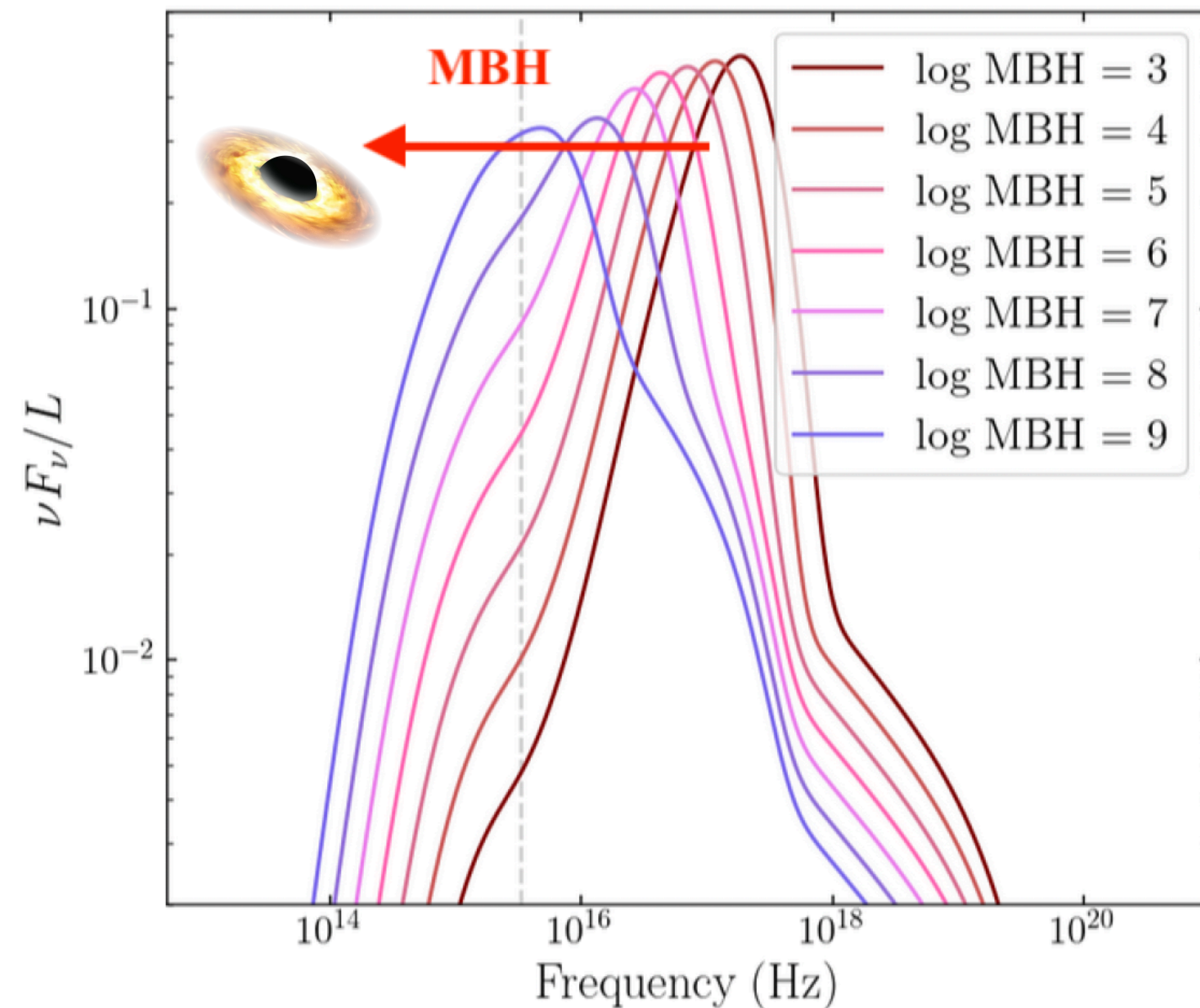
Luminosity of
an SSP of age t'

Transmission
of the ISM

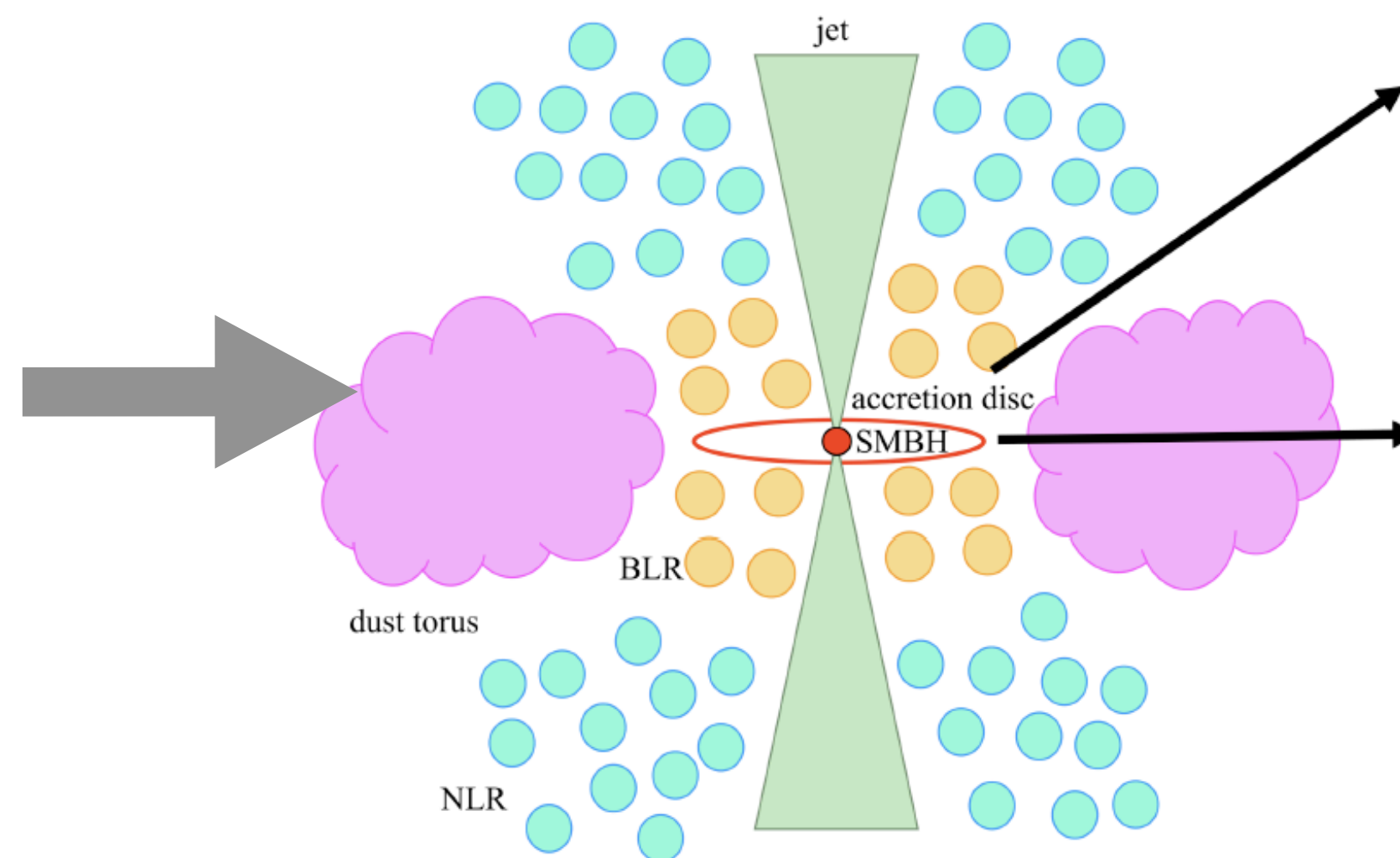
- Large model grid with free parameters being: SFH, age and stellar mass, stellar metallicity, upper mass cutoff of the IMF, gas density, gas metallicity, ionisation parameter, dust-to-metal-mass ratio, element abundances such as C/O or N/O, LyC escape fraction etc.

Components of spectra: AGN

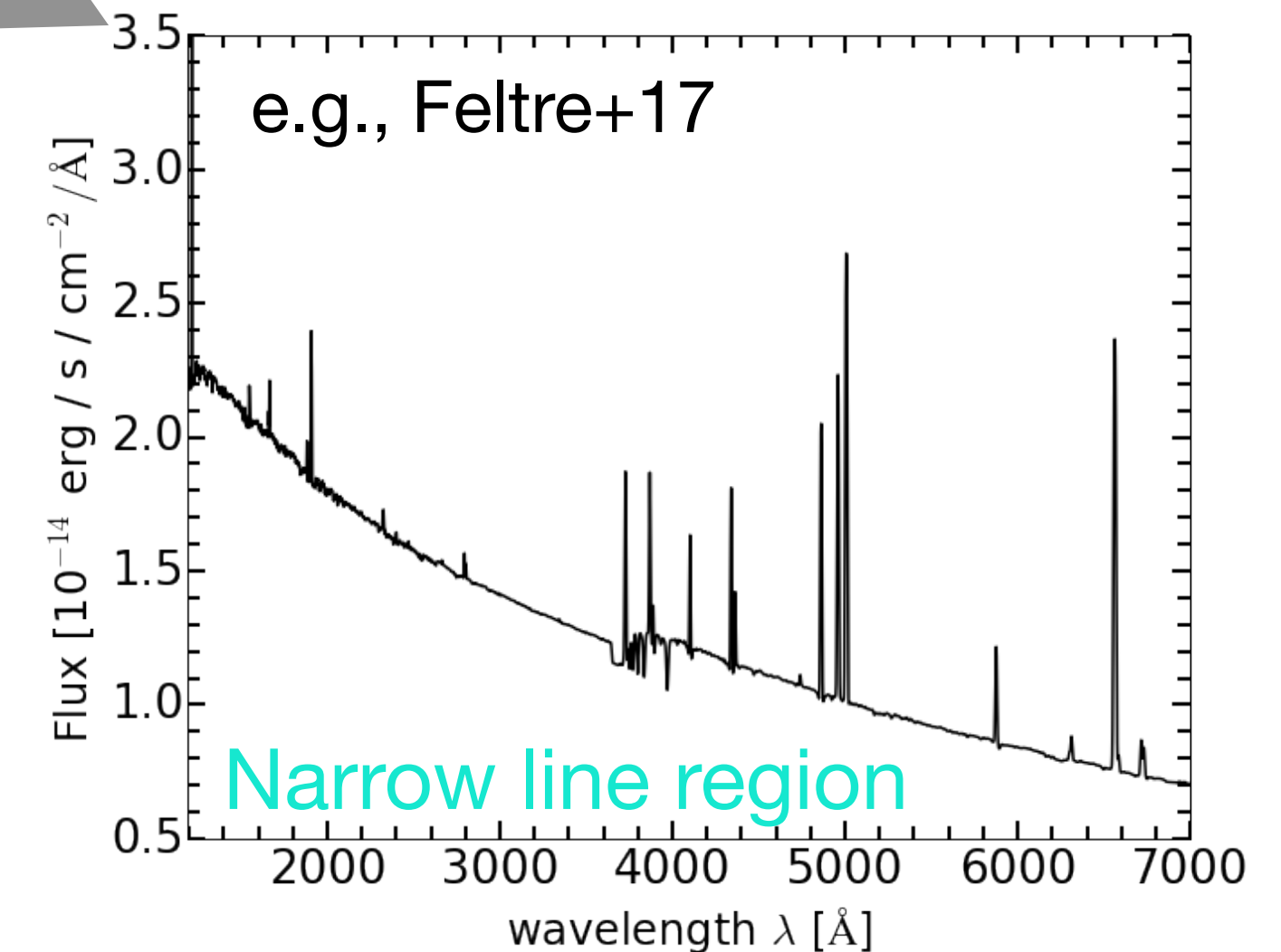
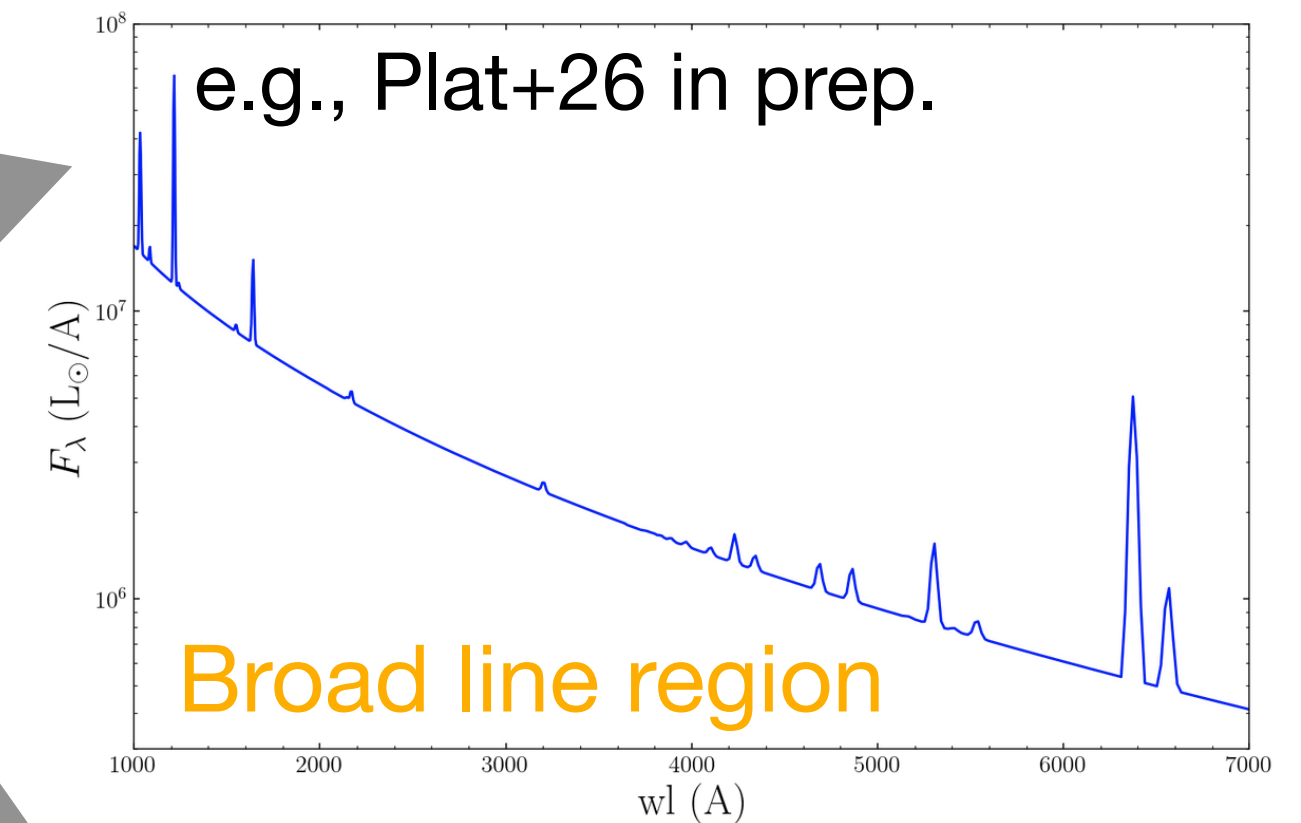
AGN continuum from theoretical models, such as Kubota & Done'18



Transmission through gas cloud with photo-ionisation code (e.g., Cloudy)



Transmitted spectrum with narrow & broad emission lines

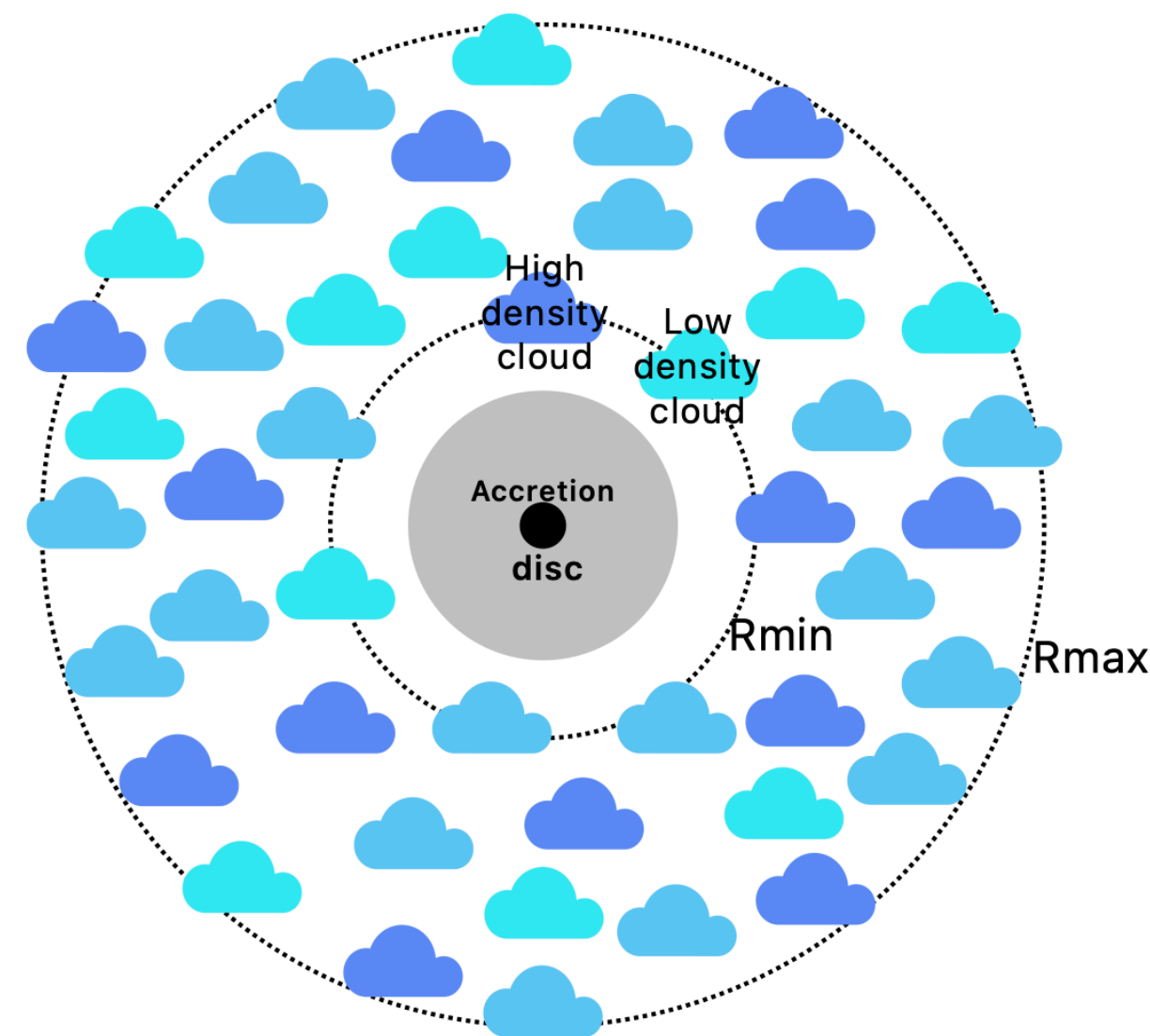


- Large model grid with free parameters being: BH mass, BH accretion rate, gas density, gas metallicity, ionisation parameter, dust-to-metal-mass ratio, element abundances such as C/O or N/O

Example of application : find the seeds of SMBH

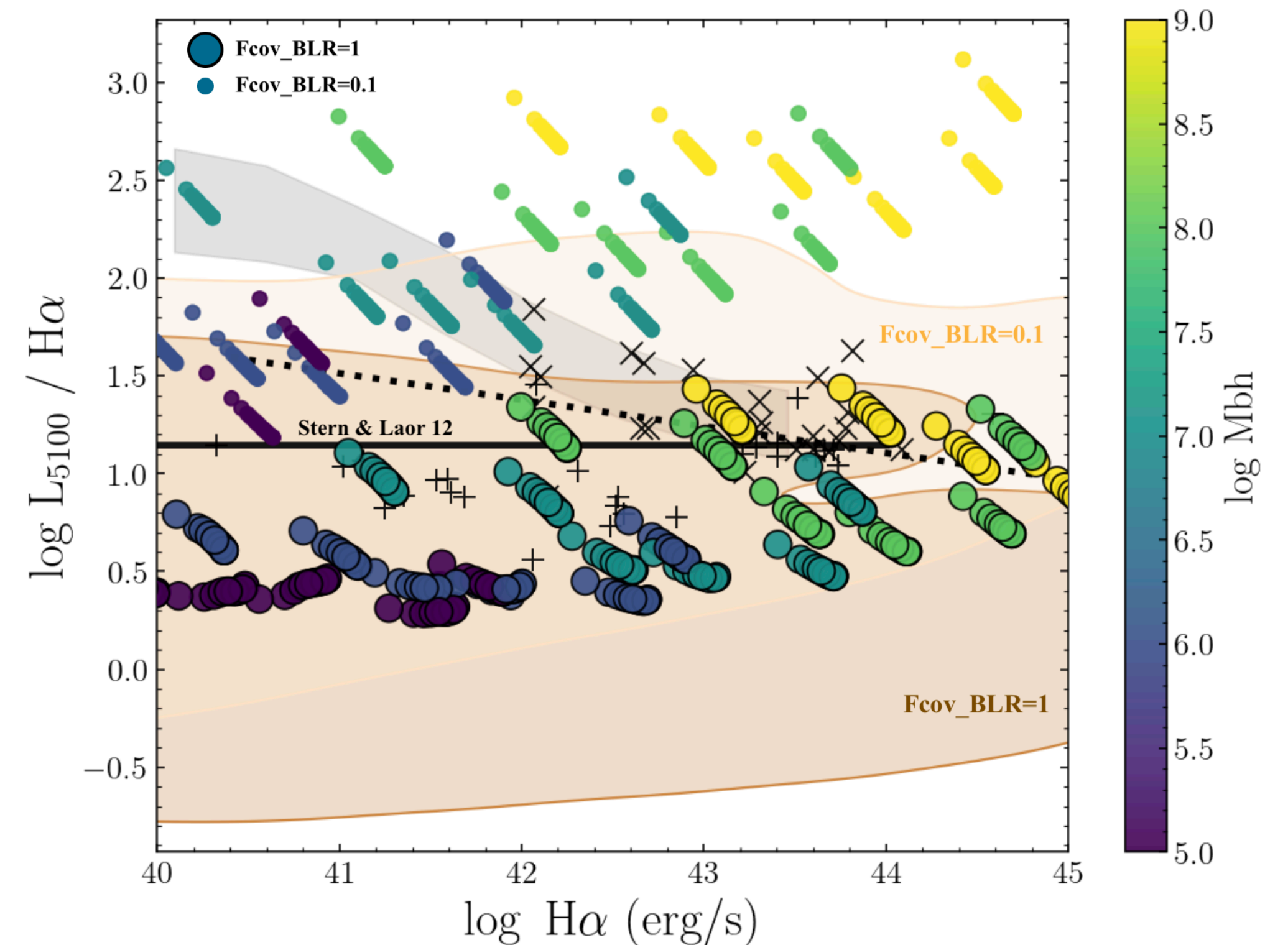
- Can we actually observe the BH seeds with JWST ?
- Can we infer accurate BH mass and Eddington ratio in high z AGN from IMBH and SMBH ?
- Can we find signatures to differentiate the different types of BH seeds at high redshift ? How to distinguish between different sources of ionization in high z galaxies, including pop II and pop III ?

BLR models



Face on view

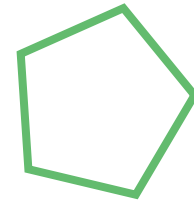
Estimates AGN accretion luminosity



Bolometric correction used to estimate accretion luminosity depends on BH mass, F_{edd} , covering factor of the BLR, gas column density of the BLR ... etc which may be very different at high redshift compared to relations calibrated at low redshift

Example of application : find the seeds of SMBH

- SF component :
C&B stellar population
(Gutkin+16)
constant SF at 3Myr and
10Myr
Z : 0.0001 to 0.060



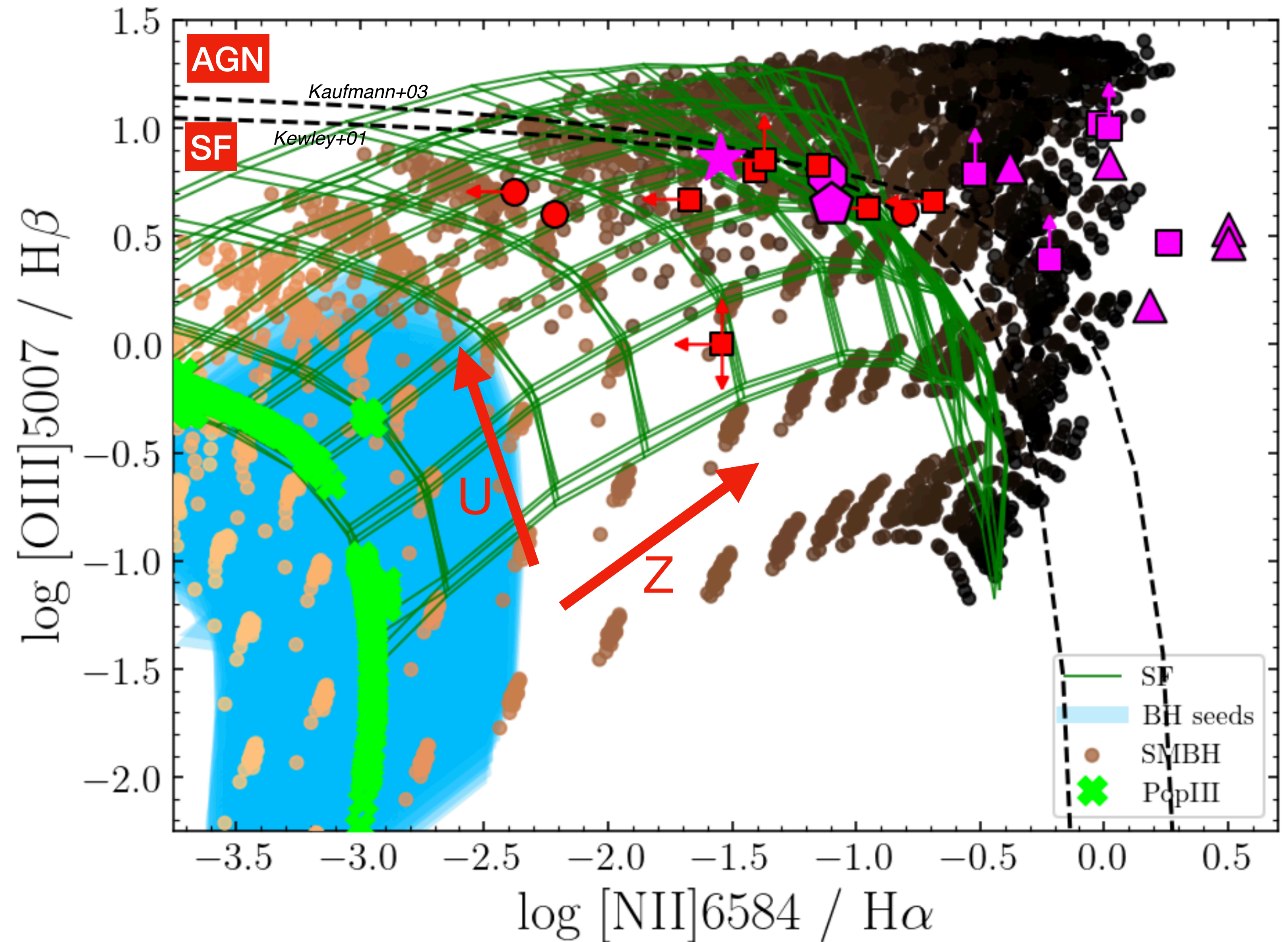
- Pop III :
GALSEVN (Lecroq+25)
SSP from 0yr to 40Myr
Z_ISM = 1e-11, 1e-4



- AGN NLR IMBH &
SMBH :
Kubota & Done SED
MBH : 10^6 to 10^9 Mo
Z : 0.0001 to 0.060

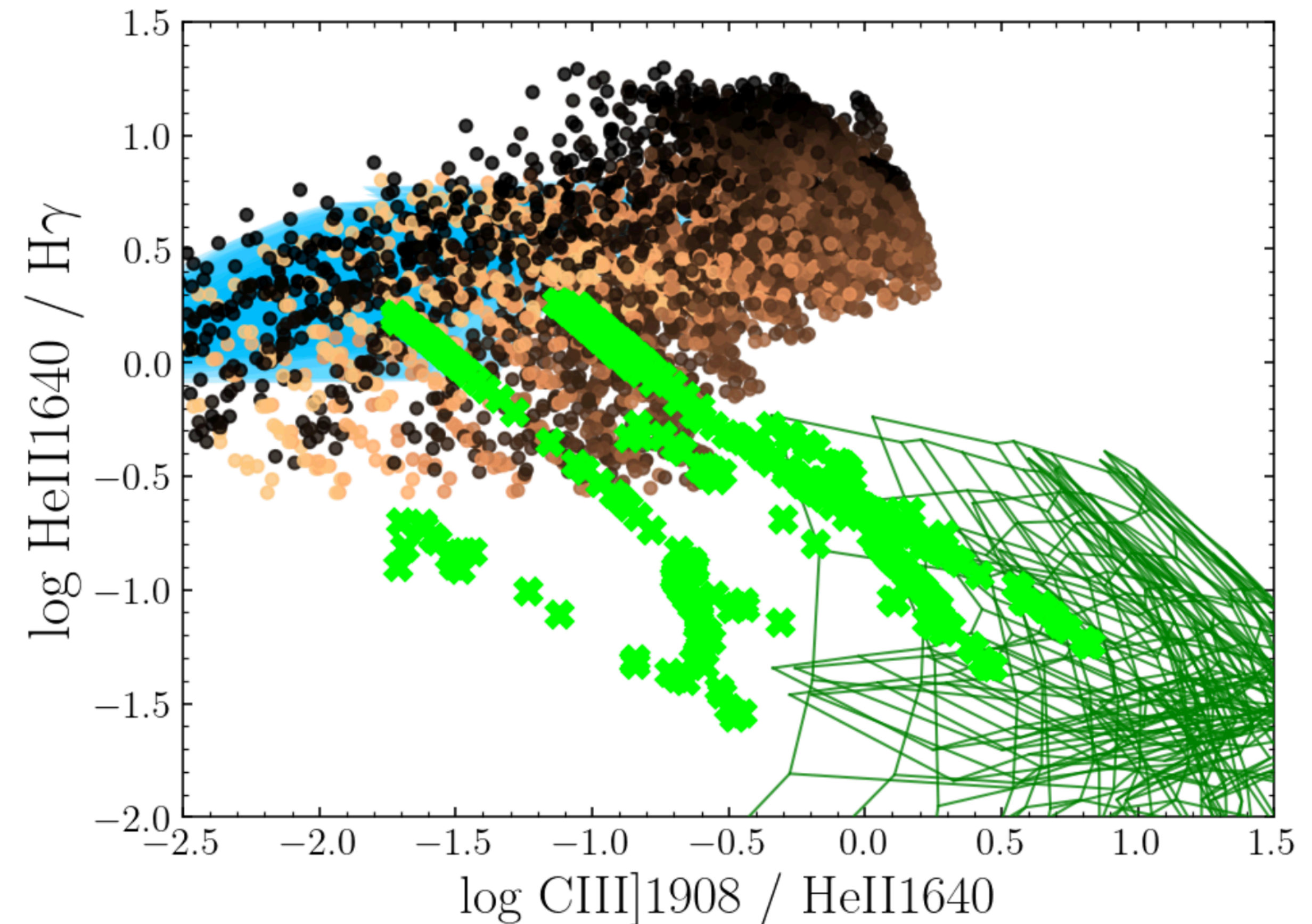
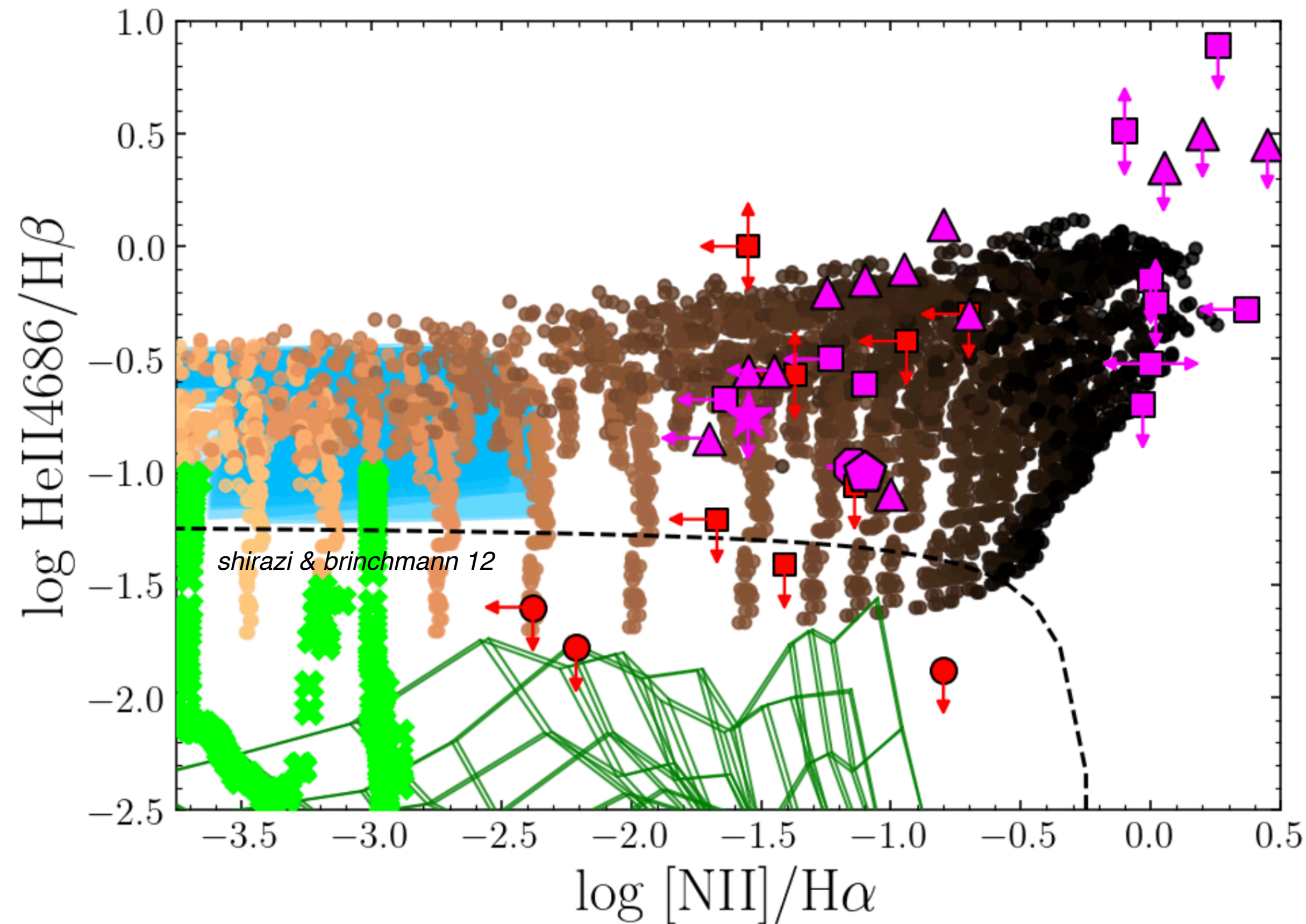


- AGN NLR BH seeds :
Z < 0.001
MBH : 10^3 to 10^5 Mo



Example of application : find the seeds of SMBH

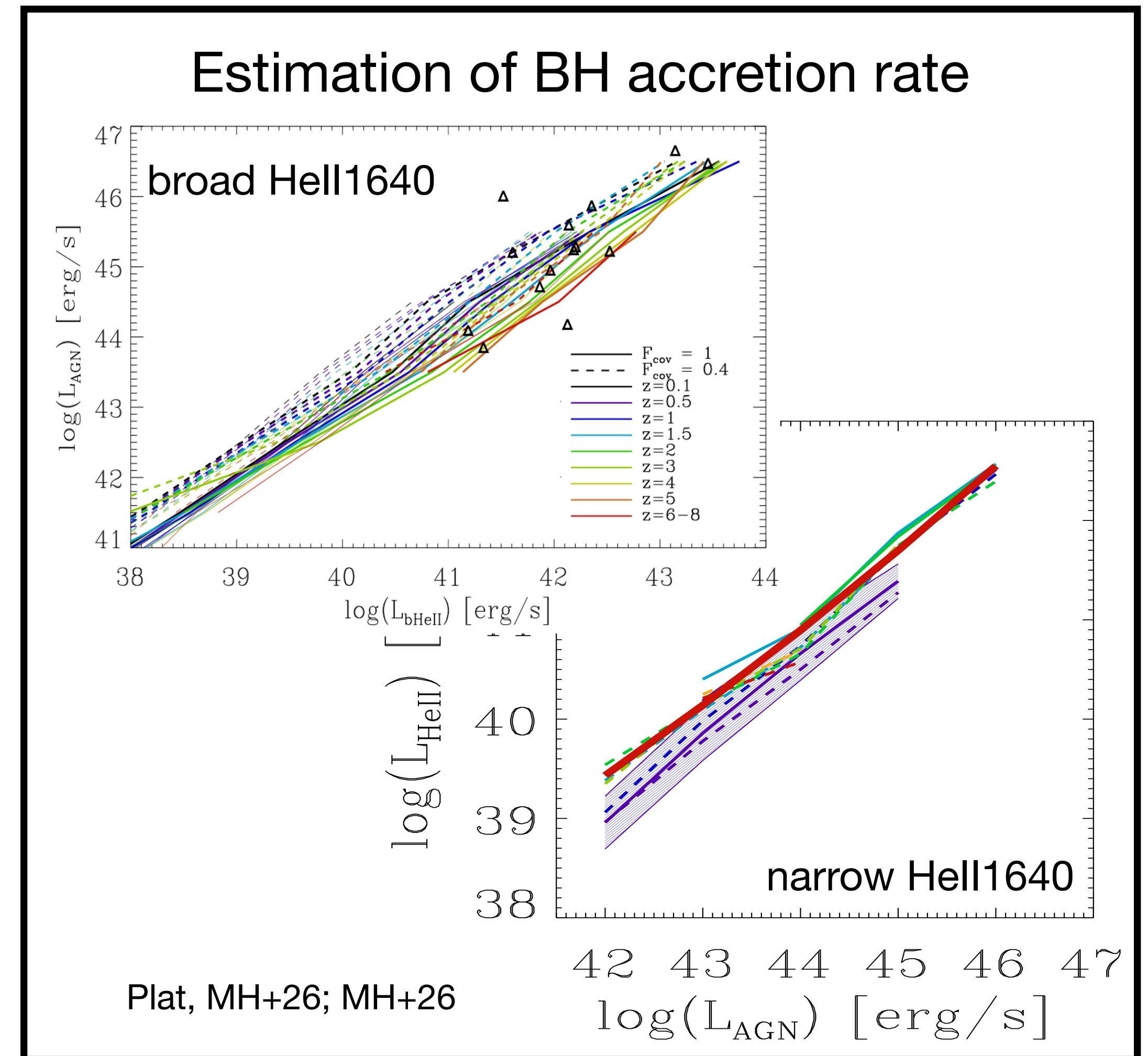
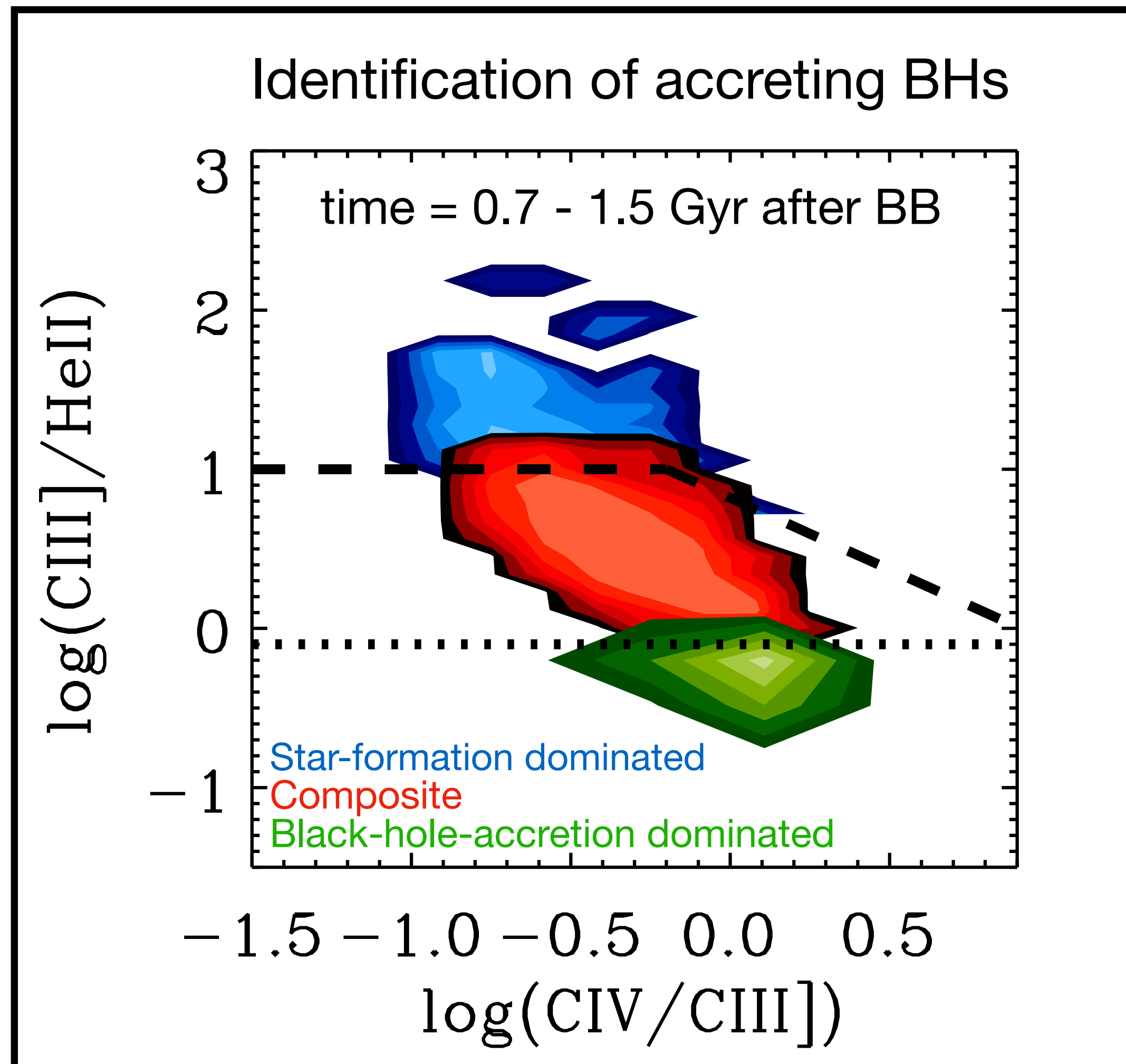
Some diagrams work a bit better to separate AGN emission from star-forming emission



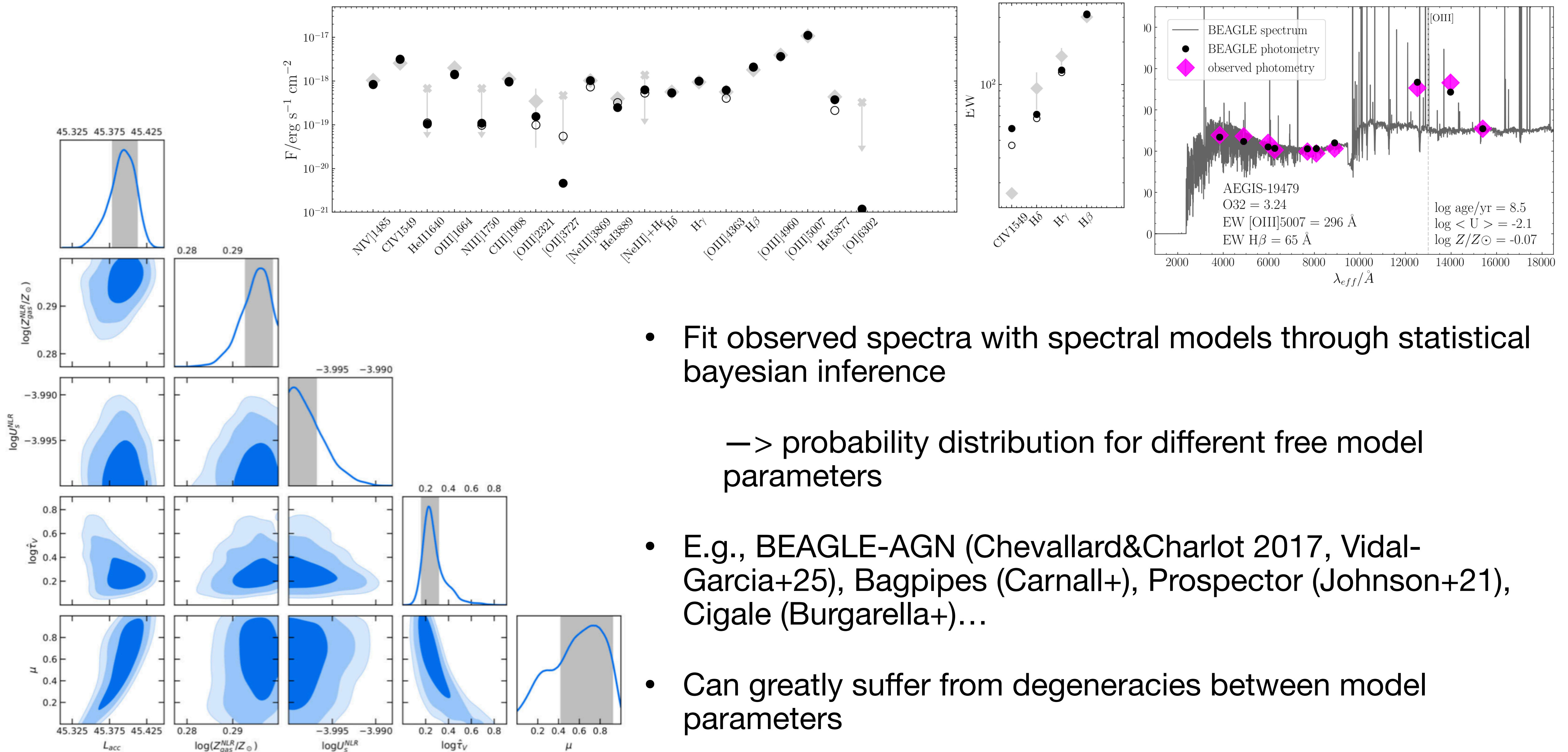
But lots of degeneracy (U, Z, C/O, N/O, n_H ...) prevent identification of BH seeds using single diagnostic diagram for type II AGN $\rightarrow$ use multiple emission lines / diagnostics at once

Emission line properties of simulated galaxies

- **Alternative:** Couple model spectra with emission lines to galaxies from cosmological simulations (line-emitting regions unresolved) by matching free parameters (e.g., Hirschmann+17, 19, 23)
- Reduces free parameters space to physically motivated combinations at given cosmic time



Estimate physical properties from observed spectra using Bayesian inference



- Fit observed spectra with spectral models through statistical bayesian inference
 - > probability distribution for different free model parameters
- E.g., BEAGLE-AGN (Chevallard&Charlot 2017, Vidal-Garcia+25), Bagpipes (Carnall+), Prospector (Johnson+21), Cigale (Burgarella+)
- Can greatly suffer from degeneracies between model parameters

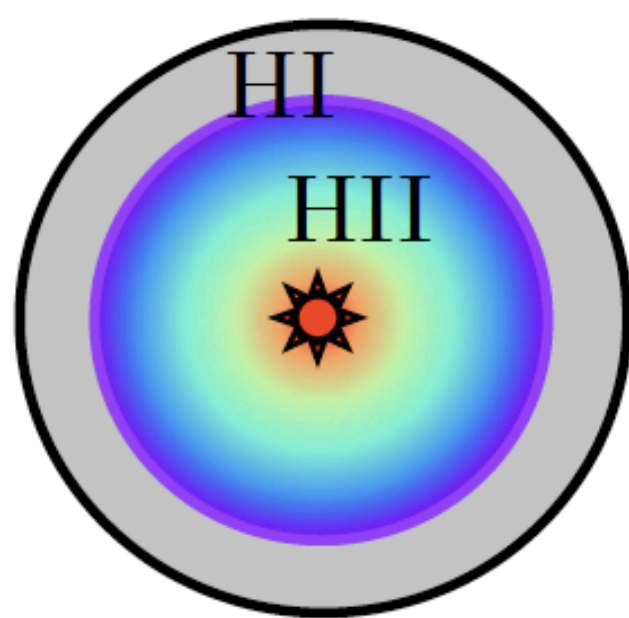
Summary and outlook

- Extensive library of ca 2 million spectra with emission lines (optical and UV)
 - for star-forming galaxies
 - for broad-line and narrow-line AGN
- Can be used to interpret galaxy spectra and infer physical properties
 - Using line ratio diagnostic diagrams (can be coupled to cosmological galaxy simulations to reduce parameter space)
 - Bayesian statistical fitting codes
- Assess “new” ways to interpret galaxy spectra with ML/AI techniques
 - Can we find new emission-line diagnostics to identify low mass BH at high z ?
 - Use AI to infer galaxy physical properties such as stellar age, metallicity, AGN fraction, ... etc

Back-up slides

What are HII regions?

Stars with effective temperature $\gtrsim 30,000$ K ionize the interiors of their birth clouds $\Rightarrow$ **HII regions**



ISM: Mixture of gas and dust composed of H, He and metals.

$\sim 70\%$

$\sim 28\%$

$\sim 2\%$ (solar)

Typical properties of HII regions:

Hydrogen density: $n_{\text{H}} \sim 10\text{--}10^4 \text{ cm}^{-3}$

Temperature: $10,000\text{--}50,000$ K

Size: $1\text{--}100$ pc

Ionization potential H: 13.6 eV, He: 24.6 eV and 54.4 eV



Emission from an HII region (see Gutkin+16):

✿ Stellar population synthesis codes:

- GALAXEV (Bruzual&Charlot03, 2019 version): single stars
- BPASS (Stanway&Eldridge18, 2019 version): single and binary stars

➡ Stellar parameters :

- metallicity $Z = 0.0005, 0.002, 0.008$ ($Z_{\odot} = 0.01524$)
- Chabrier IMF with upper mass cutoff $m_{\text{up}} = 100, 300$ and $600 M_{\odot}$

✿ Photoionization code CLOUDY (Ferland+17)

➡ Nebular parameters :

- Interstellar (gas- + dust-phase) metallicity: same as that of the stars
- Hydrogen gas density $n_{\text{H}} = 100$ and 1000 cm^{-3}
- Zero-age volume-averaged ionisation parameter $\log \langle U \rangle = -3, -2$ and -1
- Dust-to-metal mass ratio $\xi_{\text{d}} = 0.1$ and 0.3
- Carbon to oxygen abundance ratio $\text{C/O} = 0.17$ and 0.44 (solar)

Interstellar-line absorption in stellar birth clouds (HII+HI) using the prescription of Vidal-García+17.

The production of ionizing radiation by young stellar populations

✿ Main-sequence OB stars:

Massive stars are hotter than lower-mass ones.

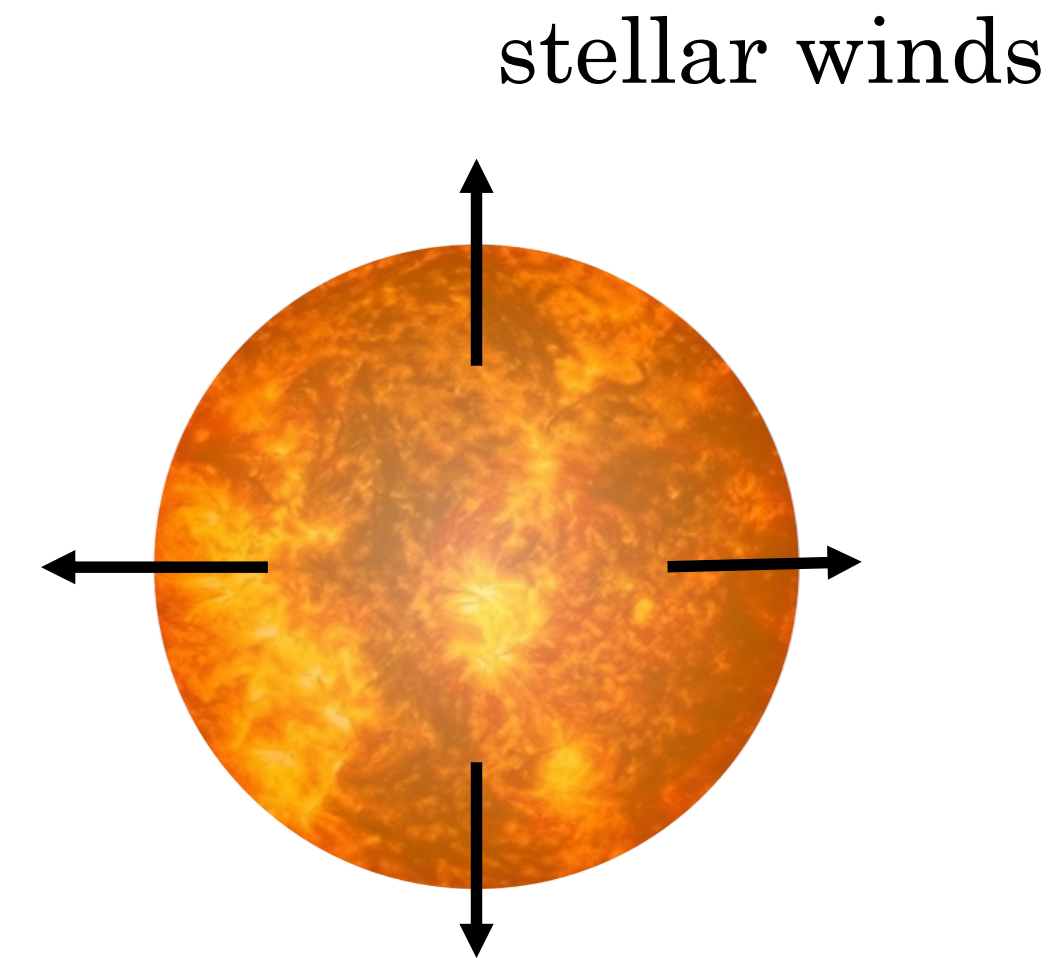
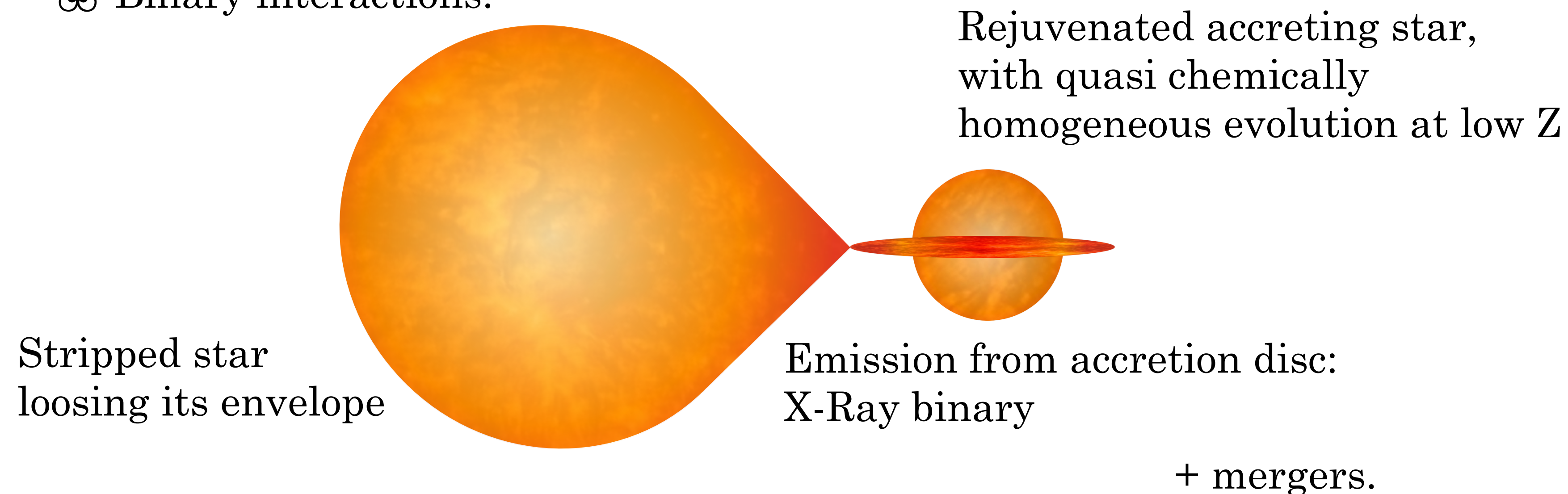
Metals increase atmosphere opacity, resulting in lower T_{eff} .

✿ Wolf-Rayet (WR) stars:

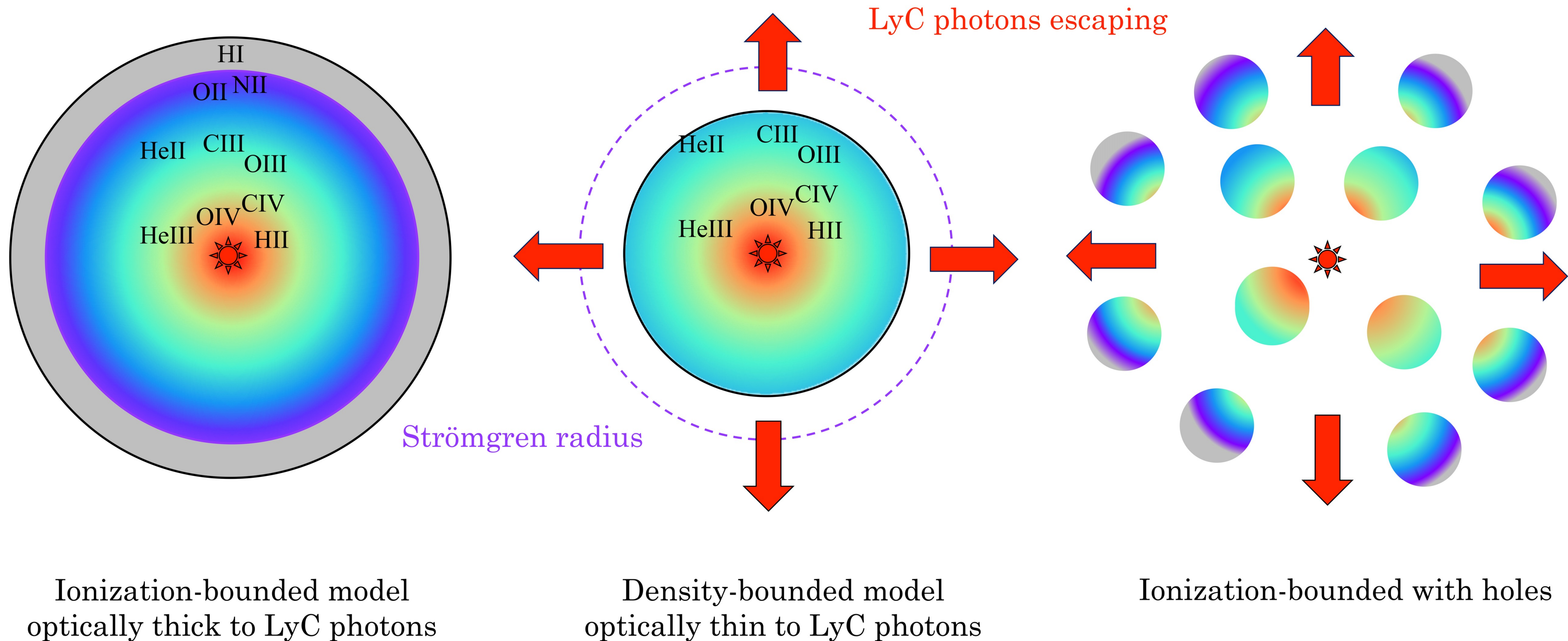
Hot massive stars, which have lost their hydrogen envelopes.

More WR stars at high Z because of higher mass-loss rate caused by line driven winds, but they produce a softer ionizing radiation.

✿ Binary interactions:



Escape of ionizing radiation: **density-bounded models**

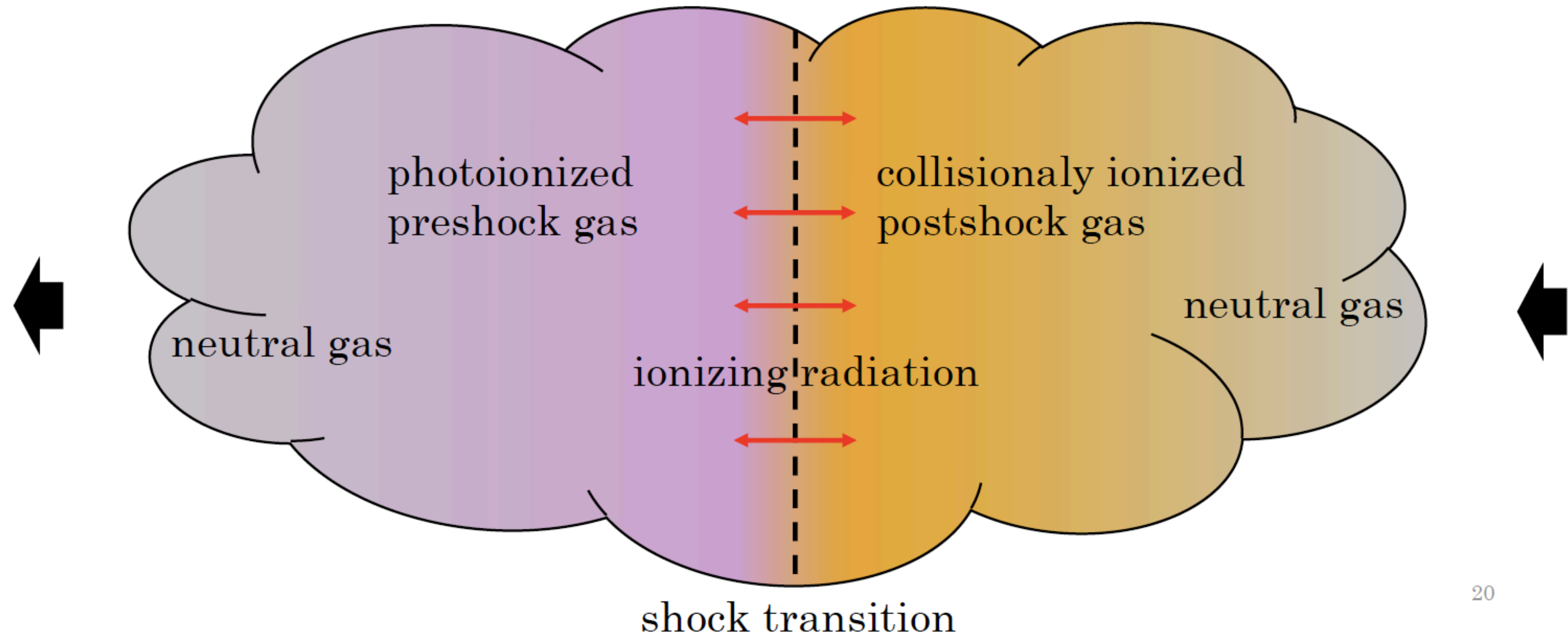


In density-bounded models, peeling off the HII region makes:

- f_{esc} larger;
- line EWs decrease (except for lines from high ionization species);
- line ratios of low-to-high ionization species decrease.

Component of **radiative shocks**: models from Alarie&Morisset19.

- Nebular emission from shock + precursor
- Magnetic field: $1\ \mu\text{G}$
- Shock velocity: from 100 to 1000 km/s
- Other nebular parameters identical to those of star-forming galaxy models
- Shocks contribute 90 % of the total HeII 1640 emission (1–60% of $\text{H}\beta$)
- No dust grains (destroyed in fast shocks)



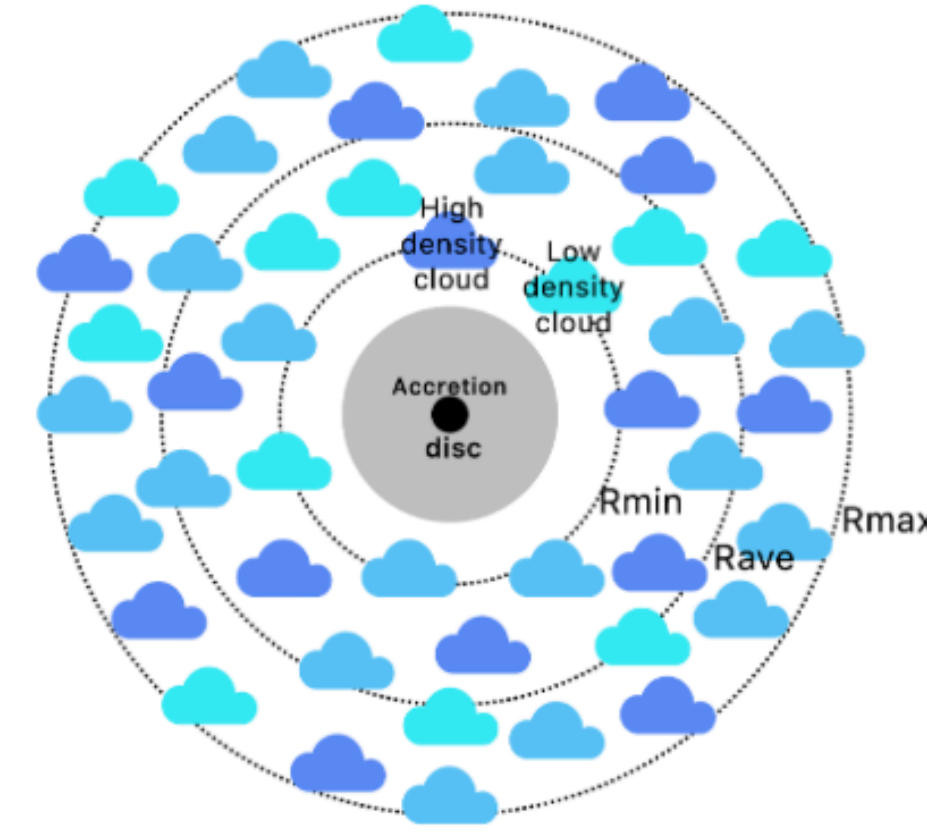
BLR parametrization :

SED parameters :

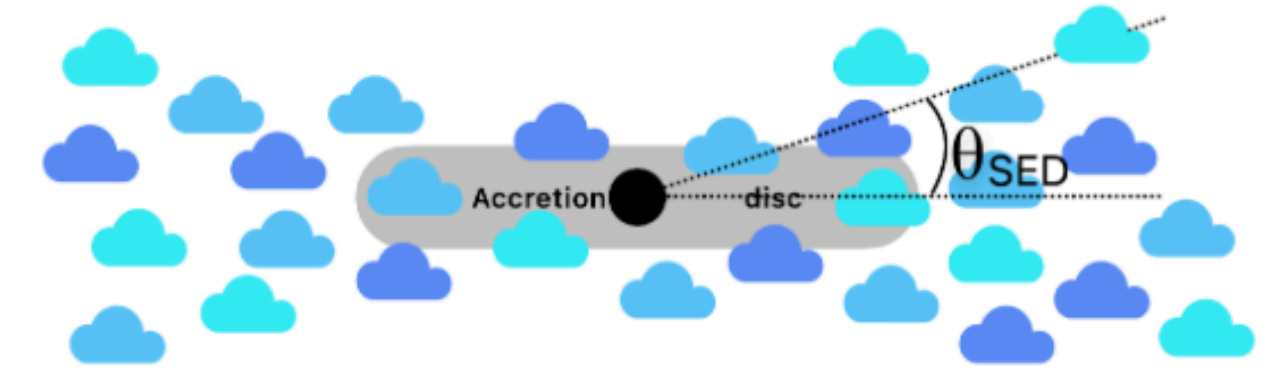
- M_{bh}
- F_{edd}
- optional parameters (spin, photon index, corona radius...)

BLR parameters :

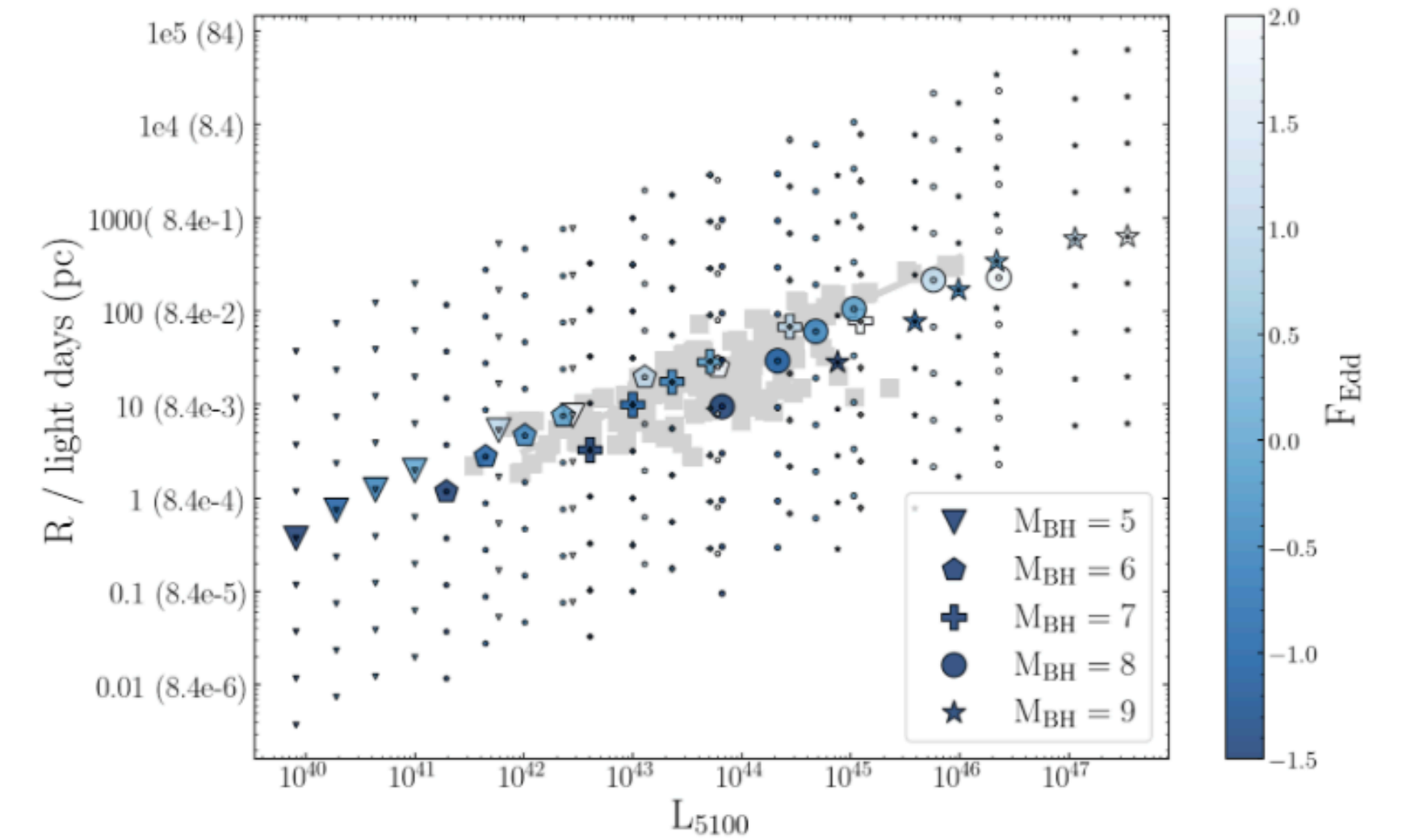
- covering factor (simply a factor for nebular continuum and line)
- Z
- N/O, C/O (or fixed to fiducial N/O ($12 + \log OH$) and $C/O = C/O_{sun}$)
- N_H (fiducial $\log N_H = 21 \text{ cm}^{-2}$)
- turbulent velocity (fiducial $v = 0 \text{ km/s}$)
- n_H distribution (fiducial from $\log n_H = 8$ to 14 cm^{-3} , equally weighted)
- BLR radius distribution (fiducial from $\log \Phi = 16$ to $24 \text{ photons/s/cm}^2$, equally weighted)



Face on view



Edge on view



Units and components of the models :

Each BLR cloud is computed in unit :

- Line flux in $\text{erg/s/cm}^2 \Rightarrow$ multiply by $4\pi \text{ rad}^2$ with $\text{rad} = \sqrt{10^{(\log Q_{agn} - \log \Phi) / (4\pi)}}$ to get erg/s
- Continuum flux in $\text{erg/s/cm}^2 \Rightarrow$ multiply by $10^{(\log Q_{agn} - \log \Phi)}$ to get erg/s

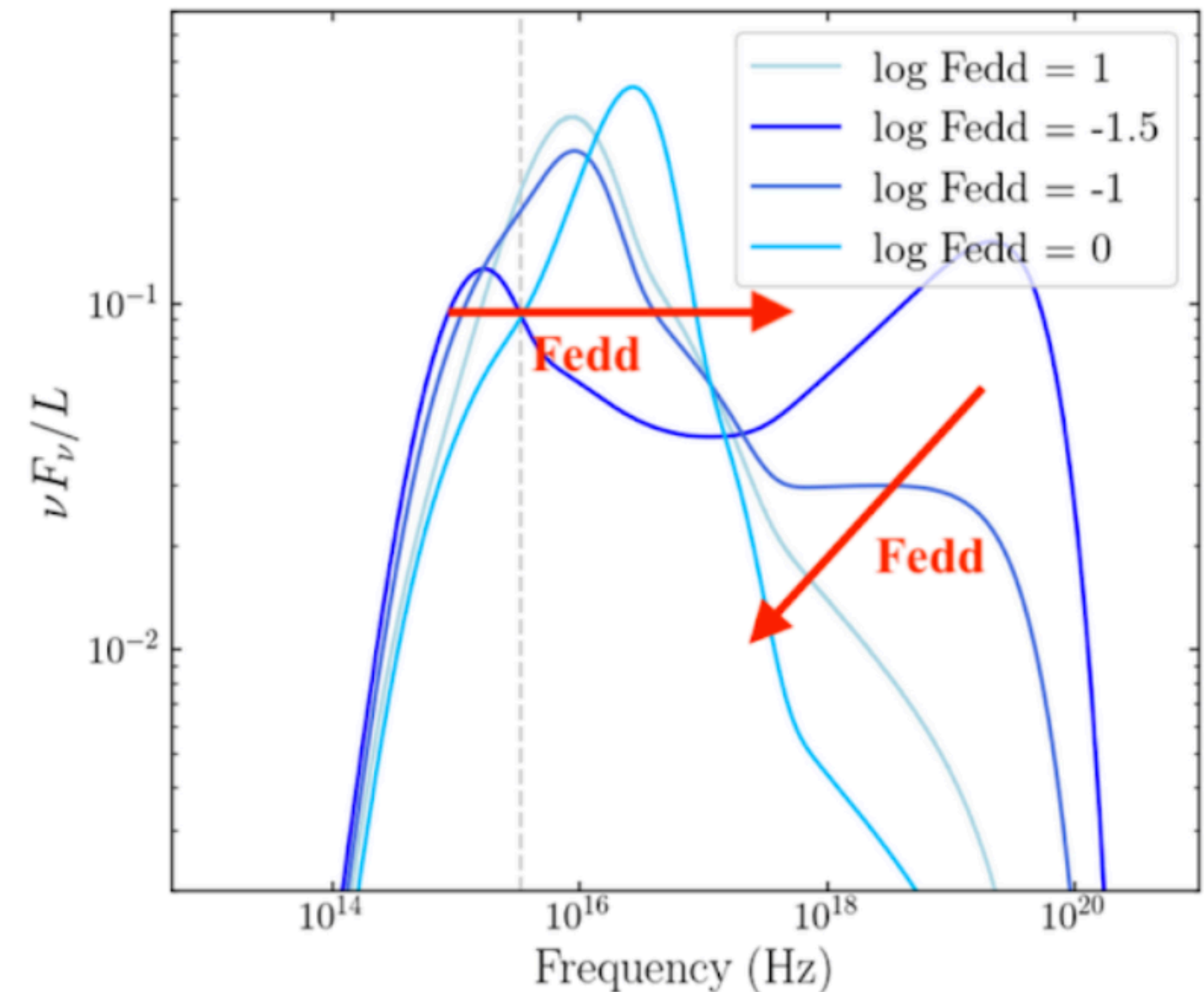
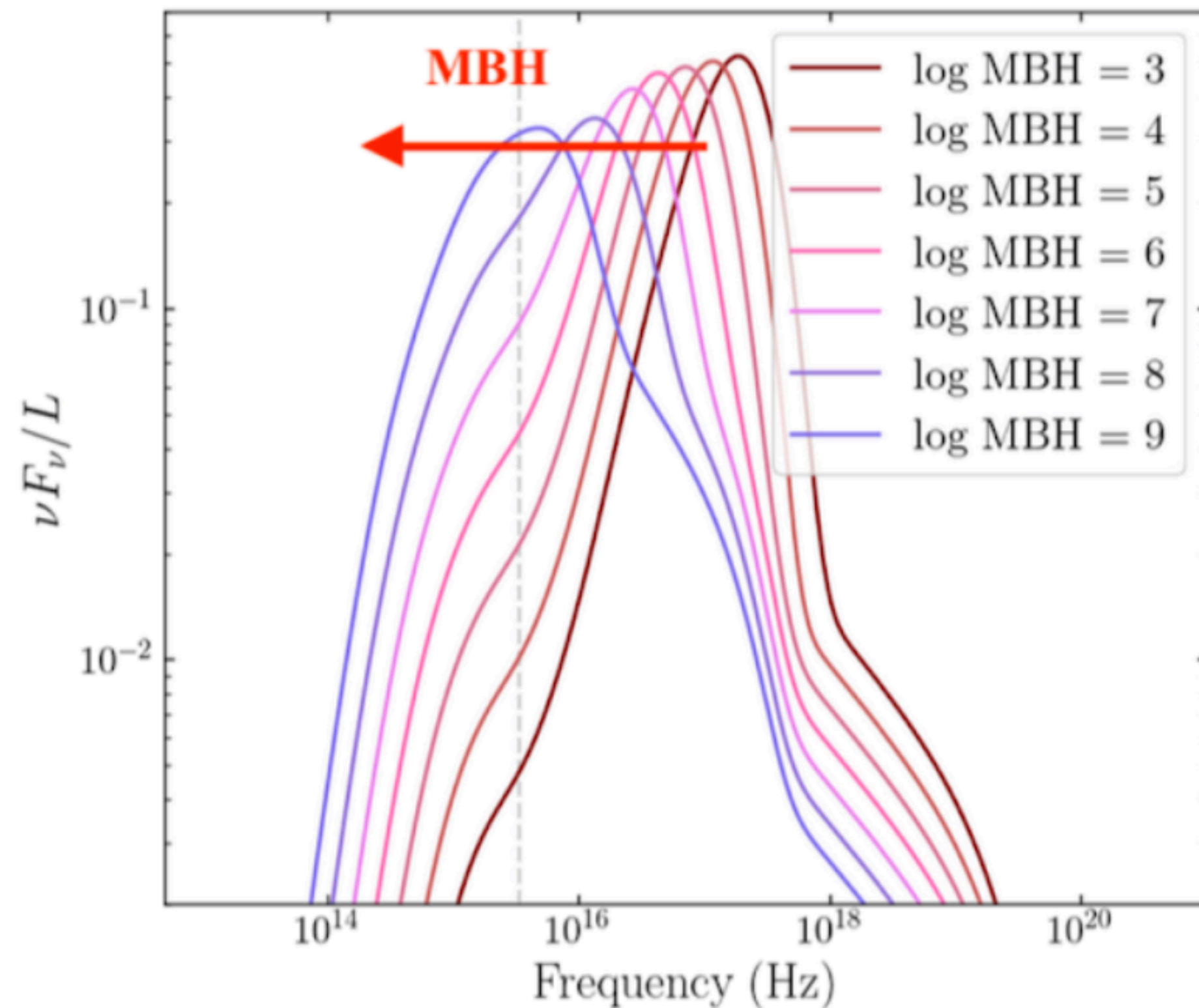
To properly combine the continuum components with the covering factor of the BLR, we need to save the incident continuum, attenuated continuum and nebular continuum.

→ Photoionization models for Type I and Type II AGN with dependence on M_{bh} and L/L_{edd}

Kubota & Done SED :

Novikov & Thorne 1973 (disc emissivities) :
the temperature of the disc depends on
 $F_{\text{edd}} / M_{\text{BH}}$

"qsosed": $L_{\text{diss,hot}} = 0.02 L_{\text{Edd}}$
→ defines r_{hot} and Γ_{hot}



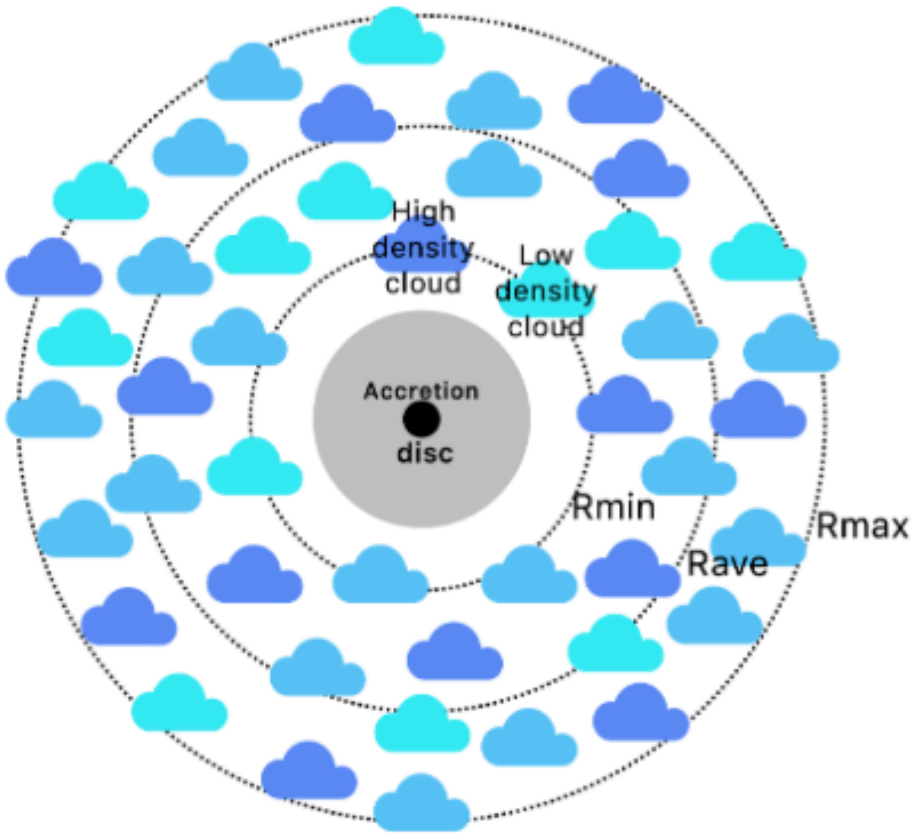
BLR parametrization :

SED parameters :

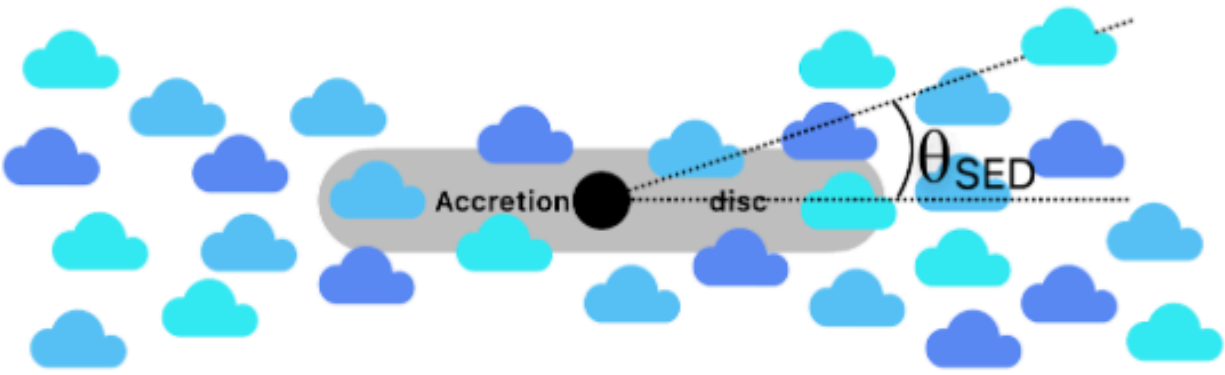
- M_{bh}
- F_{edd}
- optional parameters (spin, photon index, corona radius...)

BLR parameters :

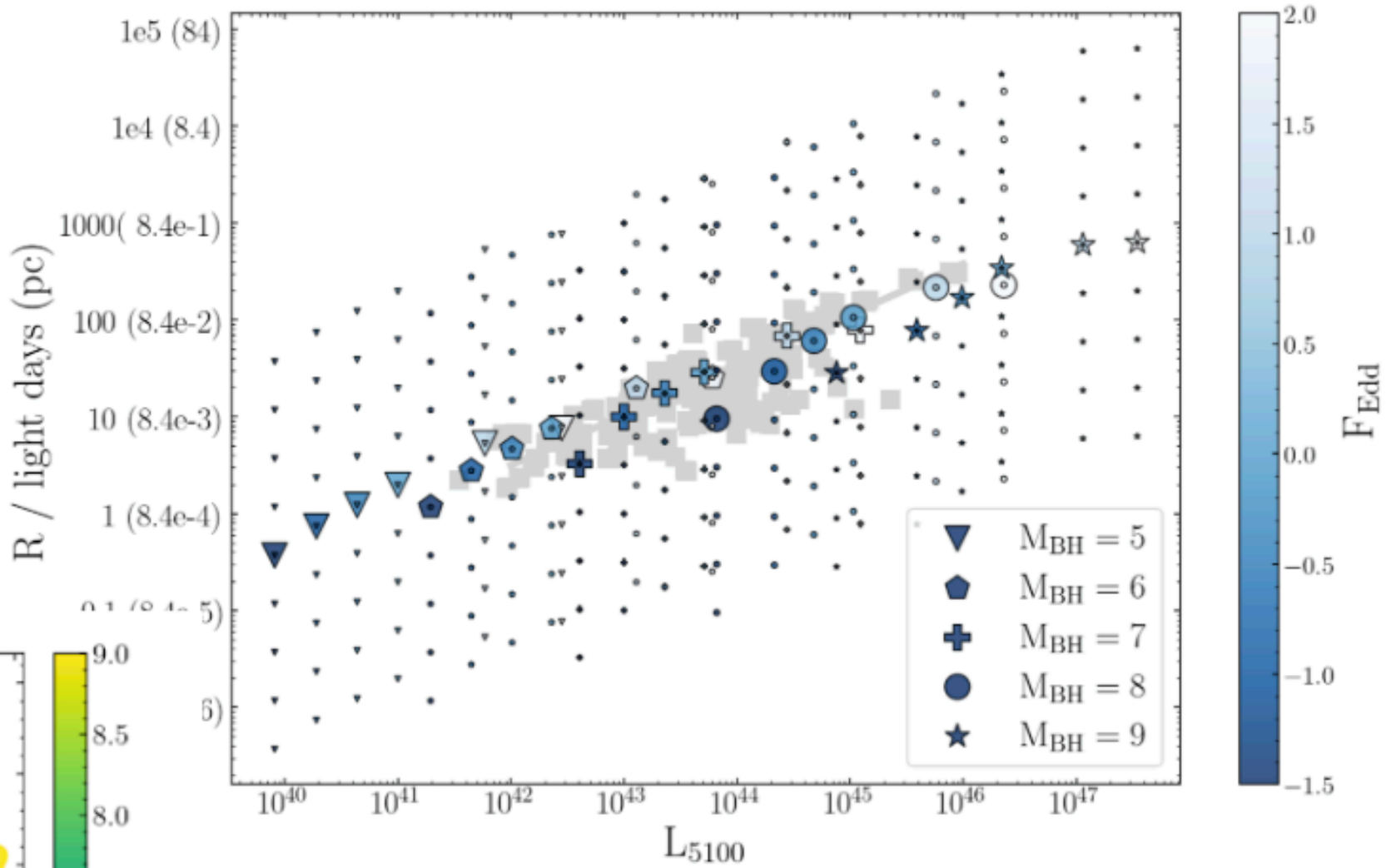
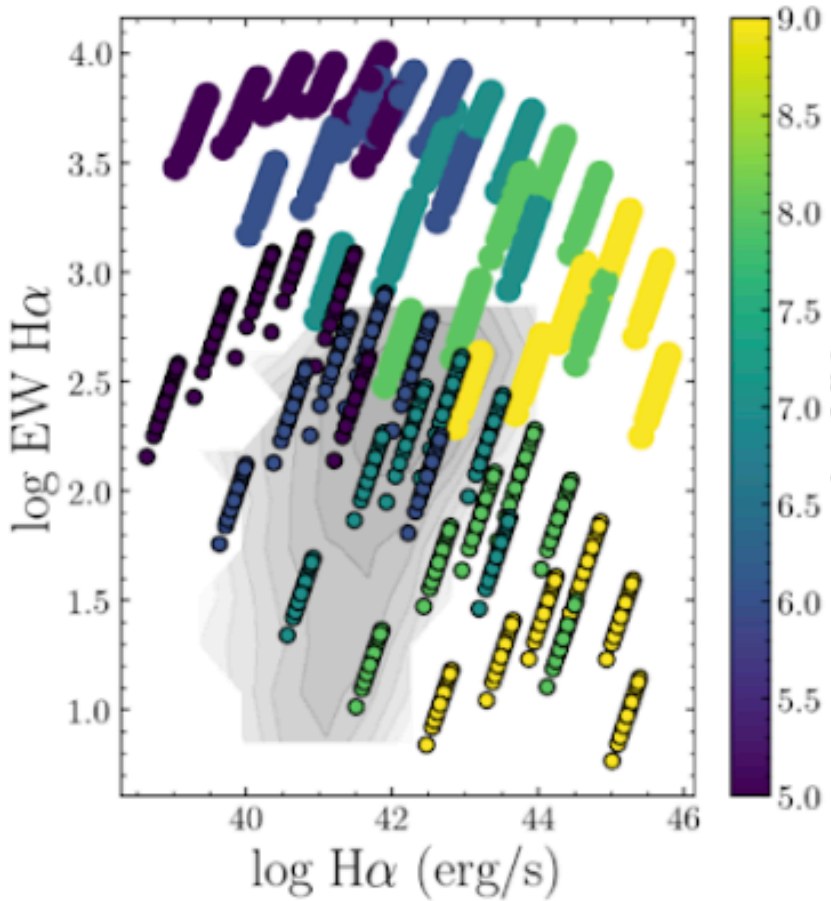
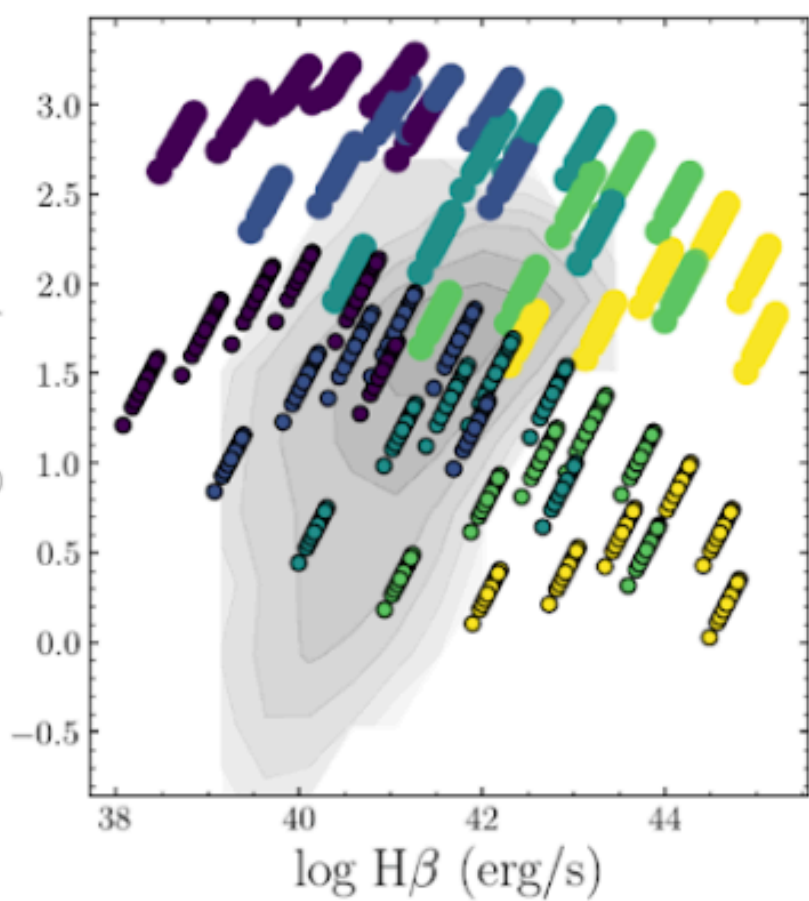
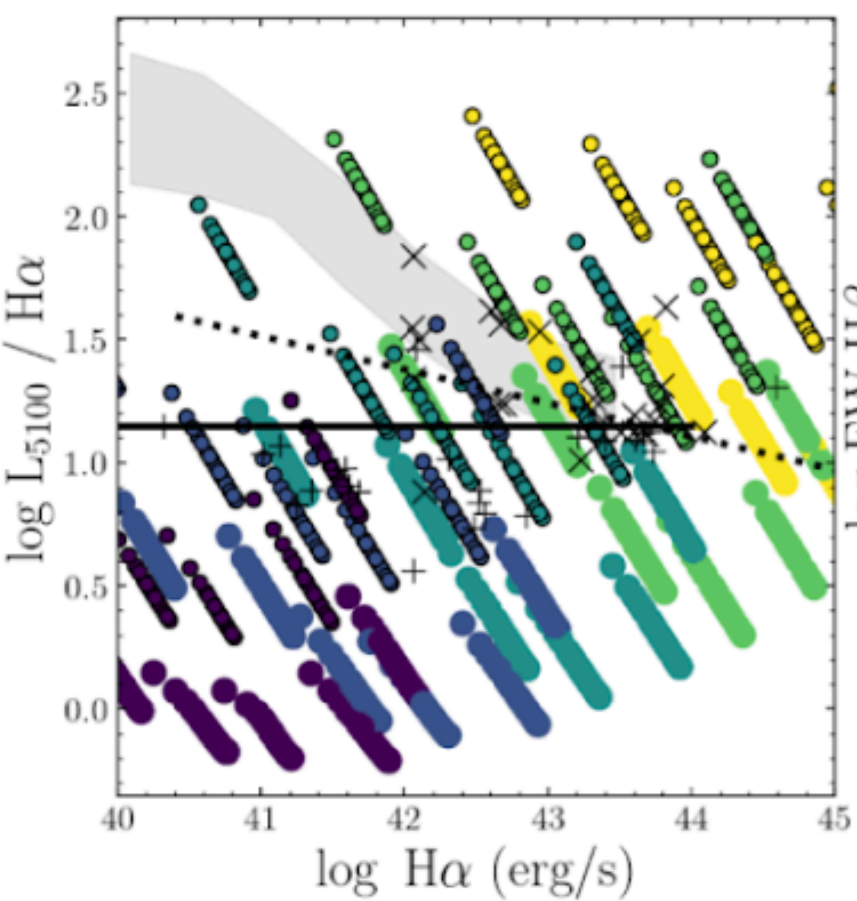
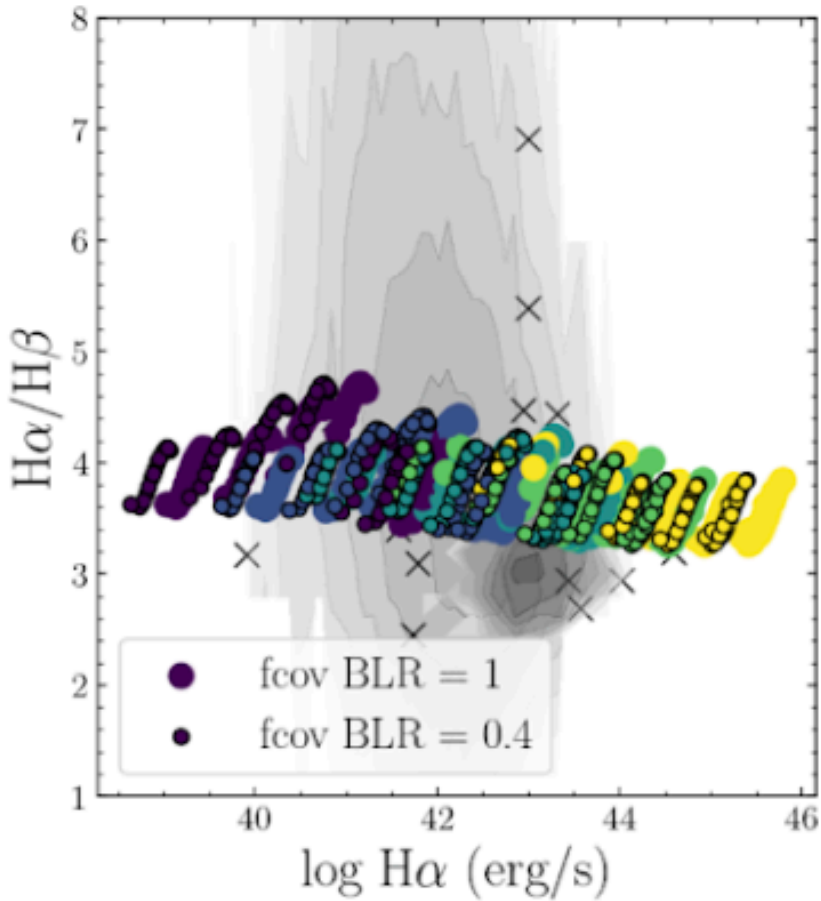
- covering factor (simply a factor for nebular continuum and line)
- Z
- N/O, C/O (or fixed to fiducial N/O ($12+\log OH$) and $C/O=C/O_{sun}$)
- N_h (fiducial $\log N_h = 21 \text{ cm}^{-2}$)
- (micro)-turbulent velocity (fiducial $v = 0 \text{ km/s}$)
- n_H distribution (fiducial from $\log n_h = 8$ to 14 cm^{-3})
- BLR radius distribution (fiducial from $\log \Phi = 16$ to $24 \text{ photons/s/cm}^2$)



Face on view



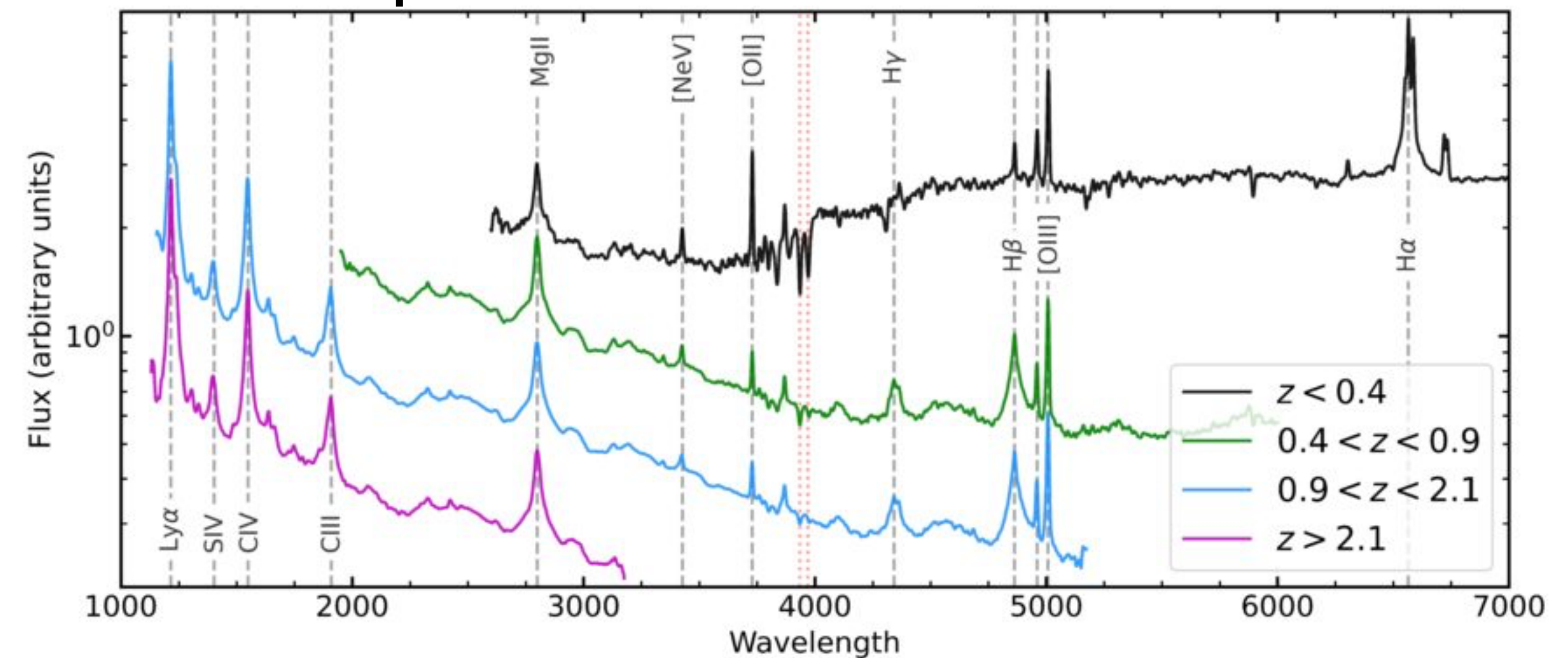
Edge on view



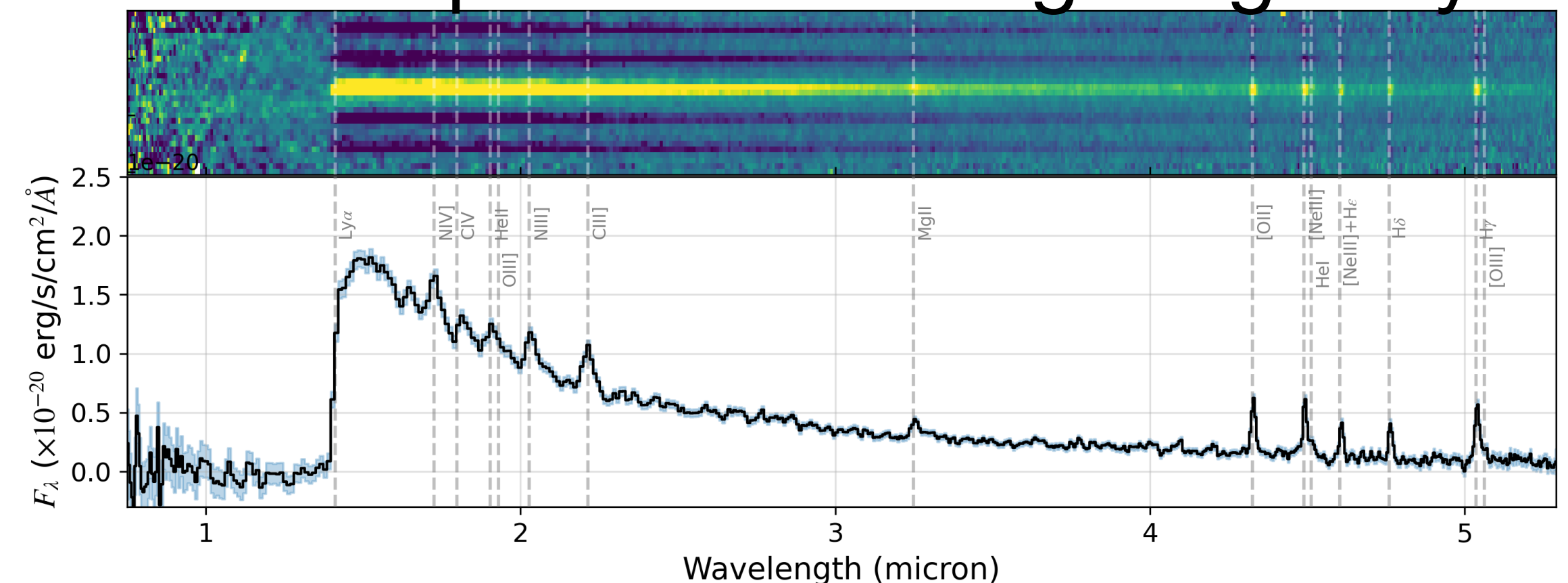
Large samples of observed galaxy spectra over time

- Current and future observatories (DESI, Euclid, 4MOST) collect large amounts of galaxy spectra out to earliest cosmic epochs
- DESI: ~30 million galaxy over its 5-year survey (DR1: 13 million)
- How can we make sense of what we see, i.e. how can we infer redshift and physical galaxy properties from spectra?

DESI Spectra



JWST Spectrum of a high-z galaxy

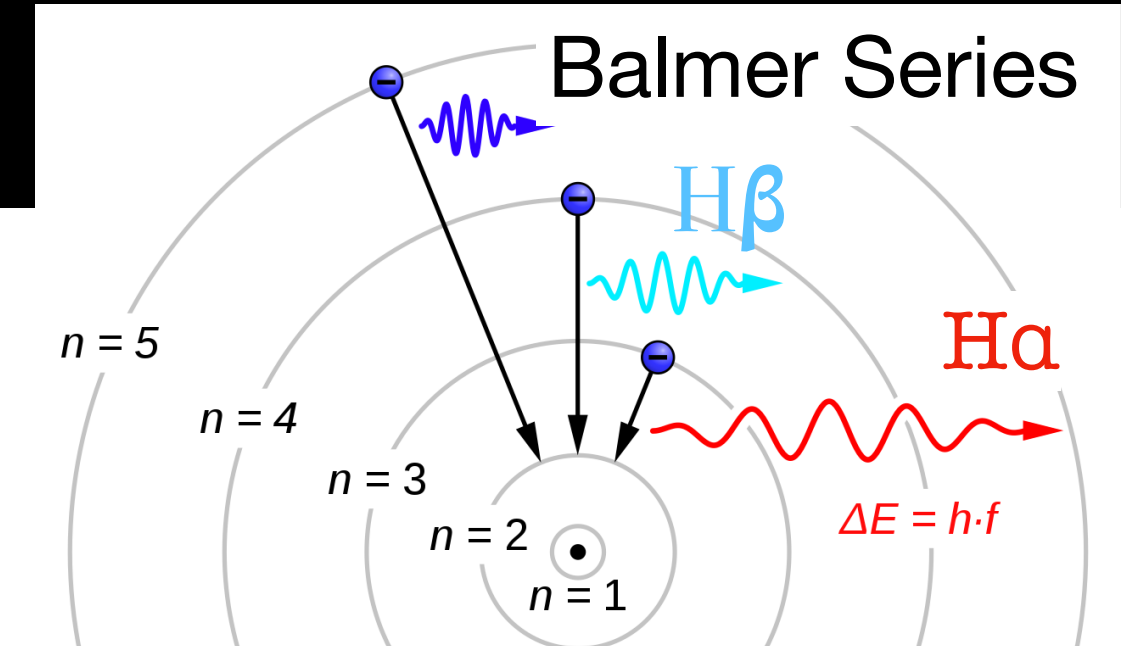
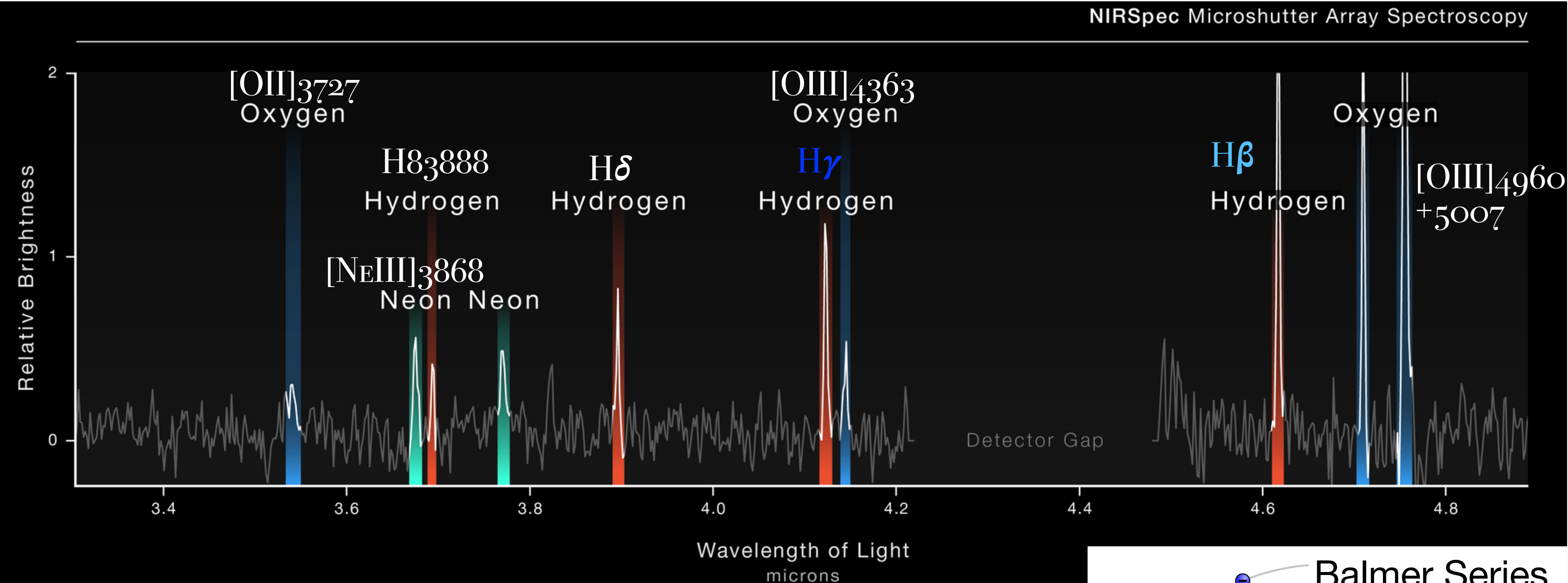


Components of spectra: emission lines

NIR image



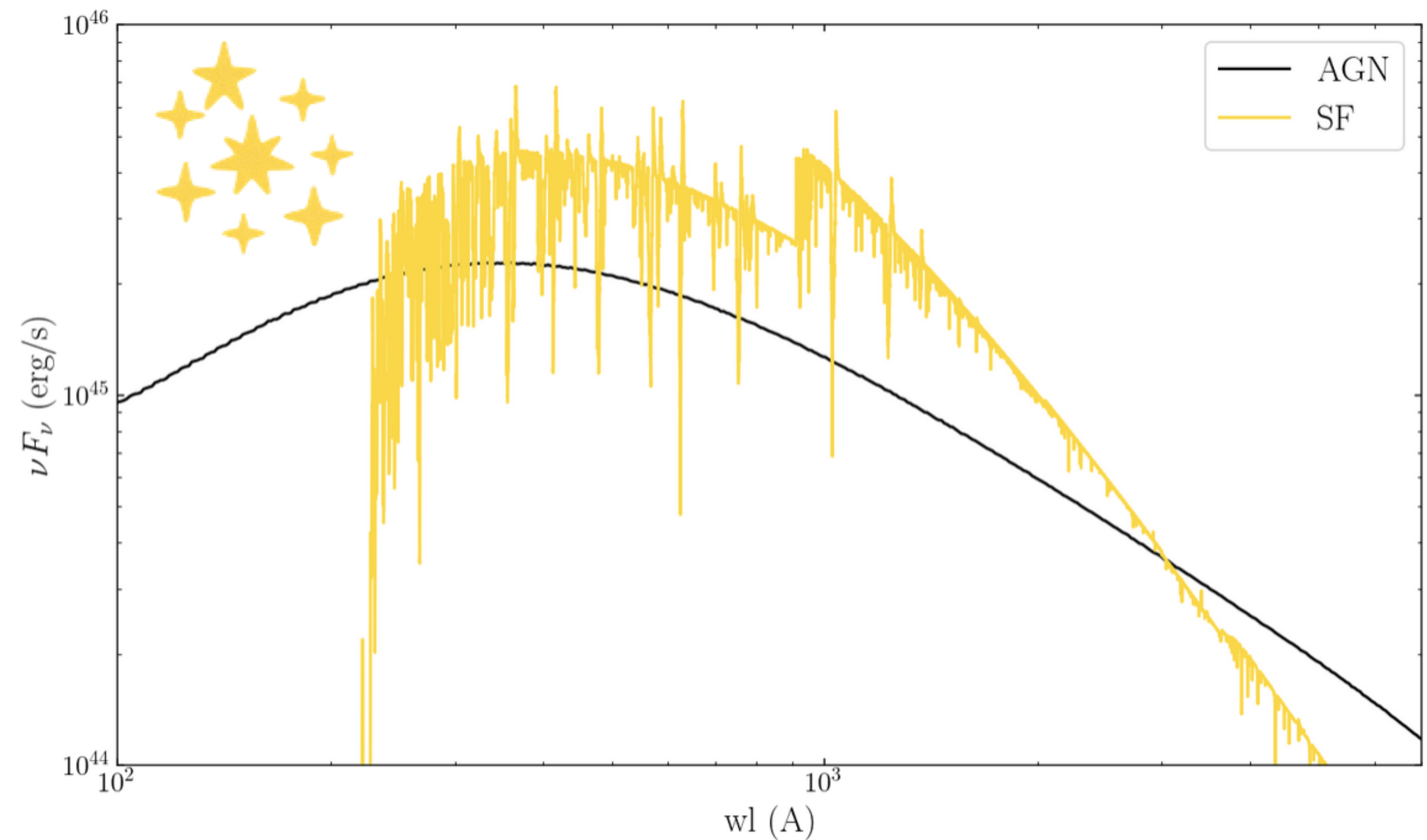
700 million years after BB



- On top of continuum: Optical and UV emission lines emerging from warm gas ionised by UV radiation of accreting black holes (BHs) and young star clusters
- **HII regions:** information on SFR, gas metallicity, gas density, gas temperature, dust, type of ionising source
- **Ionised regions around BHs:** information on BH mass, gas accretion rate, gas metallicity, gas density, gas temperature, type of ionising source

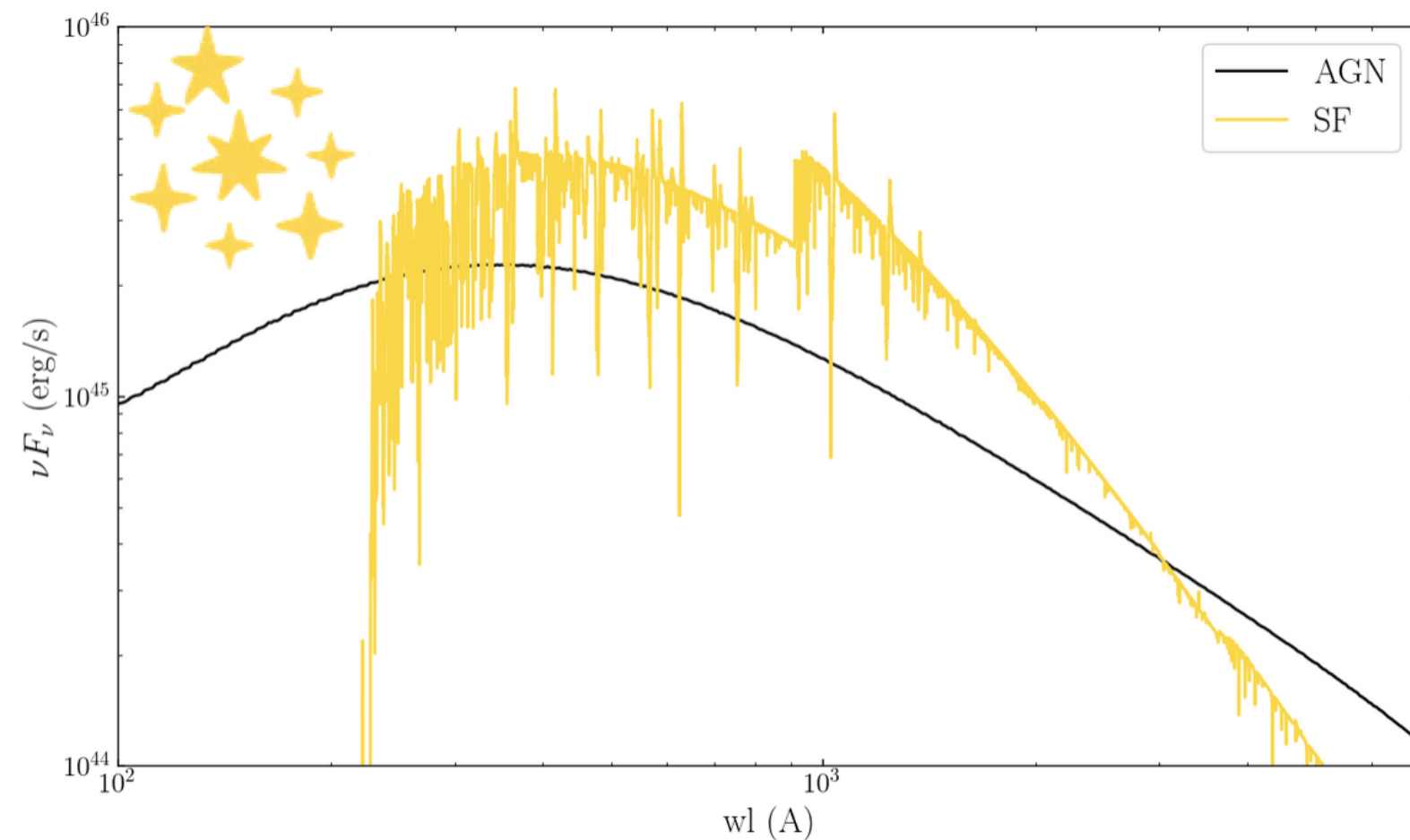
Components of spectra: stellar continuum

- **Stellar continuum**: superposition of spectra of individual stars (mainly optical, UV and NIR)
- Stellar population synthesis models (e.g. Bruzual and Charlot 2003):
 - combine initial mass function with evolution models of stars and convolve that with SFR and chemical enrichment history
- Information on stellar mass, SFR, mean stellar Z and mean age

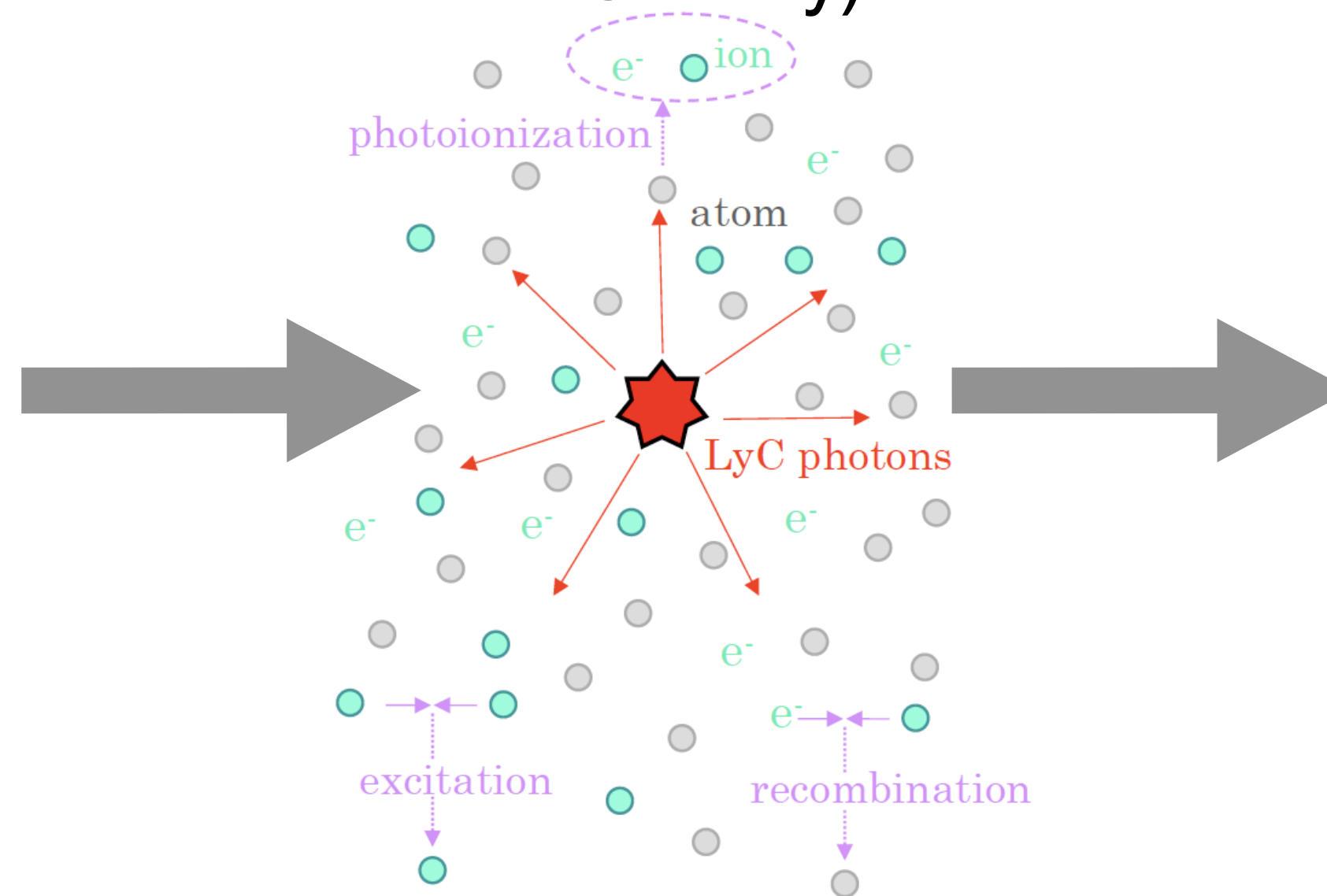


Model galaxy spectra with emission lines

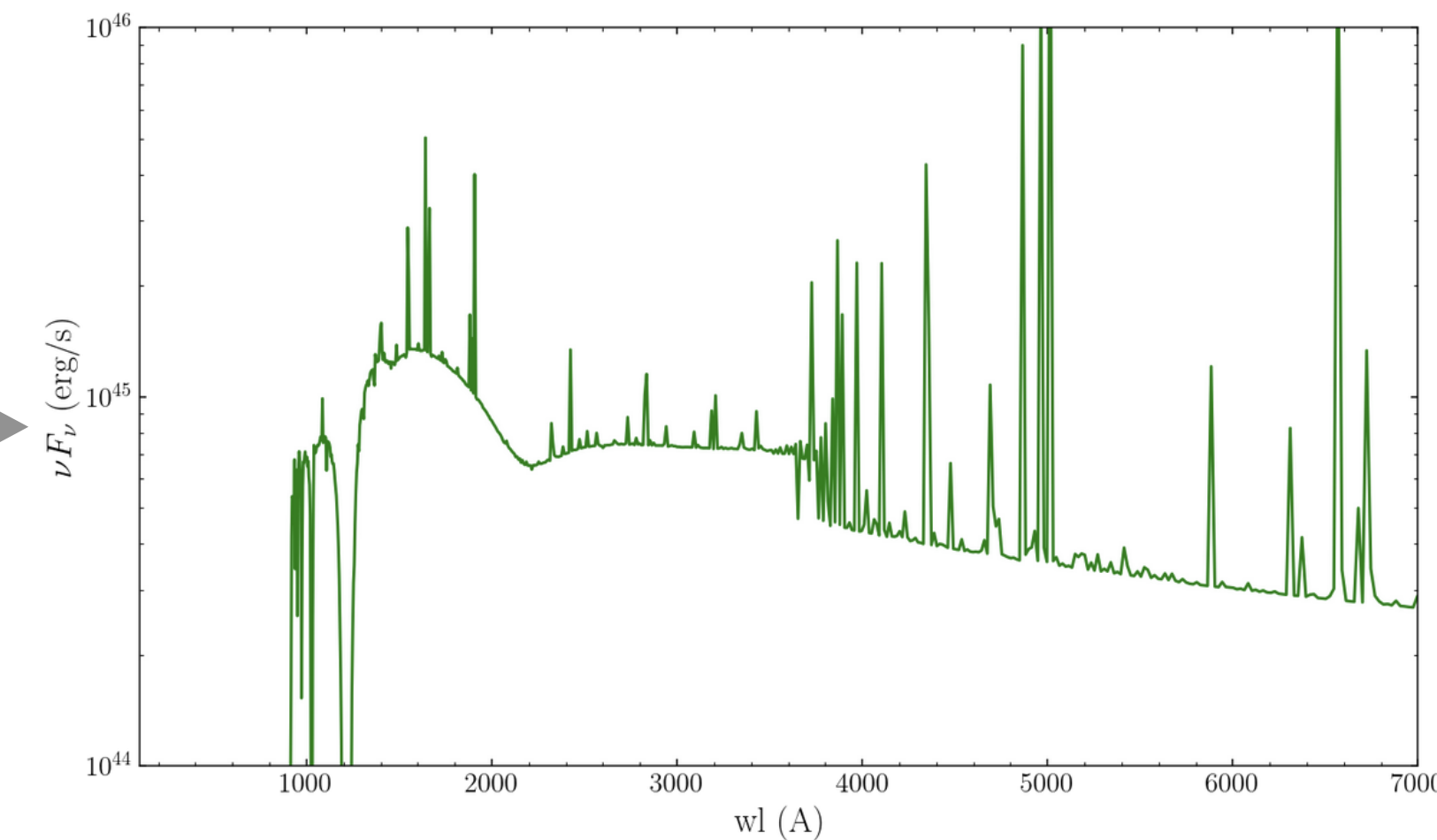
Stellar continuum from SSP models



Transmission through gas cloud with photo-ionisation code (e.g., Cloudy)



Transmitted spectrum with emission lines from HII regions

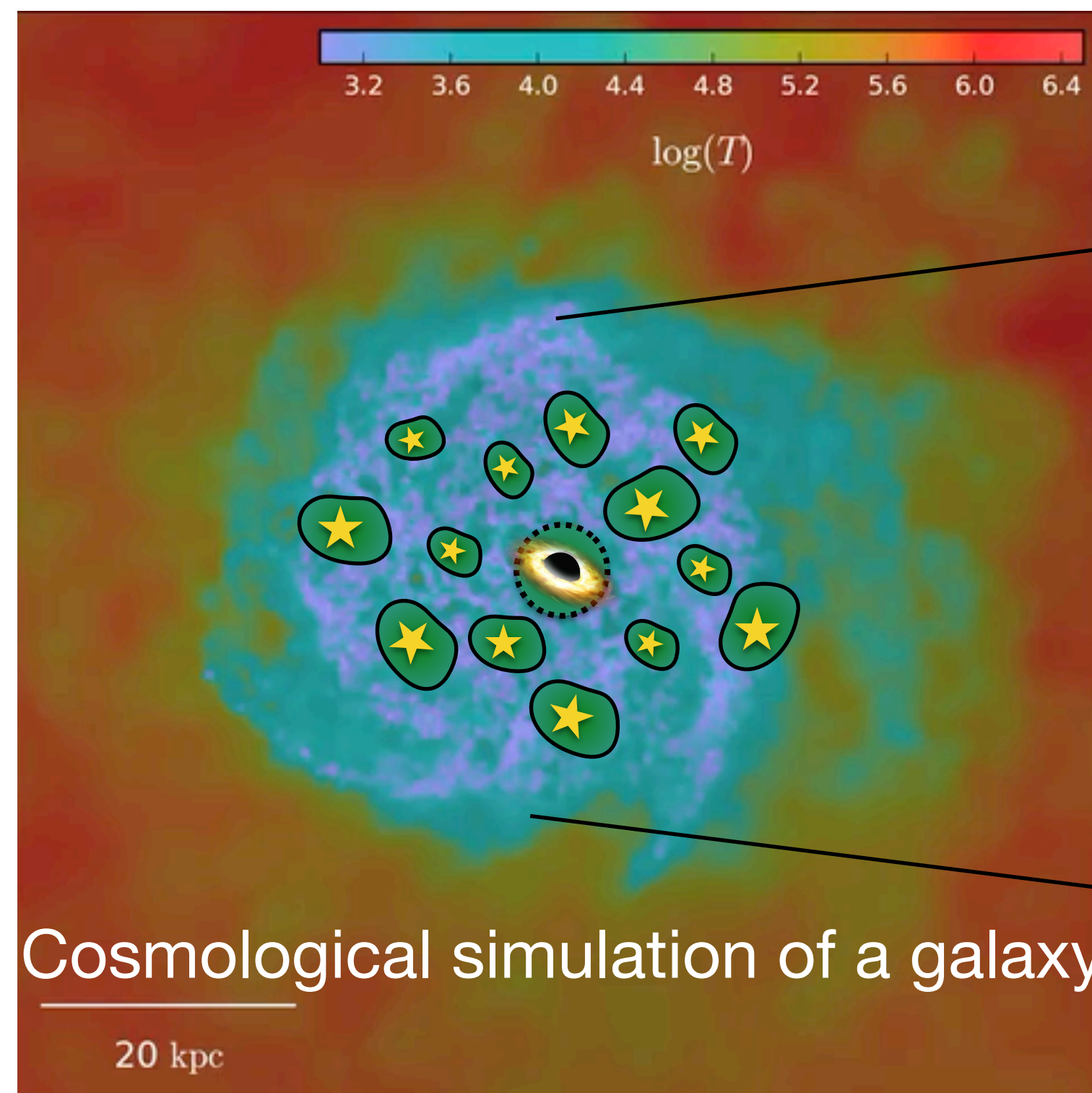


e.g., Gutkin+17, Plat+19

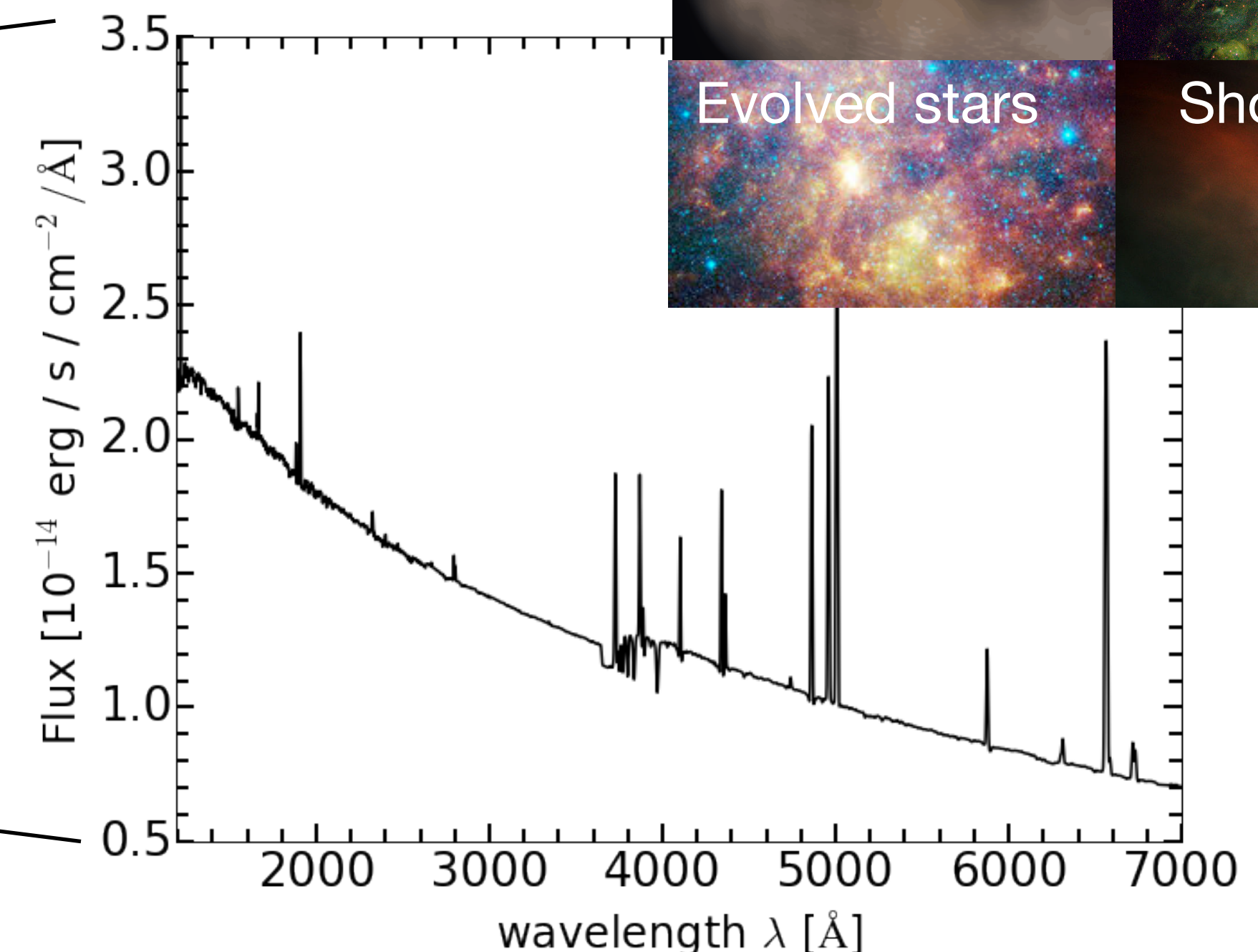
- Modelling nebular continuum and lines is complex, using basic atomic physics for ionisation and excitation of atoms and ions.
- Large model grid with free parameters being: SFR, gas density, gas metallicity, ionisation parameter, dust-to-metal-mass ratio, element abundances such as C/O or N/O, LyC escape fraction etc.

Estimate physical properties from observed spectra

- **Alternative:** Couple model spectra with emission lines to galaxies from cosmological simulations (line-emitting regions unresolved) by matching free parameters
- Reduces free parameters space to physically motivated combinations at given cosmic time
- **Our method:** only one following line emission from accretion BHs, evolved stars and fast radiative shocks



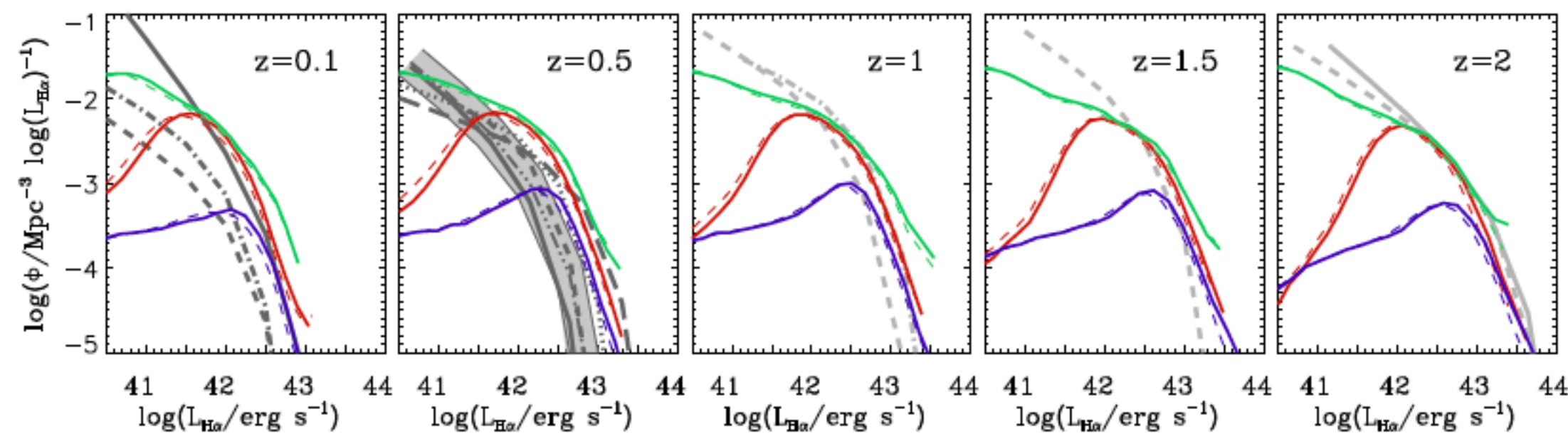
MH+17/19/23a/b/in prep;
Ceverino, MH+21; Scharre,
MH+24



Emission line properties of simulated galaxies

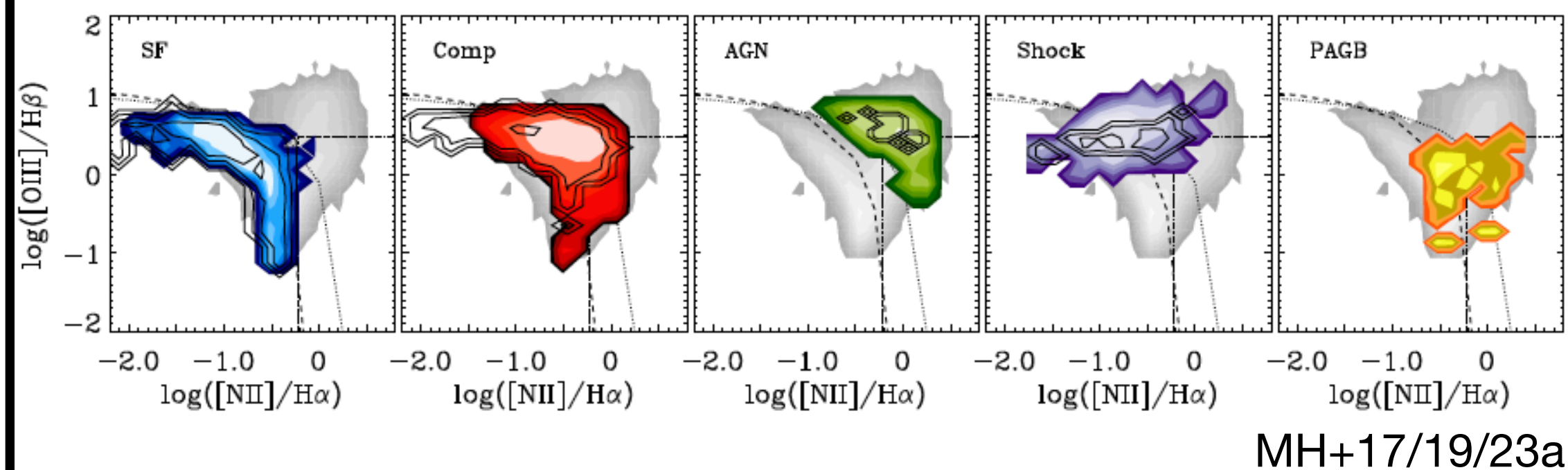
Predicted emission-line properties validated against observations of nearby galaxies

Number counts of emission-line galaxies



MH+23a, Scharre, MH+24;
see also Zhai+19, Izquierdo-Villalba, Shen+20

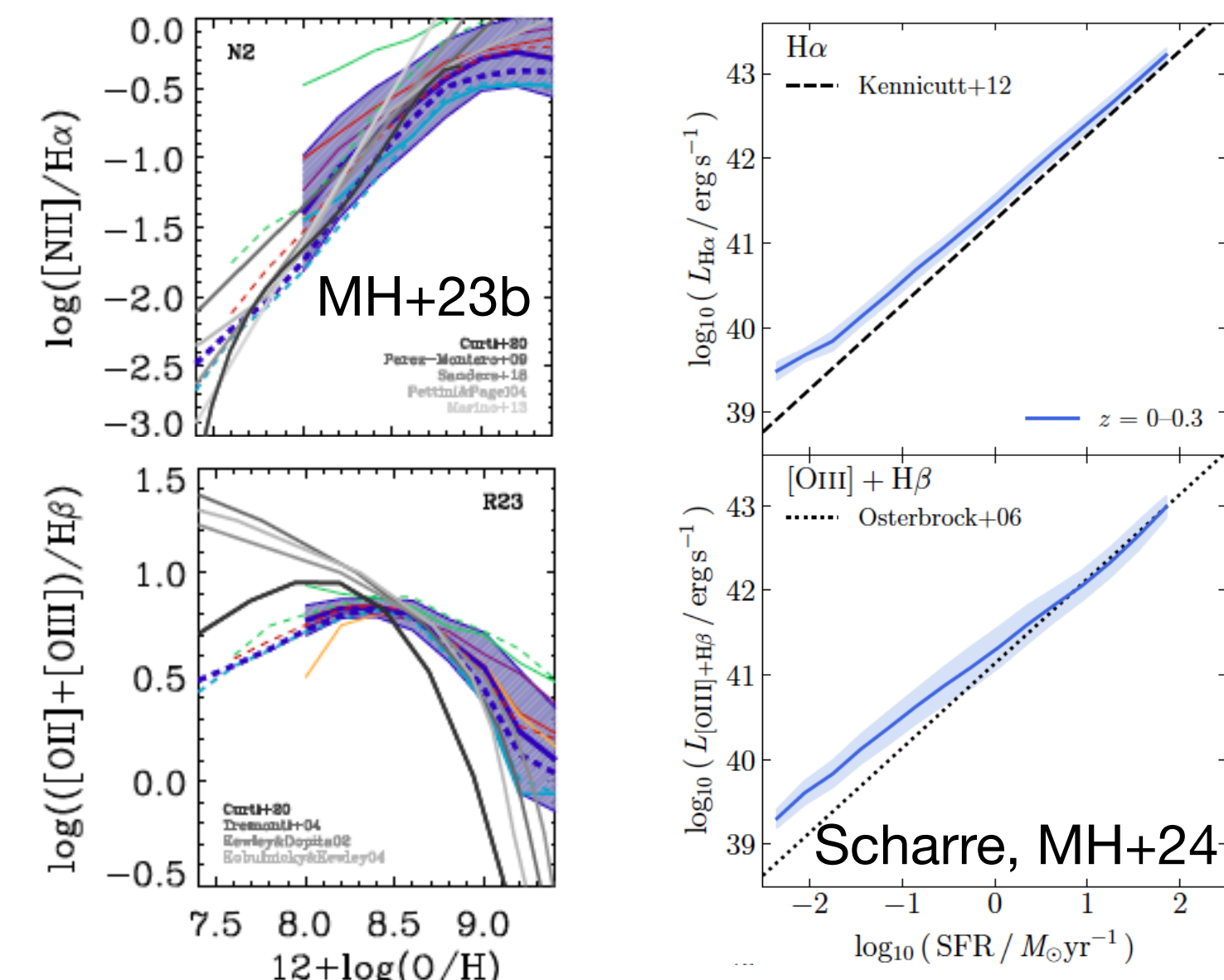
BPT diagrams



MH+17/19/23a

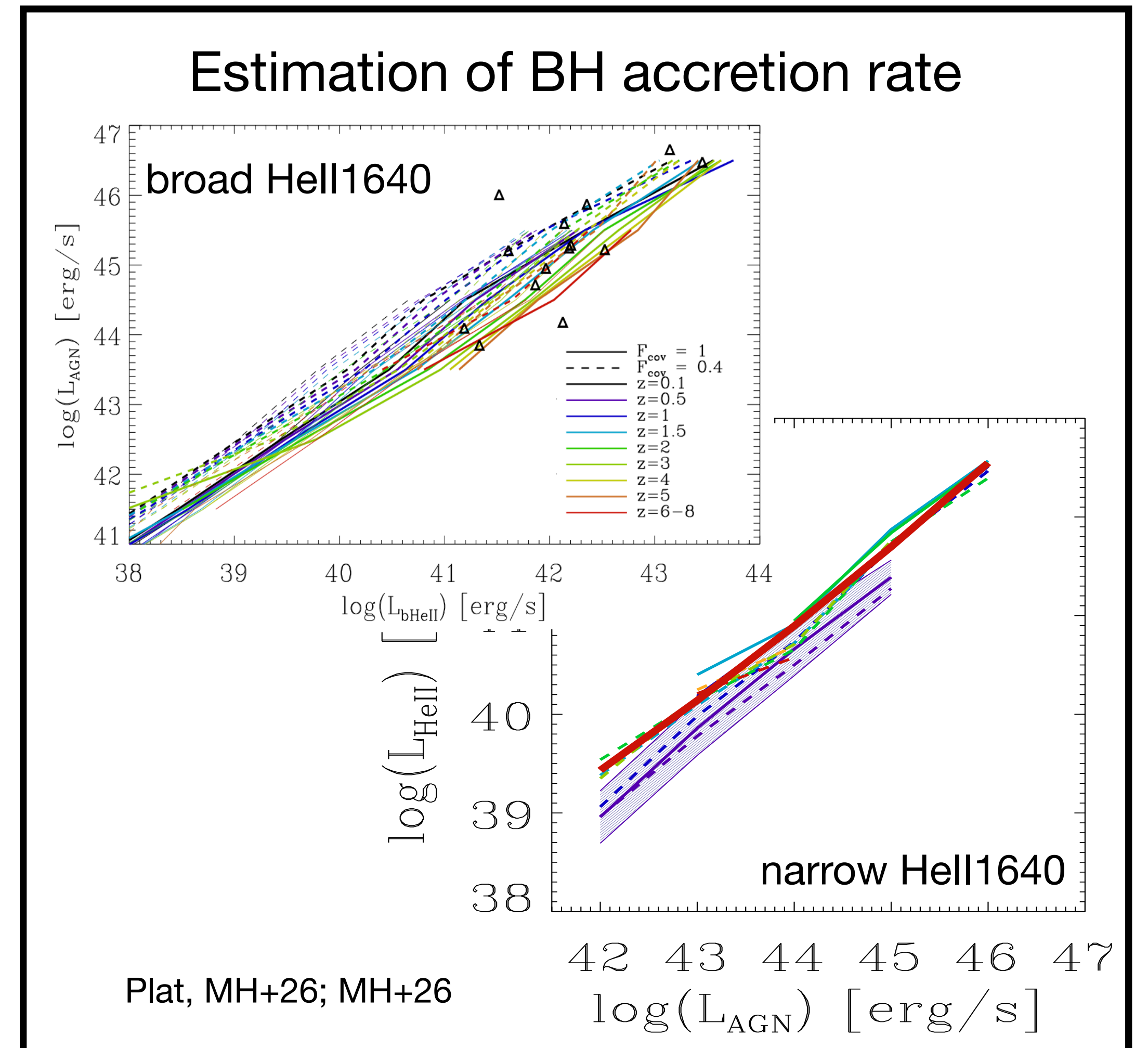
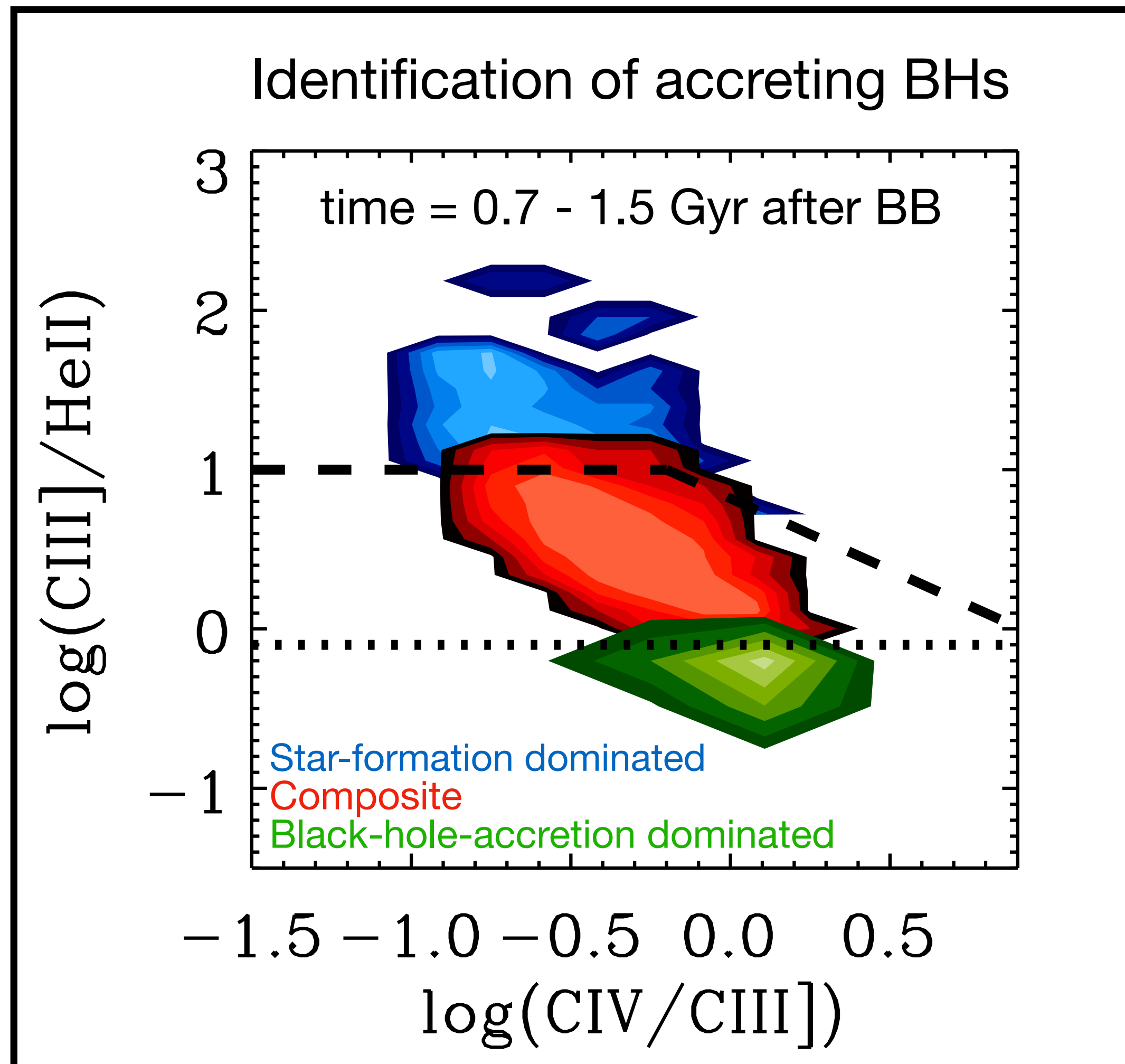
- Method applied to different cosmological simulations (e.g., MISTRAL, IllustrisTNG, Colibre) and models (e.g., GAEA & SC semi-analytic model)
- ⇒ To date, largest emission-line catalogues of 100.000s of simulated galaxies over cosmic time

Relations between line-ratios and SFR, metallicity



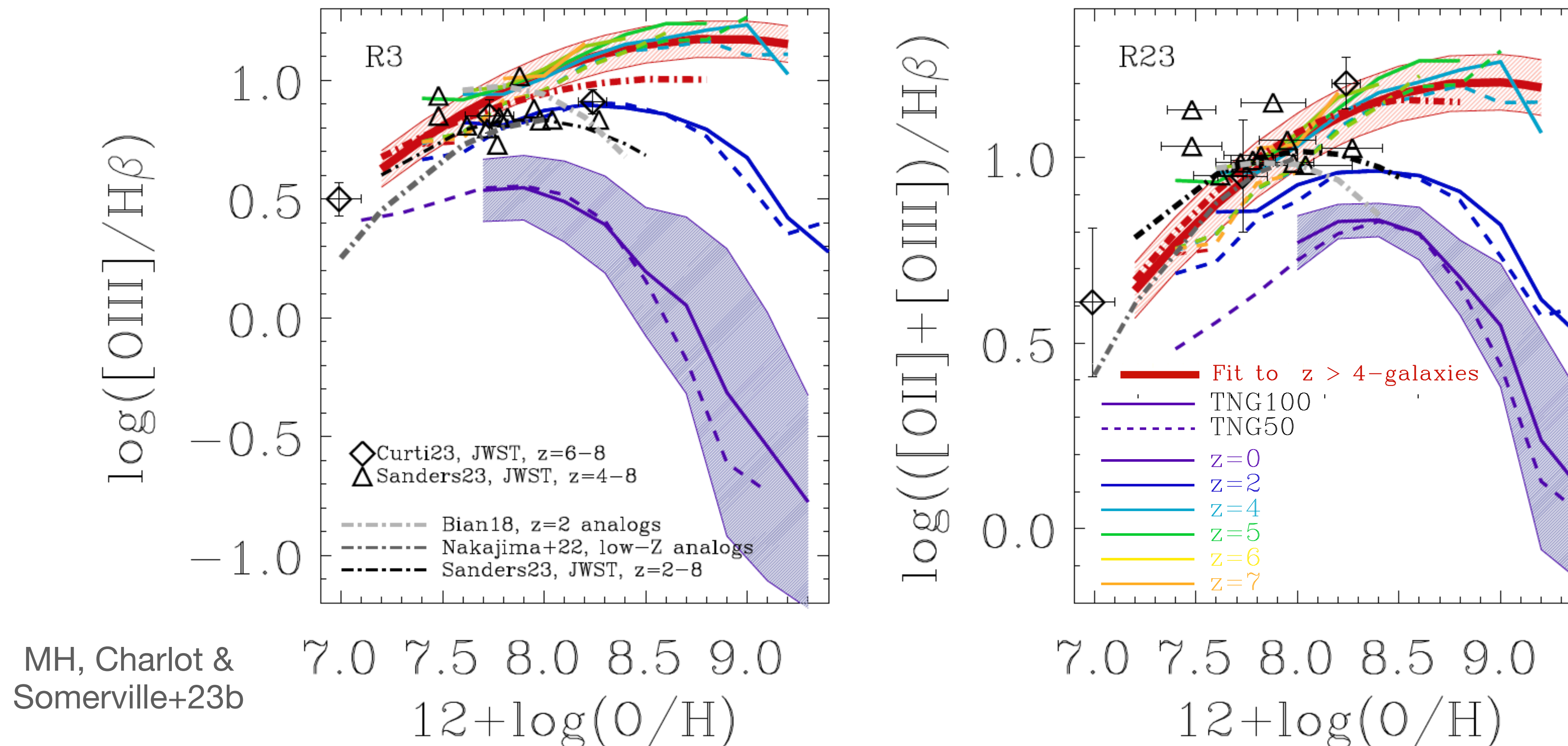
Interpreting high-z galaxy spectra from JWST

Detectability & identification of accreting black holes and estimation of their properties



Interpreting high-z galaxy spectra from JWST

Identification of metallicity diagnostics for high-z galaxy spectra to estimate gas metallicities



- Line-ratio-Z relations predicted to evolve with redshift consistent with first direct-T estimates from JWST (Curti+23, Sanders+23, Laseter+23)
- Red lines:** Predicted metallicity calibrations for galaxies at $z > 4$, can be used by observers