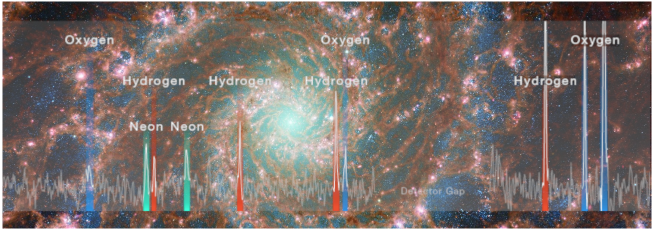
Multiline models of galaxies

Overview of modeling strategies using nearby galaxies as references

Vianney Lebouteiller (CNRS/AIM @CEA Saclay), prepared with Lise Ramambason (Univ. of Heidelberg)



△ Disclaimers

- Wide topic (like others) potentially related to many astrophysical questions
  - Despite my efforts, there will be some bias toward dwarf galaxies, low-metallicity, infrared spectroscopy, Bayesian statistics, and `Cloudy` models. . . !
- Not about ISM models per se particularly adapted to a given physical object/process within galaxies (e.g., PDR, molecular cloud etc. . . ; see presentations by B. Godard and P. Lesaffre) but about how to model multiple galactic components and processes
  - Mostly about models to match variety of lines & processes in large-scale/integrated observations
- Not about simulations, but about galaxy models that can be compared to individual, specific galaxies (and as a result also samples of galaxies)

△ Disclaimers

- Goal: provide some background and recent advances in order to
  - Be aware of biases and existing strategies/codes before choosing the modelling approach
  - Have a critical thinking of the results when reading papers or interpreting own results

- Some references/techniques presented here, but far from exhaustive

Outline

- General considerations and motivating questions
  - What are we trying to model?
  - Why are nearby galaxies useful
  - What physical processes to consider
  - What constraints
- Common diagnostics
- Modelling full galaxies, accounting for complexity of ISM and sources
  - Using 1D models
  - (BREAK?)
  - Evidence of mixing/smearing issues
  - Using >1D models & n x 1D models
  - Some results and ongoing works
- Statistical framework
  - Samplers
  - Model comparisons / decision tree
- Concluding remarks

Outline

- General considerations and motivating questions
  - What are we trying to model?

Galaxies are complex objects!

- Various kinds (dwarfs, spirals, Seyfert, mergers etc. . .)
  - What the object is when we observe it: result of integrated history of star-formation, active nuclear phases, interactions, gas exchange with CGM and IGM. . .
  - What we see: snapshot, often limited number of tracers due to incomplete wavelength coverage, extinction, signal-to-noise etc. . .

What is the object we are trying to model?

Not always easy to define a galaxy as a well-identified/circumscribed object, we limit ourselves to the object as it appears in some specific tracers or to sub-components $\Rightarrow$ Hard guess from unresolved observations

Fig.: *M82 galactic outflow observed with HST, Spitzer & Chandra.*

Fig.: *Extended UV & radio disk of M83 observed with GALEX & VLA.*

Physical processes act on various spatial scales

Observing in different tracers (e.g., Hα, CO, HI...) illustrate the complex ISM structure. Not always easy to link regions with a given excitation source

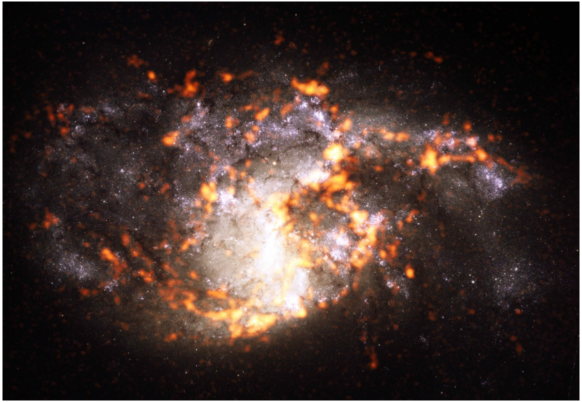


***Fig.**: NGC1385 with ALMA and HST (PHANGS; NRAO).*

Limited information

Distant galaxies are difficult to resolve (e.g., even with JWST), galaxy spectra are often spatially- and spectrally-unresolved

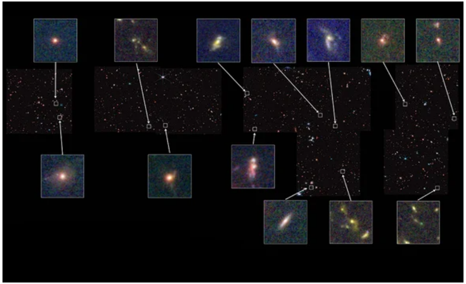
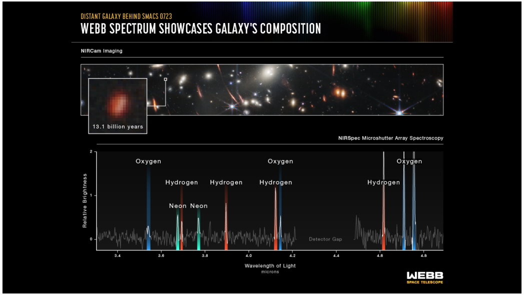


***Fig.**: JWST/NIRcam CEERS field. Credit: NASA/STScI/CEERS/TACC/S. Finkelstein/M. Bagley/Z. Levay; NASA/STScI/CEERS/TACC/S. Finkelstein/M. Bagley/J. Kartaltepe.*



Outline

- General considerations and motivating questions
  - What are we trying to model?
  - Why are nearby galaxies useful

Why study nearby galaxies?

External galaxies in general

- Galaxy evolution (SFR, Z, M*, AGNs. . .) *(e.g., Kewley+ 2019)*
  - Mass-metallicity-SFR relation *(e.g., Nakajima+ 2023)*
  - Looking for metal-free gas in the reionization epoch *(e.g., Vanzella+ 2023)*
  - (see presentations by D. Dale, K. Sandström, B. Groves)
- Multiline modeling also tackles the specific role of ISM in galaxy evolution (e.g., SF)
  - e.g., role of H_2 in SF, tracers of H_2. . . *(e.g., Madden+ 2020)*
  - Cosmic evolution of the ISM as an astrophysical object like stars and compact objects

Nearby galaxies

- Great opportunity to understand extragalactic ISM in ≠ environments and to design relevant models
- Some galaxies nearby enough to spatially disentangle physical components
  - Modelling sum of individual regions vs. full galaxy? Do results change with spatial scale considered?

Why study nearby galaxies?

External galaxies in general

- Galaxy evolution (SFR, Z, M*, AGNs. . .) *(e.g., Kewley+ 2019)*
  - Mass-metallicity-SFR relation *(e.g., Nakajima+ 2023)*
  - Looking for metal-free gas in the reionization epoch *(e.g., Vanzella+ 2023)*
  - (see presentations by D. Dale, K. Sandström, B. Groves)
- Multiline modeling also tackles the specific role of ISM in galaxy evolution (e.g., SF)
  - e.g., role of H_2 in SF, tracers of H_2. . . *(e.g., Madden+ 2020)*
  - Cosmic evolution of the ISM as an astrophysical object like stars and compact objects

Nearby galaxies

- Great opportunity to understand extragalactic ISM in $\neq$ environments and to design relevant models
- Some galaxies nearby enough to spatially disentangle physical components
  - Modelling sum of individual regions vs. full galaxy? Do results change with spatial scale considered?

Why study nearby galaxies: tracers!

- Some galaxies nearby enough to detect many tracers arising from different phases/physical processes
  - Observed tracers are signatures that reflect the complexity of the galaxy
  - Inversely: the complexity of the model physical representation of the galaxy needs to reflect these signatures (and hopefully those we don't observe)
  - What useful information from limited amount of tracers, do results change with choice of tracers? Should we consider simple models (despite unrealistic) when signatures are available?

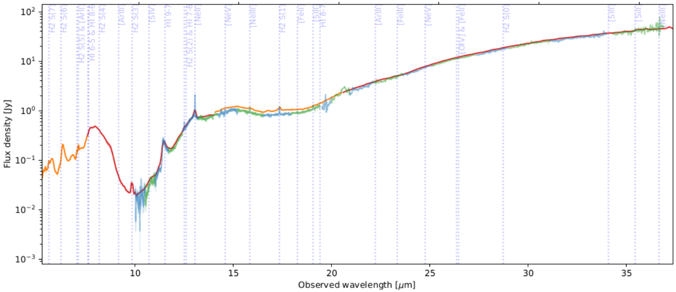


Fig.: *Arp 220 with Spitzer/IRS.*

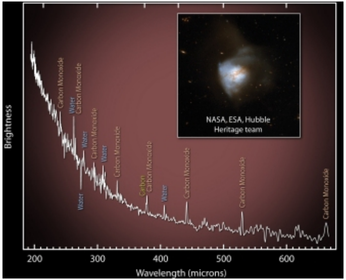


Fig.: *Arp 220 with Herschel/SPIRE (ESA).*

Why study nearby galaxies: we have interesting neighbors

- Some nearby galaxies probe quite different environments compared to MW (Z, SSCs, AGNs. . .)
  - Need some specific prescriptions (e.g., abundance patterns, D/G. . .)

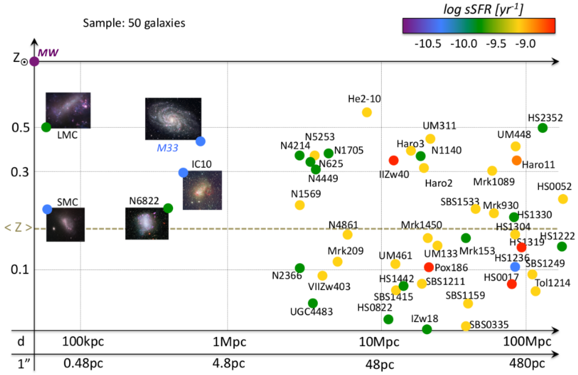


Fig.: Metallicity vs. distance for the Dwarf Galaxy Survey (Madden+ 2013; Cormier+ 2019).

Outline

- General considerations and motivating questions
  - What are we trying to model?
  - Why are nearby galaxies useful
  - What physical processes to consider

Physical processes at work

Typical model parameters to distinguish

- Parameters that describe the matter (gas, dust, composition, spatial distribution. . .)
- Parameters that control the excitation of matter (radiative / mechanical energy)
- Parameters that link both (e.g., ionization parameter U = ionizing photon flux / nc = $Q(\mathrm{H})/4nc\pi r^2$)

Variety of radiative and mechanical feedback processes

- Ionization and heating of the various ISM phases (ionized, neutral atomic, neutral molecular)
- Young stars (UV, X-rays), WR, AGNs, X-ray binaries. . .
- Cosmic rays
- Turbulence and shocks
- Magnetic field
- Molecule formation/destruction. . .

Physical processes at work

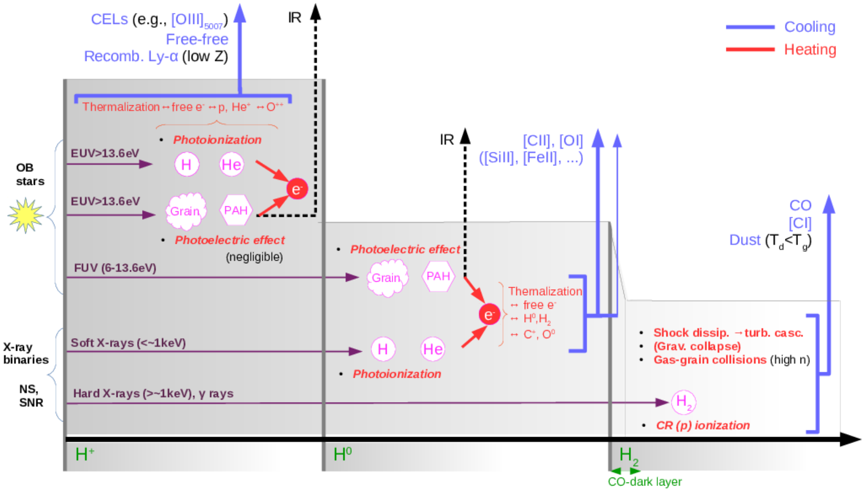
Typical model parameters to distinguish

- Parameters that describe the matter (gas, dust, composition, spatial distribution...)
- Parameters that control the excitation of matter (radiative / mechanical energy)
- Parameters that link both (e.g., ionization parameter U = ionizing photon flux / nc = $Q(\mathrm{H})/4nc\pi r^2$)

Variety of radiative and mechanical feedback processes

- Ionization and heating of the various ISM phases (ionized, neutral atomic, neutral molecular)
- Young stars (UV, X-rays), WR, AGNs, X-ray binaries...
- Cosmic rays
- Turbulence and shocks
- Magnetic field
- Molecule formation/destruction...

Illustration: 1D model



Some approximations

Assuming local conditions and snapshot

- Difficult to do everything right: turbulence, magnetic field, time evolution, chemical network. . .
  - Often relying on specific galaxy regions or galaxies dominated by some process

- Typical timescale problems: disconnection between radiative phases of AGNs or X-ray binaries (state transitions) and observed ISM tracers
  - Non-isotropic emission, light propagation, heating & cooling timescales. . .
  - For simplicity: inferred properties of transient objects reflect conditions seen by the matter when it cools down and not those inferred from the compact object itself

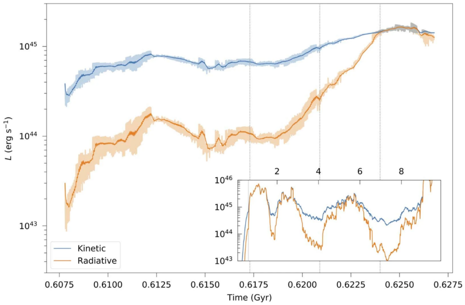


***Fig.**: Evolution of AGN feedback luminosity in zoom-in simulations for a varying SMBH accretion rate (Qiu+ 2020).*

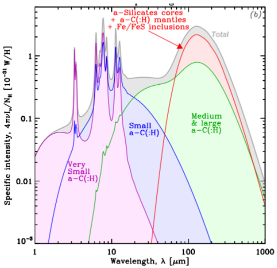
Outline

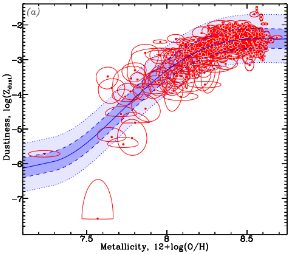
- General considerations and motivating questions
  - What are we trying to model?
  - Why are nearby galaxies useful
  - What physical processes to consider
  - What constraints

What constraints?

Dust

- Dust essential ingredient of models (Presentation by N. Ysard)
- Dust SED holds much information
  - Local physical conditions, $T_{\rm dust}$, $M_{\rm dust}$, $M_{\rm gas}$ through D/G. . . *(e.g., Galliano+ 2021)*
- Difficult to disentangle ISM phases (e.g., those associated or not with SF), especially in integrated galaxy spectra

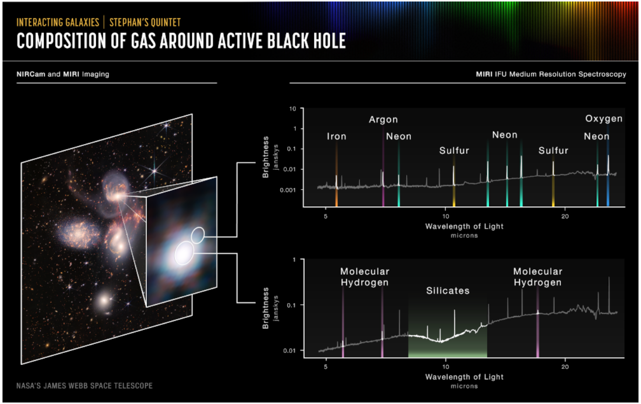




What constraints?

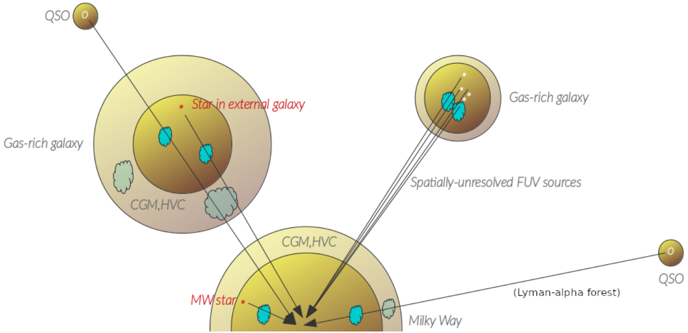
Line spectroscopy

- Gas tracers may constrain:
  - Specific phases: hydrogen (H^+, H^0, H_2), metal ionization (e.g., (OIII)/(OII)...), density (e.g., (SII), (SIII)),
  - Specific excitation mechanisms (X-rays, shocks... e.g., (NeV))
  - Sometimes in a single spectrum



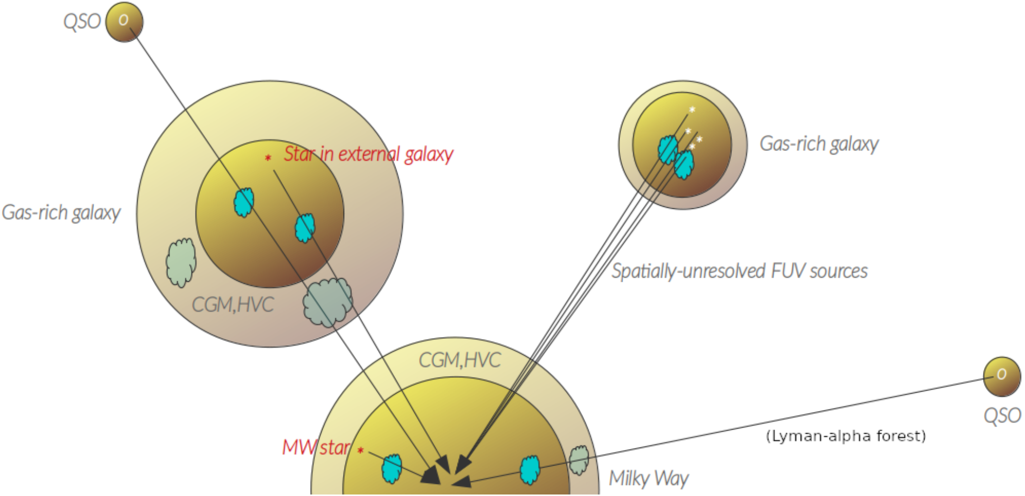
Absorption spectroscopy

- Absorption spectroscopy is very useful for chemical composition, D/H, depletion patterns, cooling rates, molecular gas fraction, even CO-dark H_2 *(Balashev+ 2017, 2020, 2022)*
  - But limited to single line of sight (LOS) or LOS averages (for which there is mixing)
  - Comparison absorption/emission not straightforward *(e.g., Arabsalmani+2023, Wilson+ 2023)*
  - $\approx 5^{+5}_{-0}$ US presidents in the future: HabWorlds Observatory for LOS mapping



- Will focus on emission lines here, it's complicated enough

Absorption spectroscopy



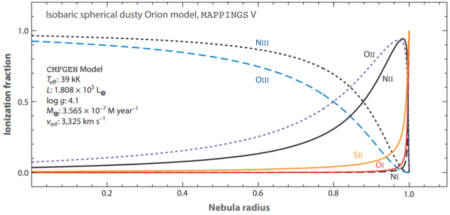
Worth noting. . .

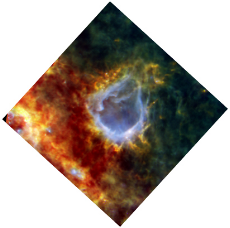
It's obvious but. . . we model what we can see. . .

- Diagnostics are valid only for the regions that are emitting!
- Results may thus be biased by selection effects of emitting components, by extinction
  - As much as possible, such effects need to be accounted for a priori or within models
- What are the possible processes (model ingredients) that can contribute to what we see?
  - We may limit ourselves to assumptions based on current knowledge, but useful to explore a priori unexpected processes as well

. . . and emission arises from different regions. . .

- For instance, the (SII) optical line ratio diagnostic is indicative of density around ionization fronts





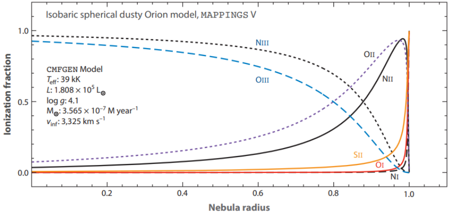
Worth noting. . .

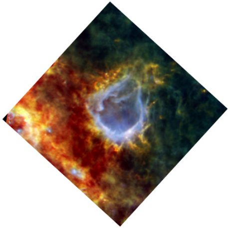
It's obvious but. . . we model what we can see. . .

- Diagnostics are valid only for the regions that are emitting!
- Results may thus be biased by selection effects of emitting components, by extinction
  - As much as possible, such effects need to be accounted for a priori or within models
- What are the possible processes (model ingredients) that can contribute to what we see?
  - We may limit ourselves to assumptions based on current knowledge, but useful to explore a priori unexpected processes as well

. . . and emission arises from different regions. . .

- For instance, the (SII) optical line ratio diagnostic is indicative of density around ionization fronts





Outline

- General considerations and motivating questions
  - What are we trying to model?
  - Why are nearby galaxies useful
  - What physical processes to consider
  - What constraints
- Common diagnostics
- Modelling full galaxies, accounting for complexity of ISM and sources
  - Using 1D models
  - Evidence of mixing/smearing issues
  - Using >1D models & n x 1D models
  - Some results and ongoing works
- Statistical framework
  - Samplers
  - Model comparisons / decision tree
- Concluding remarks

Typical empirical diagnostics using emission lines

Many signatures potentially available to constrain galactic parameters and physical processes

- Primary ingredients of models may be the ultimate goal (e.g., gas density n, metallicity Z)
- Or other physical parameters can be deduced if the "right" processes have been considered
  - either in a relatively trivial way: e.g., SFR, $M(\mathrm{H}^+)$, AGN fraction...
  - or not: f_{esc}, $M(\mathrm{H}_2)$...

Many, many potential diagnostics through spectroscopy of galaxies

- Historically, long-slit or integrated spectroscopy of galaxies used to probe average physical conditions / chemical composition / dominant excitation sources :
  - Gas (electron) density & pressure, ionization parameter from line ratios *(e.g., Kewley+ 2019)*
  - Chemical composition (abundances, metallicity, depletion patterns) *(e.g., Dopita+ 2016)*
  - SFR (e.g., Hα, far-IR lines, 24 μm...) *(e.g., de Looze+ 2017)*
  - BPT (Baldwin–Phillips–Terlevich) & AGN fraction (excitation diagram) vs. mass, vs. z
- Coronal lines indicating unambiguous AGN activity *(e.g., CLASS survey, Reefe+ 2022)*
- UV ($\sim$1400-1900Å) diagnostics to distinguish SF, AGN, and shocks *(e.g., CLASSY survey, Mingozzi+ 2023)*

Typical empirical diagnostics using emission lines

Many signatures potentially available to constrain galactic parameters and physical processes

- Primary ingredients of models may be the ultimate goal (e.g., gas density *n*, metallicity *Z*)
- Or other physical parameters can be deduced if the "right" processes have been considered
  - either in a relatively trivial way: e.g., SFR, $M(\mathrm{H}^+)$, AGN fraction. . .
  - or not: f_{esc}, $M(\mathrm{H}_2)$. . .

Many, many potential diagnostics through spectroscopy of galaxies

- Historically, long-slit or integrated spectroscopy of galaxies used to probe average physical conditions / chemical composition / dominant excitation sources :
  - Gas (electron) density & pressure, ionization parameter from line ratios *(e.g., Kewley+ 2019)*
  - Chemical composition (abundances, metallicity, depletion patterns) *(e.g., Dopita+ 2016)*
  - SFR (e.g., Hα, far-IR lines, 24 μm. . .) *(e.g., de Looze+ 2017)*
  - BPT (Baldwin–Phillips–Terlevich) & AGN fraction (excitation diagram) vs. mass, vs. z
- Coronal lines indicating unambiguous AGN activity *(e.g., CLASS survey, Reefe+ 2022)*
- UV ($\sim$1400-1900Å) diagnostics to distinguish SF, AGN, and shocks *(e.g., CLASSY survey, Mingozzi+ 2023)*

Illustrations: electron/gas density diagnostics

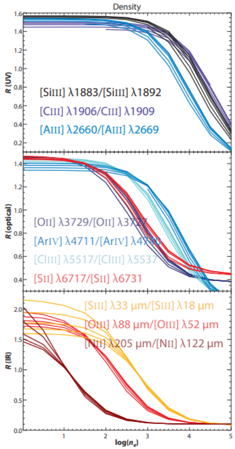


***Fig.**: Density diagnostics for UV, optical, and IR line ratios (Kewley+ 2019).*

Illustrations: excitation mechanisms

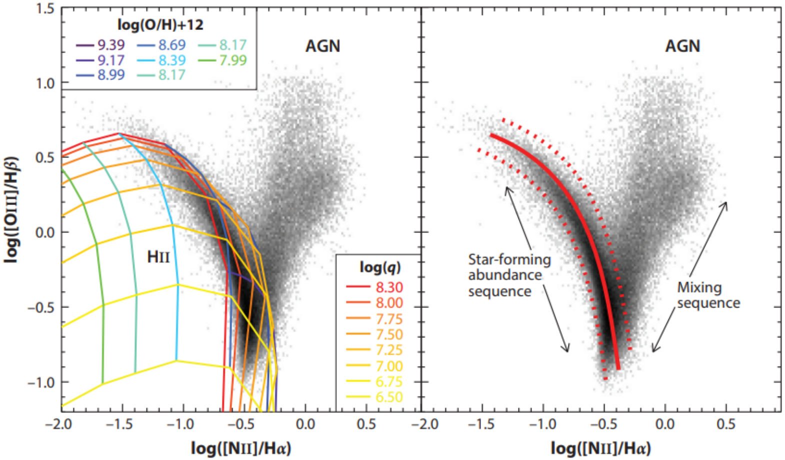


***Fig.**: BPT diagram with optical lines (Kewley+ 2019).*

Illustrations: AGN fraction

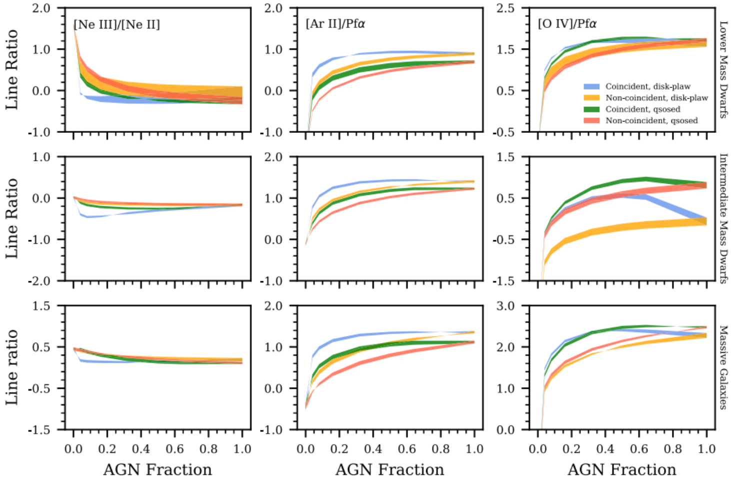


***Fig.**: JWST MIRI line ratios vs. AGN fraction for a given BH mass and for various geometries (Richardson+ 2022).*

Illustrations: AGN activity

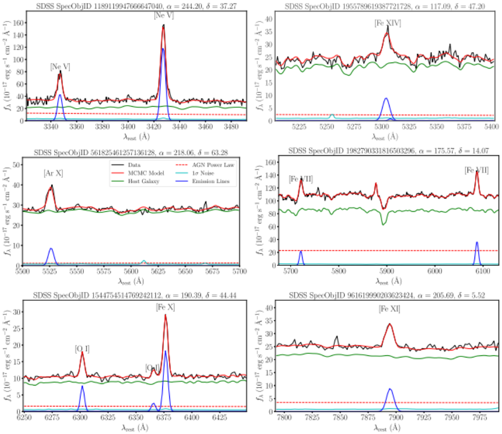


Fig.: Coronal lines in SDSS probe the relatively harder BH spectrum in low-mass galaxies and probe AGN activity where optical diagnostics suggest SF (Reefe+ 2022).

Illustrations: disentangling excitation mechanisms

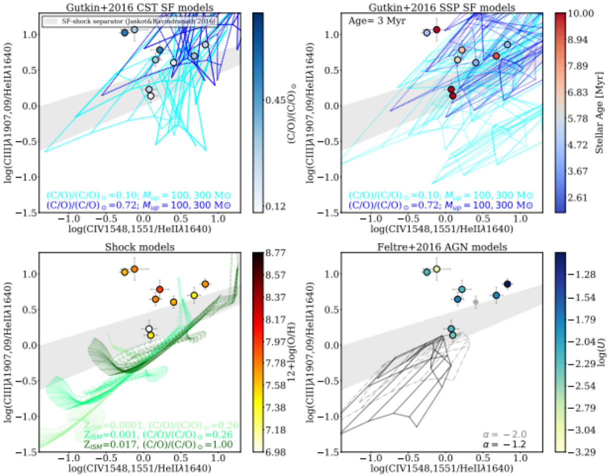


Fig.: *UV diagnostics to distinguish SF, AGN, and shocks (Mingozzi+ 2023).*

Spectroscopic diagnostics have a bright future

Wide-field/all-sky optical and near-IR spectroscopic surveys

- SDSS-V 2020-, Euclid 2023-2030, Rubin 2024-2033, Roman 2026-2032, SPHEREx >2025
- Millions of spatially-resolved and integrated spectra, including mostly dwarf galaxies as well as low-surface brightness galaxies for a wide redshift range

IR spectroscopy

- JWST is observing much fainter IR lines compared to *Spitzer* $\Rightarrow$ new diagnostics
- Mid/far-IR spectra @z=0 mostly not resolved with *Spitzer*, *Herschel*, same for a potential future IR NASA probe-class mission (waiting for IR space interferometry. . .)

High-z spectroscopy

- JWST is providing optical diagnostics at very high-z, spatially unresolved $\rightarrow$ ELT
- Exciting JWST+ALMA synergies, ALMA high-z galaxies already show multi-phase ISM *(e.g., Fujimoto+ 2022)*
- UV diagnostics shift to NIRspec when optical ones shift to MIRI *(CLASSY; Mingozzi+ 2023)*

Spectroscopic diagnostics have a bright future

Wide-field/all-sky optical and near-IR spectroscopic surveys

- SDSS-V 2020-, Euclid 2023-2030, Rubin 2024-2033, Roman 2026-2032, SPHEREx >2025
- Millions of spatially-resolved and integrated spectra, including mostly dwarf galaxies as well as low-surface brightness galaxies for a wide redshift range

IR spectroscopy

- JWST is observing much fainter IR lines compared to *Spitzer* ⇒ new diagnostics
- Mid/far-IR spectra @z=0 mostly not resolved with *Spitzer*, *Herschel*, same for a potential future IR NASA probe-class mission (waiting for IR space interferometry. . .)

High-z spectroscopy

- JWST is providing optical diagnostics at very high-z, spatially unresolved → ELT
- Exciting JWST+ALMA synergies, ALMA high-z galaxies already show multi-phase ISM *(e.g., Fujimoto+ 2022)*
- UV diagnostics shift to NIRspec when optical ones shift to MIRI *(CLASSY; Mingozzi+ 2023)*

Spectroscopic diagnostics have a bright future

Wide-field/all-sky optical and near-IR spectroscopic surveys

- SDSS-V 2020-, Euclid 2023-2030, Rubin 2024-2033, Roman 2026-2032, SPHEREx >2025
- Millions of spatially-resolved and integrated spectra, including mostly dwarf galaxies as well as low-surface brightness galaxies for a wide redshift range

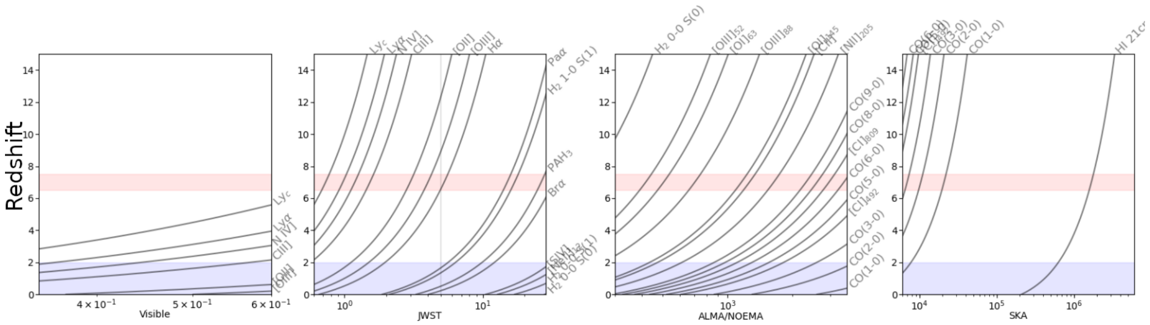
IR spectroscopy

- JWST is observing much fainter IR lines compared to *Spitzer* ⇒ new diagnostics
- Mid/far-IR spectra @z=0 mostly not resolved with *Spitzer*, *Herschel*, same for a potential future IR NASA probe-class mission (waiting for IR space interferometry. . .)

High-z spectroscopy

- JWST is providing optical diagnostics at very high-z, spatially unresolved → ELT
- Exciting JWST+ALMA synergies, ALMA high-z galaxies already show multi-phase ISM *(e.g., Fujimoto+ 2022)*
- UV diagnostics shift to NIRspec when optical ones shift to MIRI *(CLASSY; Mingozzi+ 2023)*

Tracers vs. z



- Sweet spot $z \lesssim 2$ (UV → opt., opt. → NIRspec, near-IR → MIRI, CO ladder with ALMA)
- Sweet spot $z \sim 7$ (UV → NIRspec, far-IR → ALMA, CO → SKA)
- Far-IR diagnostics will remain unvailable at z∼0 until potentially PRIMA
- Mid-IR diagnostics will remain unavailable at z∼2-10 until potentially PRIMA

Same diagnostics with integral field spectroscopy (IFS)

Some instruments and surveys

- SDSS-IV MaNGA, VLT/MUSE, GTC/MEGARA, JWST/MIRI MRS. . .
- Local Universe: SAURON *(de Zeeuw+ 2002)*, ATLAS3D *(Cappellari+ 2011)*, CALIFA *(Sánchez+ 2012)*, SAMI *(Croom+ 2012)*, MaNGA *(Bundy+ 2015)*, PHANGS-MUSE *(Emsellem+ 2021)*. . .
- High-redshift: KMOS3D *(Wisnioski+ 2015)*, SINS/zC-SINF *(Förster Schreiber+ 2018)*. . .
- Future: VLT/BlueMUSE, SDSS-V Local Volume Mapper, ELT/METIS,HARMONI. . .

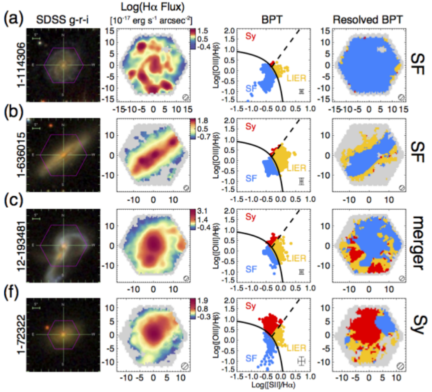
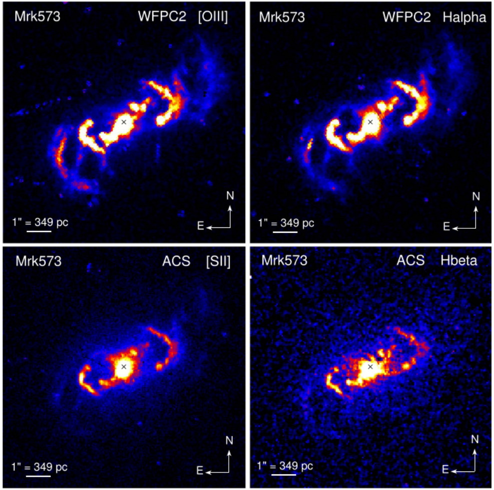


Fig.: Spatially resolved excitation properties of the ionized gas with SDSS/MaNGA (Belfiore+ 2016). LIER component in SF galaxies ∼ DIG.

Narrow-band imaging



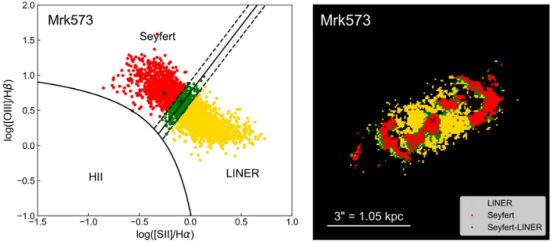


***Fig.**: Spatially resolved BPT mapping of the extended narrow-line regions of nearby Seyfert 2 galaxies with HST (Ma+ 2021).*

How to make the best of existing and future observations

Empirical diagnostics

- We need to understand potential biases, selection effects, and the meaning of average quantities
- We need to design new empirical diagnostics for future observatories with the help of state-of-the-art models

Some complex parameters require full-on models

- Masses (H^+, H^0, H_2...)
- Tracers of $M(H_2)$: (CII) 158 μm, (OI) 63 μm, OH, HD...
- ISM structure (clump distribution, escape fraction of ionizing photons...)
- Multi-phase observations in general
- Ideally at any redshift!

How to make the best of existing and future observations

Empirical diagnostics

- We need to understand potential biases, selection effects, and the meaning of average quantities
- We need to design new empirical diagnostics for future observatories with the help of state-of-the-art models

Some complex parameters require full-on models

- Masses (H^+, H^0, H_2...)
- Tracers of $M(H_2)$: (CII) 158 μm, (OI) 63 μm, OH, HD...
- ISM structure (clump distribution, escape fraction of ionizing photons...)
- Multi-phase observations in general
- Ideally at any redshift!

Outline

- General considerations and motivating questions
  - What are we trying to model?
  - Why are nearby galaxies useful
  - What physical processes to consider
  - What constraints
- Common diagnostics
- Modelling full galaxies, accounting for complexity of ISM and sources
  - Using 1D models
  - Evidence of mixing/smearing issues
  - Using >1D models & n x 1D models
  - Some results and ongoing works
- Statistical framework
  - Samplers
  - Model comparisons / decision tree
- Concluding remarks

Outline

- General considerations and motivating questions
  - What are we trying to model?
  - Why are nearby galaxies useful
  - What physical processes to consider
  - What constraints
- Common diagnostics
- Modelling full galaxies, accounting for complexity of ISM and sources
  - Using 1D models

Some specific "favorable" cases (still complicated!)

Favorable geometries

- Spatially-resolved individual regions (single HII region, single molecular cloud)
- HII galaxies, AGN-dominated galaxies. . . ⇒ single ionizing source

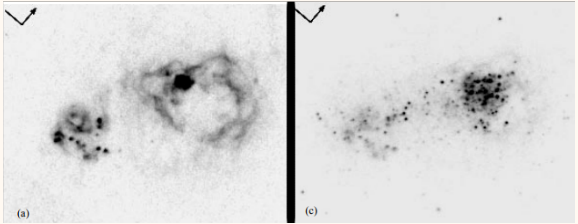


***Fig.**: Ionized gas filaments surrounding a young stellar cluster in the dwarf SF galaxy IZw18, very favorable geometry for models! (Cannon+ 2004).*

- 1D model with spherical geometry or full 3D model can be envisioned (e.g., `Cloudy 3D`, M^3)

Messenger Interface Monte Carlo MAPPINGS V (M^3; *Jin+ 2022*)

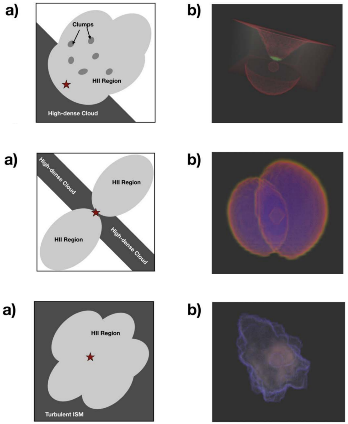


***Fig.**: Modeling nebulae with arbitrary 3D geometries (Jin+ 2022).*

Cloudy 3D/pyCloudy *(Morisset+ 2013)* and PyCROSS *(Fitzgerald+ 2020)*

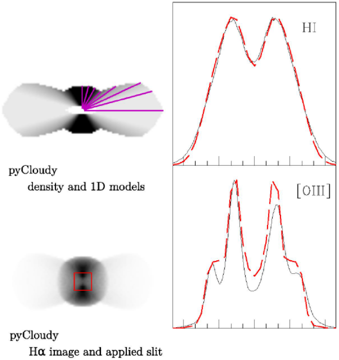


Fig.: *Pseudo-3D models: set of n 1D models following angular laws, populating emissivity cube and projecting (Morisset+ 2013; PN application in Gesicki+ 2016).*

Modeling full galaxies with single 1D models: some codes

1D (or 2D) line and dust RT models

- RADEX *(van der Tak+ 2007)*, RADMC *(Dullemond and Dominik 2004)*. . .
- With LTE or simple non-LTE approximations (such as escape probability or Large velocity gradient - LVG methods)
- No photoionization and no chemistry
- Potentially spherically symmetric
- ⇒ Constraints on physical conditions such as N, n, and T through generation of synthetic spectra & grids

1D photoionization and photochemistry steady-state models

- Cloudy *(Ferland+ 2017)*, MAPPINGS V *(Binette+ 1985, Sutherland+. 2018)*, Meudon PDR *(Le Petit+2006, Bron+ 2016)*. . .
- ⇒ long to run! (e.g., OTF MCMC is difficult)
- Main focus in this presentation

Modeling full galaxies with single 1D models: some codes

1D (or 2D) line and dust RT models

- `RADEX` *(van der Tak+ 2007)*, `RADMC` *(Dullemond and Dominik 2004)*...
- With LTE or simple non-LTE approximations (such as escape probability or Large velocity gradient - LVG methods)
- No photoionization and no chemistry
- Potentially spherically symmetric
- ⇒ Constraints on physical conditions such as N, n, and T through generation of synthetic spectra & grids

1D photoionization and photochemistry steady-state models

- `Cloudy` *(Ferland+ 2017)*, `MAPPINGS V` *(Binette+ 1985, Sutherland+. 2018)*, `Meudon PDR` *(Le Petit+2006, Bron+ 2016)*...
- ⇒ long to run! (e.g., OTF MCMC is difficult)
- Main focus in this presentation

Modeling full galaxies with single 1D models: applications

Applications

- Useful to link observables to "average" physical conditions
- Physical conditions may be interesting by themselves (n, U...) but we're also eventually interested in other resulting parameters (mass of gas, SFR, f_{esc}...)
  - Some codes provide plenty of interesting output quantities from which we can examine things like the formation pathways of H_2, X-ray photoionization etc...

- Single 1D: assuming co-spatial excitation sources (all stellar clusters, potentially AGN – i.e., with coincident mixing)

Prescriptions for photoionization/photodissociation codes

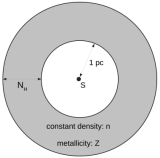
Required prescriptions for models

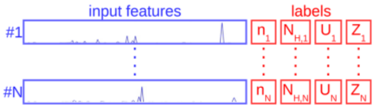
- Abundance patterns, D/G, dust properties. . .
- Equation of state
  - Constant pressure (e.g., Orion, but also for diffuse and translucent clouds; *van Dishoeck and Black 1986*)
  - High densities quickly reached and not well adapted to average galaxy properties. Some alternatives: density scaling with N(H), magnetic field pressure term, pseudo constant pressure. . .
- Shocks: complicated, can use a mechanical heating term (e.g., from SNe rate) flat or not with depth
- CR: nearby starbursts and ULIRGs all suggest MW-like range values $10^{-16,-13}$ s^{-1} with higher values in nuclei and regions with intense SF *(Indriolo+ 2012, 2018; Oka+ 2019; van der Tak+ 2016; Holdship+ 2022; Gonzalez-Alfonso+ 2018)*
- Heavy dependence on stellar atmospheres: BPASS, new versions to test each time. . .

- Extremely low Z: little knowledge on dust opacity curves, CR, stellar atmospheres

Examples: machine learning (ML) techniques

- Supervised ML technique with GAME (GAlaxy Machine learning for Emission lines; *Ucci+ 2017*)





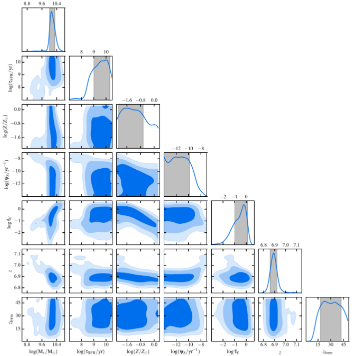
Details

- Library of synthetic spectra assuming very simple, single, 1D models (spherical geometry)
- Z, U, $N(\mathrm{H})$ predictions from an arbitrary suite of emission lines
- Great performance for a large number of tracers (better than Bayesian techniques in that regard)
- Possible application to IFS observations *(Ucci+ 2019)*

Examples: probablistic methods

Going Bayesian with BEAGLE (BayEsian Analysis of GaLaxy sEds; *Chevallard and Charlot 2016*)

- Using *Gutkin+ (2016)* models that combine stellar population synthesis and photoionization codes to describe an ensemble of HII regions and diffuse gas ionized by young stars
- Effective HII region: all HII regions and DIG ionized by a single stellar generation with a set of effective parameters
  - Strong – though classic – assumption
  - Collection of isolated HII regions currently being investigated
- Geometry accounted for by dust attenuation for stellar+nebular emission (inclination, disk, bulge)
- Powerful algorithm including instrumental effects
- Using nested sampling techniques to account for multi-modal posterior distributions

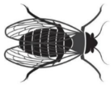


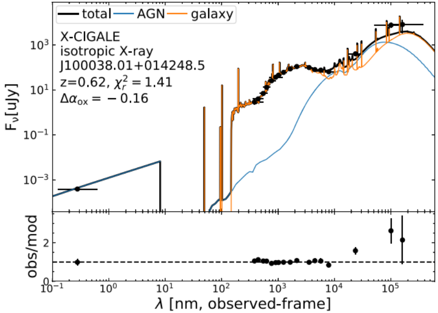
Examples: probablistic methods

Going Bayesian with the Code Investigating GALaxy Emission (`CIGALE`; *Burgarella+ 2005*)

- Relying on energy balance principle (absorption by dust in UV-optical vs. re-emission in IR)
- Using geometry templates for dust attenuation
- Detailed treatment of X-ray sources: `X-CIGALE` *(Yang+ 2022)*
- Nebular emission treated *(Boquien+ 2019)* to decontaminate broadband photometry
  - Line predictions for HII region + PDR under development

- ⇒ Hands-on project @GISM2

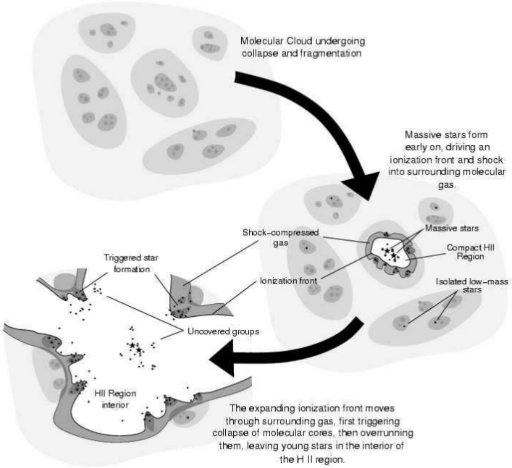




Limitations of single 1D models

- Like all static nD models, cannot capture the complex and dynamic structure of the ISM along with all of the relevant, time-varying star formation and feedback
- 1D models assume co-spatial sources / effective galactic-wide parameters
- Like all parameterized models: wide range of theoretically allowed parameters
- Distribution/geometry of gas is difficult to implement (e.g., HII region, PDR, molecular cloud)

Limitations of single 1D models



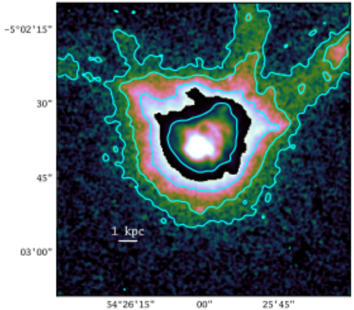
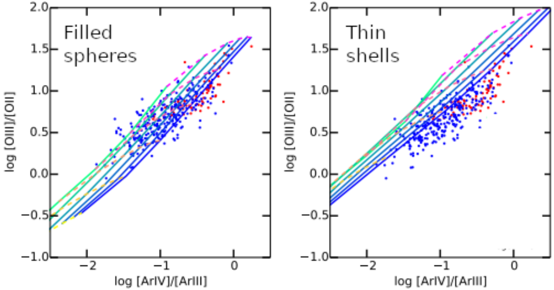


Fig.: *Left: evolution of HII regions (Hester and Desch 2005). Right: Escaping photons channels through (OIII) with MUSE (Herenz+ 2020).*

Possible tweaks to single 1D model approaches

- Accounting for different (spherical) geometries
- Accounting for holes and/or matter-bounded (aka density-bounded) models
- Accounting for time evolution

Assumption of geometry



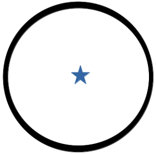


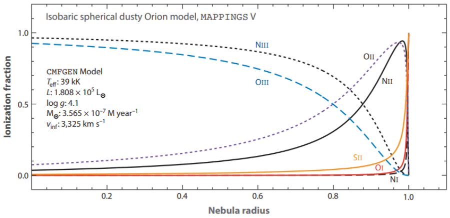
***Fig.**: Line ratio diagnostic line assuming different spherical geometries (Stasinska+ 2015).*

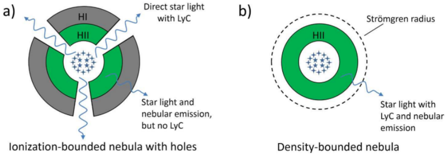
- Playing with geometry not enough to reproduce low values of (OI)/(OIII) in LyC-leaking galaxies or (CII)/(OIII) in high-z galaxies

Accounting for holes and/or matter-bounded models

BPT-like diagnostics & metallicity diagnostics

- Matter-bounded nebula produce the "normal" amount of (OIII) close to the stars, but some H RL-emitting regions are missing further out ⇒ (OIII)/Hβ $\nearrow$
- Matter-bounded nebula lack low-ionization lines (e.g., (NII), (SII)) emitting regions





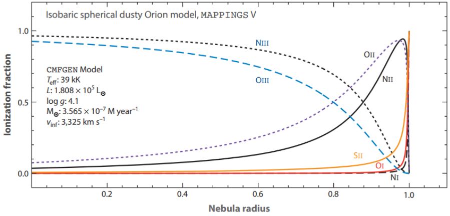
Escape fraction of ionizing photons

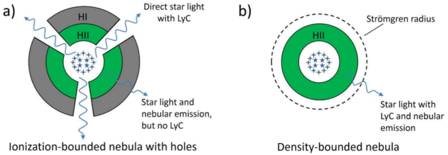
- As a result, weak low-ionization lines may be used to identify potential leakers (e.g., Wang+ 2020, 2021, Zackrisson+2013, Ramambason+ 2020, 2022)

Accounting for holes and/or matter-bounded models

BPT-like diagnostics & metallicity diagnostics

- Matter-bounded nebula produce the "normal" amount of (OIII) close to the stars, but some H RL-emitting regions are missing further out $\Rightarrow$ (OIII)/Hβ $\nearrow$
- Matter-bounded nebula lack low-ionization lines (e.g., (NII), (SII)) emitting regions





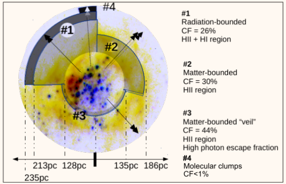
Escape fraction of ionizing photons

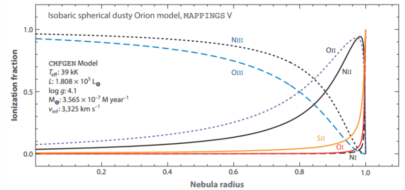
- As a result, weak low-ionization lines may be used to identify potential leakers *(e.g.,, Wang+ 2020, 2021, Zackrisson+2013, Ramambason+ 2020, 2022)*

Accounting for matter-bounded models

Once upon a time

- Long-standing high T_e([OIII]) problem in dwarf galaxy IZw18: extra heating due to stellar winds, SNe, shocks?
  - T_e constrained by [OIII]λ4363Å/([OIII]λ5007Å+[OIII]λ4959Å) and depends on local density
  - Density from [SII] is not representative of the ionized nebula





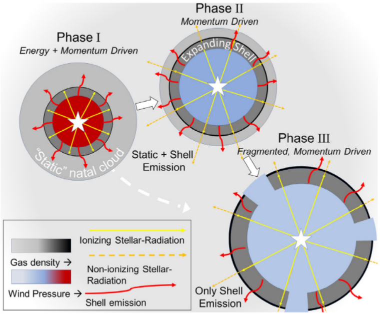
- *Péquignot (2008)*: radiation-bounded shells embedded in a matter-bounded medium produces a lower mean density and higher $T_e \Rightarrow$ photoionization by (non-population III) hot stars is enough!
- Introducing "topological models", i.e., combination of 1D models

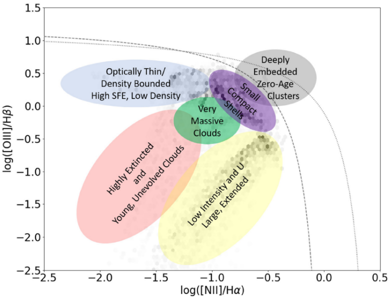
Accounting for time evolution of single HII region + PDR

△ Requiring coupling with time-dependent models

1D spherical with time-evolution

- WARPFIELD-EMP *(Pellegrini+ 2020)* couples the 1D stellar feedback code WARPFIELD with the Cloudy HII region/PDR code and the POLARIS line and continuum RT code, in order to make detailed predictions for the time-dependent emission arising from the HII region and PDR surrounding an evolving star cluster





Break

- Questions
- Take a good breath
- (Wake up?)

Outline

- General considerations and motivating questions
  - What are we trying to model?
  - Why are nearby galaxies useful
  - What physical processes to consider
  - What constraints
- Common diagnostics
- Modelling full galaxies, accounting for complexity of ISM and sources
  - Using 1D models
  - Evidence of mixing/smearing issues

Mixing

- Eventual goal: build a comprehensive model able to explain the multi-phase signatures and able to account for complex ISM/source geometries
  - Unresolved spectroscopy is often inevitable (e.g., high-z galaxies, some specific wavelength domains)

Components

- Galaxies in general do include:
  - A collection of HII regions following some luminosity function, some leaking ionizing photons – possibly super-stellar clusters as well ⇒ Q, U mixing
  - A distribution of gas following some density PDF related to turbulence, self-gravitation, and rotational support *(e.g., Khullar+ 2021)* ⇒ n, P mixing (biases depend on critical densities)
  - A collection of molecular clouds, some associated with recent SF
  - WR stars, high-mass X-ray binaries and possibly AGN

Mixing: distributions

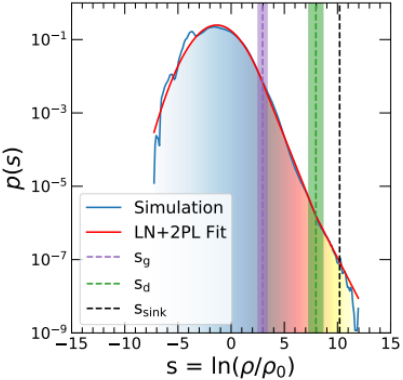


***Fig.**: Density PDF from simulations with 3 regimes associated with turbulence, self-gravitation, and disc/rotation (Khullar+ 2021).*

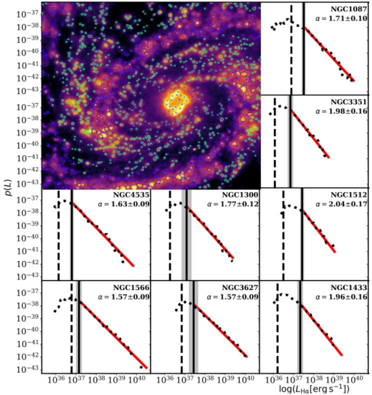
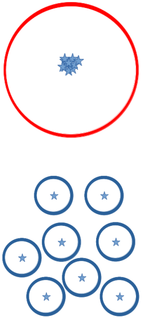


***Fig.**: HII region luminosity function in PHANGS-MUSE galaxies (Santoro+ 2022).*

Mixing: effective HII region vs. collection of HII regions

- Stellar population radiation field (BPASS), potential X-ray source, fixing all but *U* and SED shape (age)
- (Don't read too much into this, depends a lot on how models are designed)



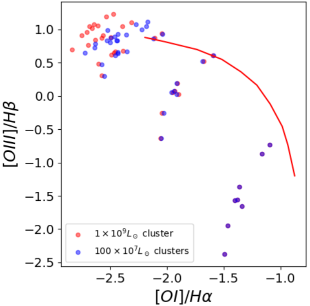


***Fig.**: Globally high-density regime, same volume. But same parameters lead to different line fluxes for high U values (thinnest nebulae; dust absorption).*

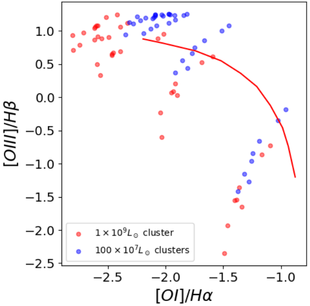
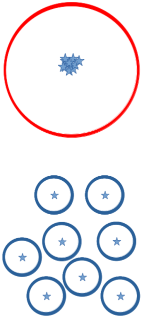


***Fig.**: Absorption of X-rays doesn't follow the same "rules" as UV photons. Line ratios cannot be recovered even choosing a different U value, this is a geometry effect.*

Mixing: effective HII region vs. collection of HII regions

- PDR diagnostics (no X-ray source)



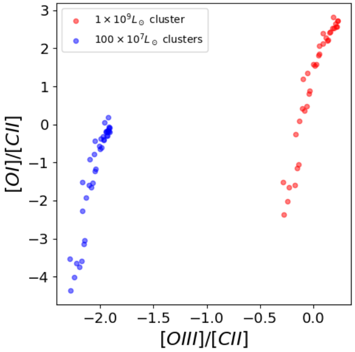


Fig.: *The effective region model would need a much lower density to match better with collection of models.*

Mixing: (non-)coincident AGN/SF mixing

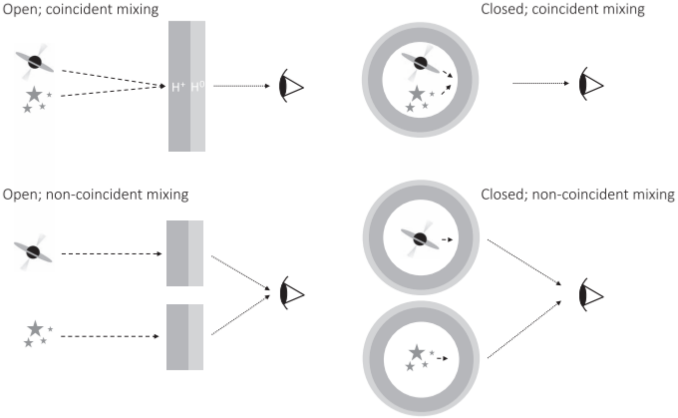
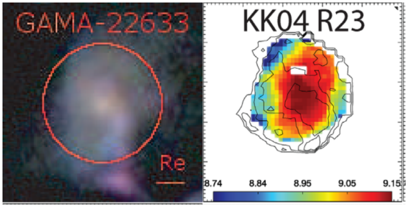


Fig.: *Different geometries for AGN and SF excitation (Richardson+ 2022).*

Evidence of inhomogeneities

- IFS observations reveal the mixing, AGN contribution, metallicity variations. . . within galaxies. This should be kept in mind when modelling a spatially-unresolved galaxy
- Kewley+ (2019): *"For example, the global metallicity of a high-redshift galaxy may not be the true mean metallicity but may be weighted toward specific HII regions with certain sets of properties."*



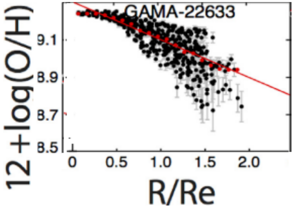


***Fig.**: Metallicity gradient and dispersion in SAMI galaxies (SF face-on spirals; Poetrodjojo+ 2018).*

Biases due to smearing and selection effects

IFS results

- M* can be severely underestimated (factors up to 5) using the integrated SED due to the bias of young stars dominating the SED *(Sorba and Sawicki 2015, 2018)*
  - See also *Galliano+ (2011)* for dust mass estimates vs. spatial resolution in LMC
- Detailed study of biases due to beam smearing for spectroscopic diagnostics still limited (e.g., Z, SFR, $M(\mathrm{H}^+)\ldots$)
  - Note that a single 1D component always imply some kind of bias for a single pixel in IFS observations (e.g., SAMI, MUSE. . .) ⇒ Longitudinal mixing and spatial disconnection between excitation source and matter that may lie in different pixels

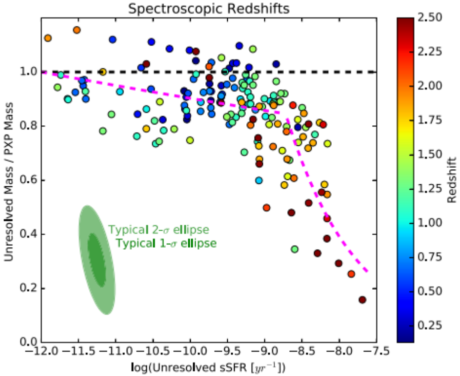
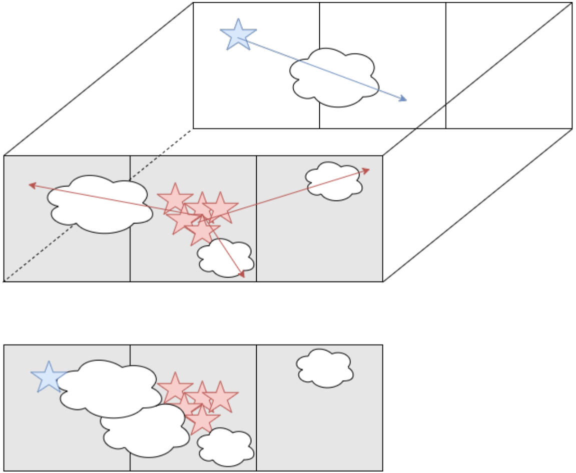


***Fig.**: Missing mass for high sSFR galaxies (Sorba and Sawicki 2018)*

Beam smearing and LOS mixing in IFS observations



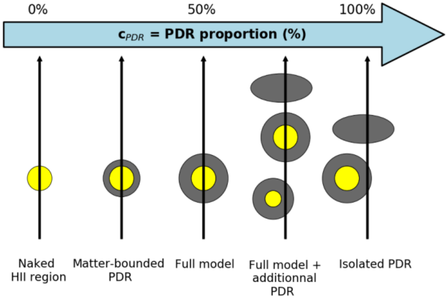


***Fig.**: Line of sight mixing (Lambert-Huyghe+ 2022).*

Outline

- General considerations and motivating questions
  - What are we trying to model?
  - Why are nearby galaxies useful
  - What physical processes to consider
  - What constraints
- Common diagnostics
- Modelling full galaxies, accounting for complexity of ISM and sources
  - Using 1D models
  - Evidence of mixing/smearing issues
  - Using >1D models & n x 1D models

Beyond single 1D models: options (and difficulties)

- Combination of 1D models
  - Pseudo 3D
  - Combination of independent 1D models representing galaxy components
- Full 3D RT
- Simulations

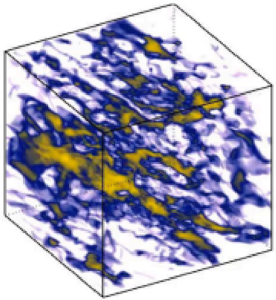
(Focus on nebular emission, for panchromatic SED models, see, e.g., *Conroy+ 2013; Baes+ 2019*)

Simulations

Simulations

- Include dynamical effects, a realistic/consistent ISM structure and distribution of sources
  - Chemistry is numerically expensive (most simulations do not include any form of non-equilibrium metal chemistry)
  - Need to rely on subgrid models to account for the physics on sub-pc scales (including resolving the ionization fronts)

- Post-processing: feed numerical simulations to photoionization codes or chemical networks in order to measure the metal ionization states and their relevant emission after the simulation has been run *(e.g., Jonsson+ 2010, Melekh+ 2015, Vandenbroucke and Wood 2019)*
  - See Hirschmann+ (2017, 2019, 2022,2023) for post-processing with `Cloudy` (photoionization) and `MAPPINGS V` (fast radiative shocks)



Simulations with large chemical networks

Simulations

- Solving a large chemical network within a 3D simulation, e.g., combination of thermochemical network PRISM with on-the-fly radiation hydrodynamics code RAMSES−RTZ
  - Full 3D cosmological or isolated galaxy simulations *(e.g., Katz+ 2022)*
- Study of cooling and heating processes in the ISM, synthetical observations. . .
- Prescriptions usually limited to general properties rather than individual objects or even samples, exploration of large regions of parameter space remains difficult
- Ideally: grids of 1000s-100,000s simulations with varied parameters to produce synthetic library of spectra to compare to observations! (can heat entire labs in the winter. . .)

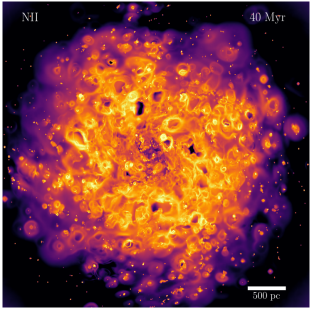


***Fig.**: SMC-like simulated galaxy, N^{+} column density (Katz+ 2022).*

Simulations with large chemical networks

Simulations

- Very useful to test "unmixing" techniques or spatial-resolution biases!
  - Known distributions of density, metallicity…
  - We would like to reproduce the average parameters and their dispersion

- △ Comparison of intrinsic parameter average values or variations in simulations vs. parameters derived from emitting regions is not trivial
  - Not all cells in simulations lead to emitting species
  - Biases due to instrumental uncertainties
  - What internal distributions should be used in models?

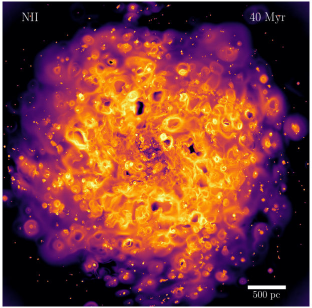


***Fig.**: SMC-like simulated galaxy, N^+ columndensity (Katz + 2022).*

Full 3D RT (w/o photoionization)

Full 3D RT

- Adapted to objects with known (potentially complex) geometries
- 3D Monte-Carlo RT codes can handle complex geometries and density structures
  - e.g.: `RADMC-3D` *(Dullemond+ 2012)*, `SKIRT` *(Baes+ 2003, 2011)*: produce synthetic images/spectra from an arbitrary distribution of stars, dust, and gas density distribution from 1- to 3-D
  - Applications often limited to stellar populations and dust-heating processes *(e.g., de Looze+ 2014)*
- Future is fully self-consistent 3D RT models, which will allow detailed dust and gas distributions to be embedded within the photoionized nebula with arbitrary T, n, and dust distributions
  - Promising avenues with `SKIRT+Cloudy` *(Romero+ 2023)*

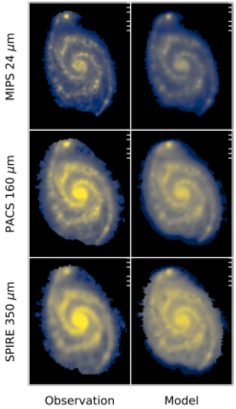
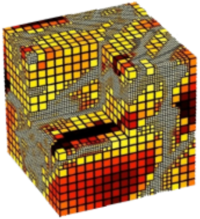


***Fig.**: Comparison simulated/observed images for M51 (Nersesian+ 2020).*

Full 3D RT (w/ photoionization)

Full 3D RT

- Full 3D photoionization is great but expensive (MOCASSIN, TORUS−3DPDR, SOC/LOC, RASCAS, ART2, M^3...)
  - Need to know the distribution of matter and sources (geometry is not a free parameter)
  - Particularly adapted to PNe, bipolar HII regions, fractal HII regions...



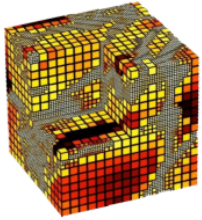
Overall pros and cons of 3D methods

- Great to treat the transfer and deal with projection effects
- Great to test the impact of geometry or model specific objects
- So far impractical to explore large parameter space
- Still cannot capture the complex and dynamic structure of the ISM along with all of the relevant, time-varying star formation and feedback

Full 3D RT (w/ photoionization)

Full 3D RT

- Full 3D photoionization is great but expensive (MOCASSIN, TORUS-3DPDR, SOC/LOC, RASCAS, ART2, M^3...)
  - Need to know the distribution of matter and sources (geometry is not a free parameter)
  - Particularly adapted to PNe, bipolar HII regions, fractal HII regions. . .



Overall pros and cons of 3D methods

- Great to treat the transfer and deal with projection effects
- Great to test the impact of geometry or model specific objects
- So far impractical to explore large parameter space
- Still cannot capture the complex and dynamic structure of the ISM along with all of the relevant, time-varying star formation and feedback

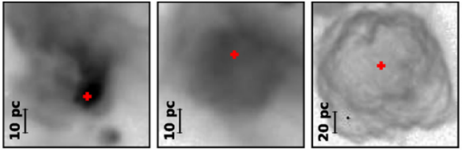
Back to 1D: combination of independent 1D models

Principle

- Integrated spectrum of a galaxy ∼ sum of many emitting stars+ISM components that are correlated or that may even share similar properties (not a new idea)

Example: ensemble averages of aging HII regions *(Dopita+ 2006; Groves+ 2008)*

- Removing the (single) age parameter
- Continuous formation of stellar clusters (each cluster forming stars coevally) ⇒ Evolutionary track of an HII region with given parameter sets
- Flux-averaged spectra along this track



Combination of independent 1D models

Locally optimally emitting cloud (LOC; *Ferguson+ 1997; Richardson+ 2016*)

- Assumes that the cumulative observed emission from each individual emission-line galaxy is the result of selection effects stemming from various emission lines optimally emitted by a large number of gas clouds spanning a large range in physical conditions
  - Fully parameterized, useful for non trivial components like AGNs
  - Potentially >100s of models

$$L_{line} = \underbrace{\int ... \int}_{\mathrm{n}} L(p_1, ..., p_n)\psi(p_1, ..., p_n)dp_i...dp_n \qquad \psi = U^{\alpha_U} n^{\alpha_n}...$$

General model "architecture": topological models

- Linear combination of independent 1D models
  - "Topology" vs. geometry: the exact way in which the components are distributed doesn't matter
  - Many models (grids) but less computationally intensive than simulations or full 3D models

Combination of independent 1D models

Locally optimally emitting cloud (LOC; *Ferguson+ 1997; Richardson+ 2016*)

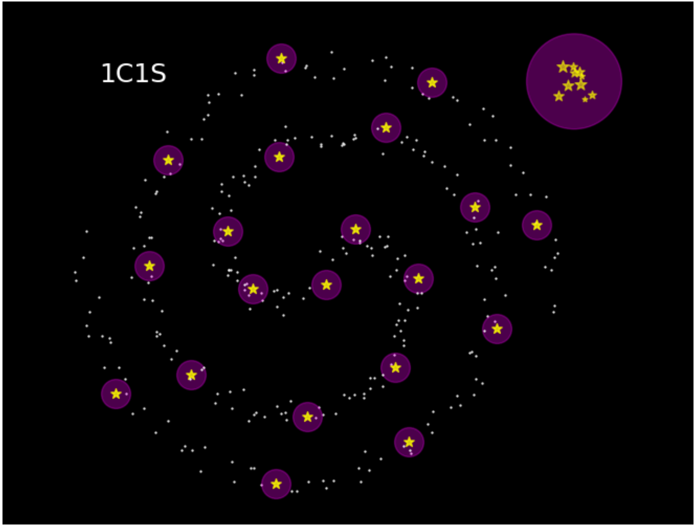
- Assumes that the cumulative observed emission from each individual emission-line galaxy is the result of selection effects stemming from various emission lines optimally emitted by a large number of gas clouds spanning a large range in physical conditions
  - Fully parameterized, useful for non trivial components like AGNs
  - Potentially >100s of models

$$L_{line} = \underbrace{\int ... \int}_{\text{n}} L(p_1, ..., p_n)\psi(p_1, ..., p_n)dp_i...dp_n \qquad \psi = U^{\alpha_U} n^{\alpha_n}...$$

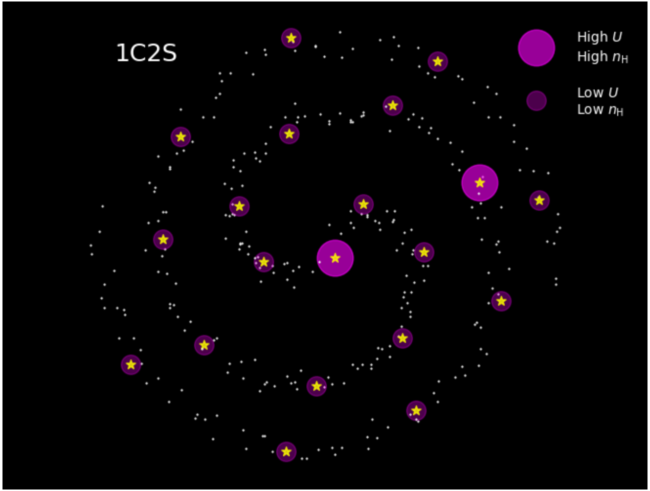
General model "architecture": topological models

- Linear combination of independent 1D models
  - "Topology" vs. geometry: the exact way in which the components are distributed doesn't matter
  - Many models (grids) but less computationally intensive than simulations or full 3D models

Single effective representative cluster, single ISM component

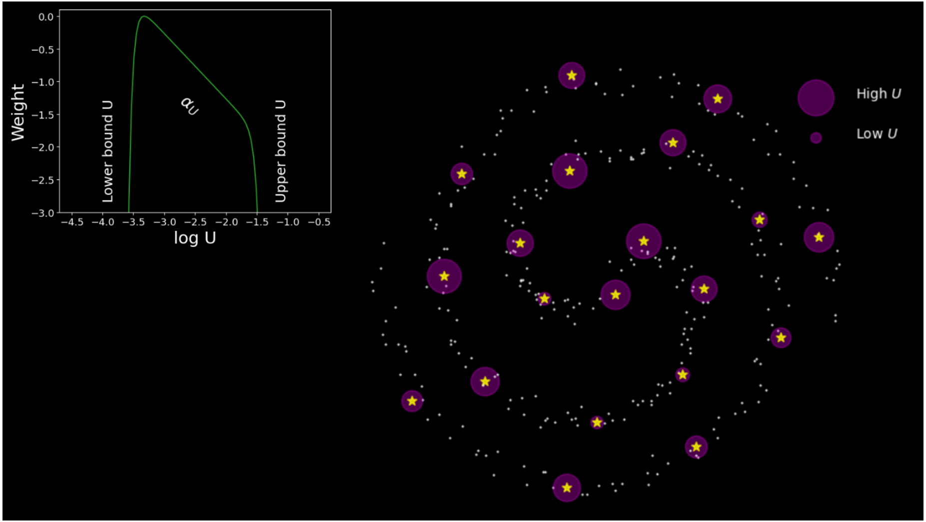


Single effective representative cluster, two ISM components

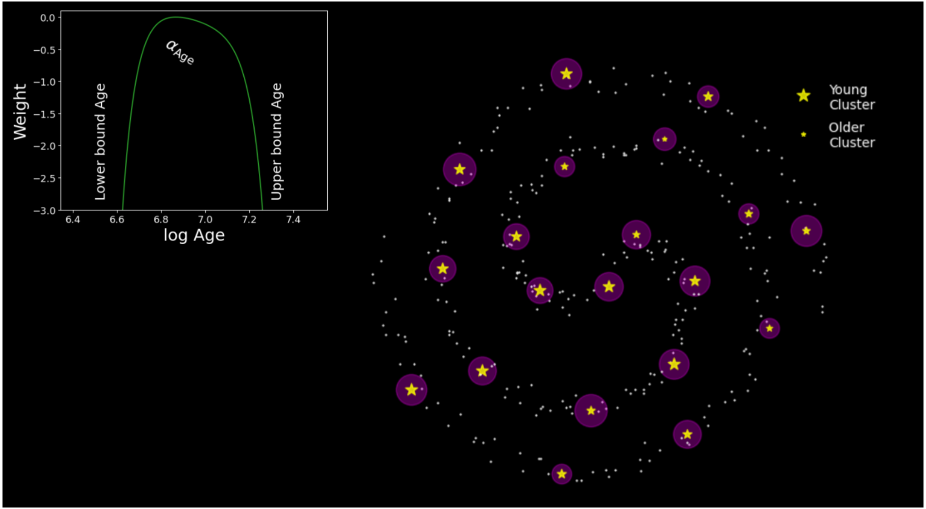


Each component is described by a single value, this doesn’t sound very realistic. . .

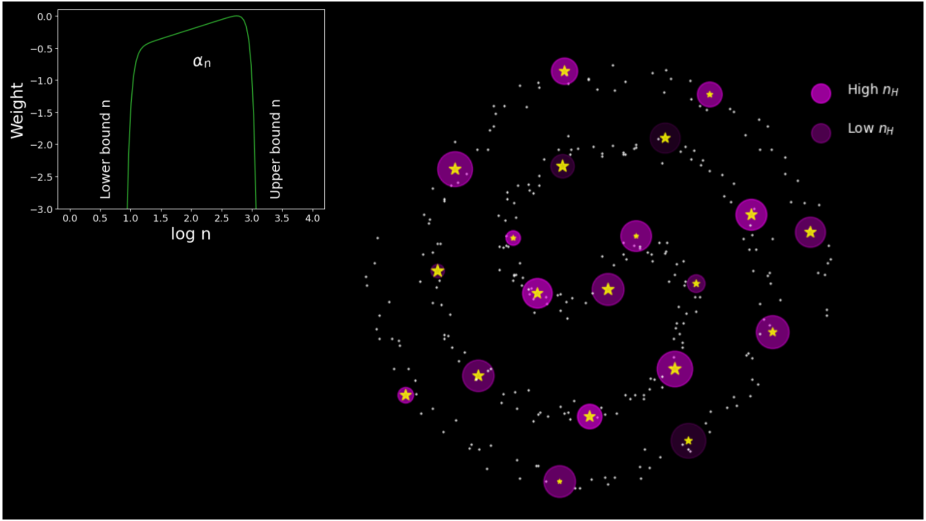
Single effective representative cluster, U distribution



U & age distribution

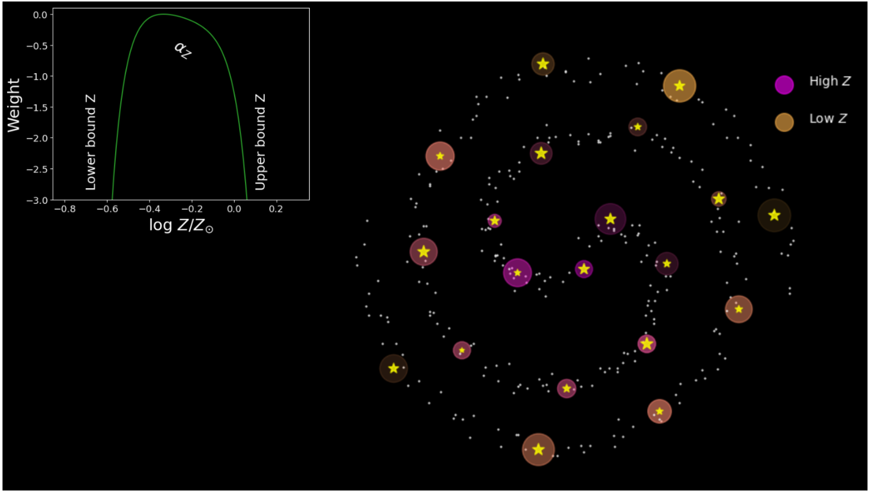


U & age & n distribution

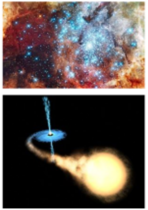


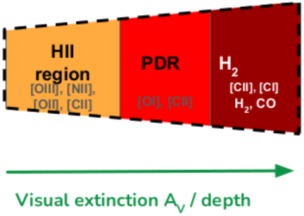
U & age & n & Z distribution

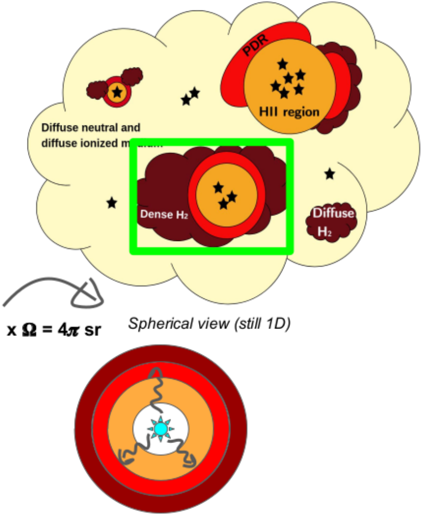
(Actual spatial distribution is not important)



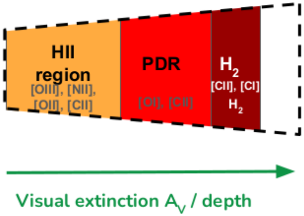
Accounting for physical depth of clouds

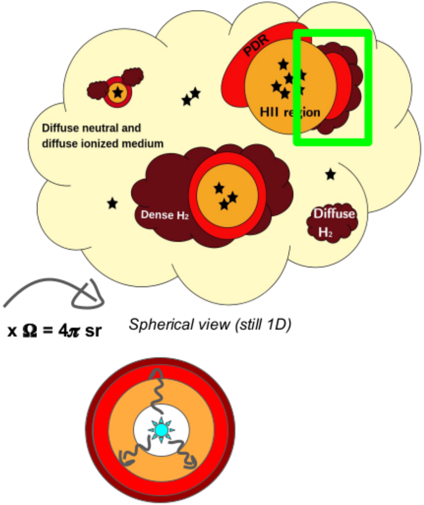




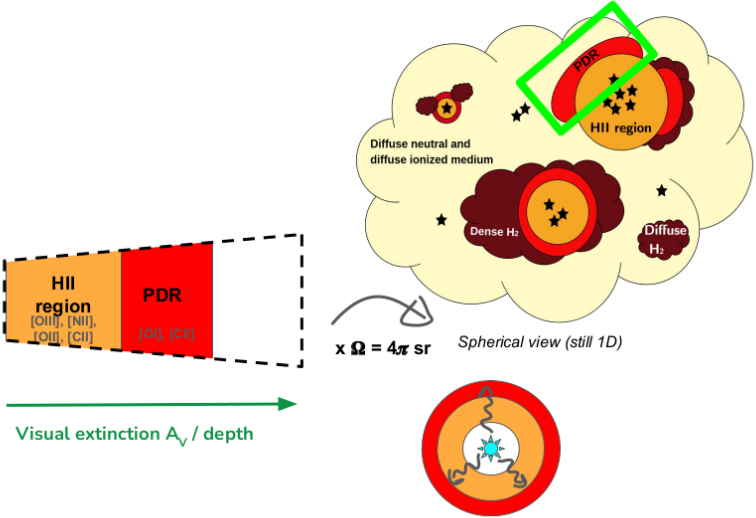


Accounting for physical depth of clouds

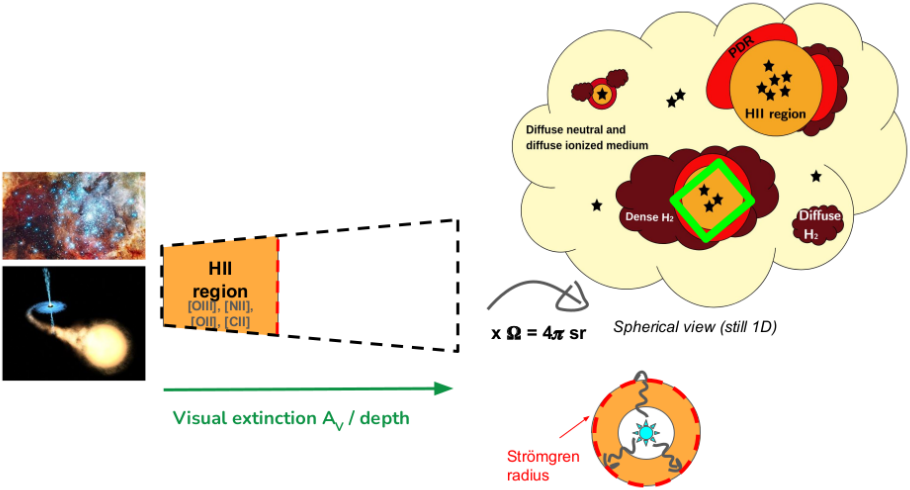




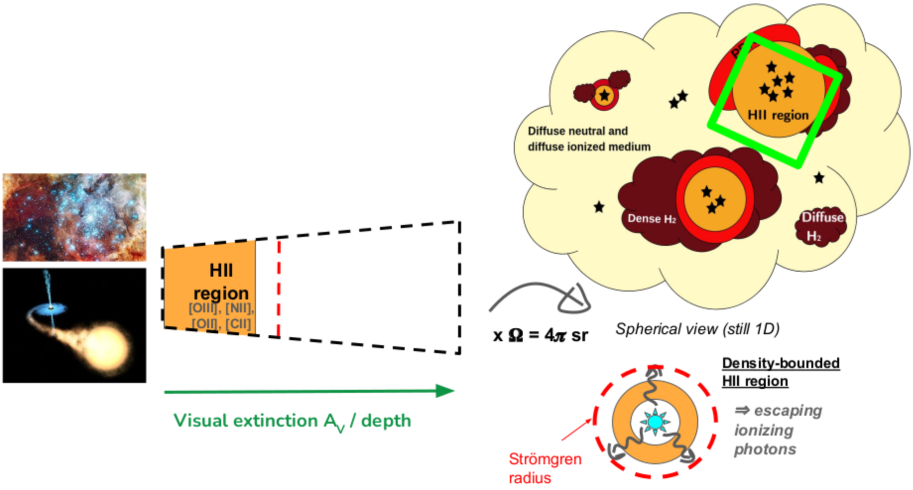
Accounting for physical depth of clouds



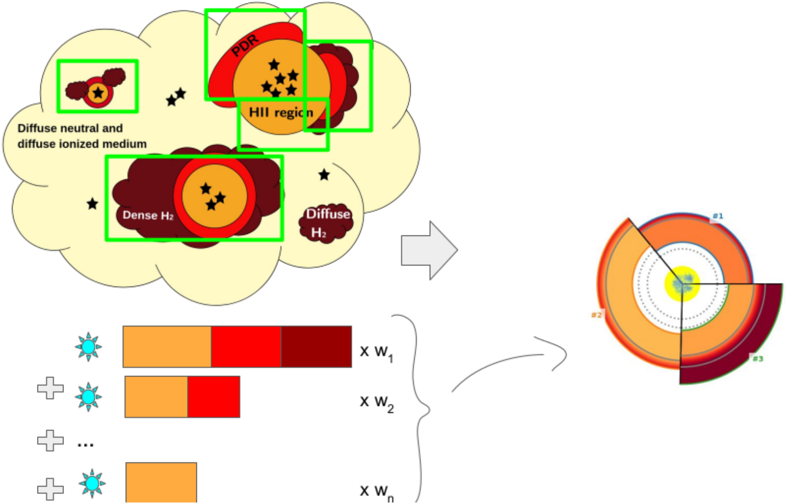
Accounting for physical depth of clouds



Accounting for physical depth of clouds



Accounting for physical depth of clouds



Iterations of (simple) combinations

IZw18
(Lebouteiller+ 2017)

3 & 4-sector radiation- or matter-bounded
CFs tailored
Specific model

Haro 11
(Cormier+ 2012)

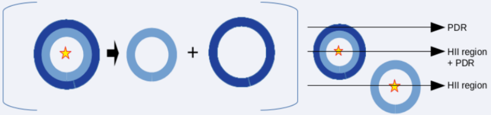
HII region + PDR CF scaled a posteriori

Diffuse ionized gas

IC10
(Polles 2017; Polles+ 2019)

Radiation-bounded or matter-bounded

SMC/LMC SF regions
(Lambert-Huyghe+ 2022)



Combination of HII region + PDR from single radiation-bounded sector

DGS
(Cormier+ 2019)

2 radiation-bounded sectors, Only dense HII region has a PDR CFs from grid

Outline

- General considerations and motivating questions
  - What are we trying to model?
  - Why are nearby galaxies useful
  - What physical processes to consider
  - What constraints
- Common diagnostics
- Modelling full galaxies, accounting for complexity of ISM and sources
  - Using 1D models
  - Evidence of mixing/smearing issues
  - Using >1D models & n x 1D models
  - Some results and ongoing works

Combinations of 1D models for nearby galaxies: flash results

Matching suite of lines

- Reproducing T_e([OIII]) in IZw18 *(Péquignot 2008)*
- Matching many IR and optical lines (PDR+HII region) at once in an unresolved galaxy *(Cormier+ 2012)*

PDR and CO-dark gas

- PDR "covering factor" ↘ when Z ↘ *(Cormier+ 2019)*
- CO-dark gas fraction is a function of Z and geometry *(Ramambason+ in prep.)*
- Origin of [CII] in the neutral atomic gas and influence of X-ray sources in ISM heating *(Lebouteiller+ 2017)*
- Evidence of wide range of PDR fractions in spatially-resolved SF regions *(Lambert-Huyghe+ 2022)*

Escape fraction of ionizing photons

- Fraction of escaping photons ↘ when integrating larger spatial scales *(Polles+ 2017, 2019)*
- Fraction of escaping photons ↗ when Z ↘ *(Ramambason+ 2022)*

Illustration of ongoing works: LOC models

- (Testing phase!) Comparing (single) parameter values from single 1D models to average LOC parameters: single 1D model captures well the average *U* and *n*, not so good for age (i.e., stellar population SED shape) and *Z*

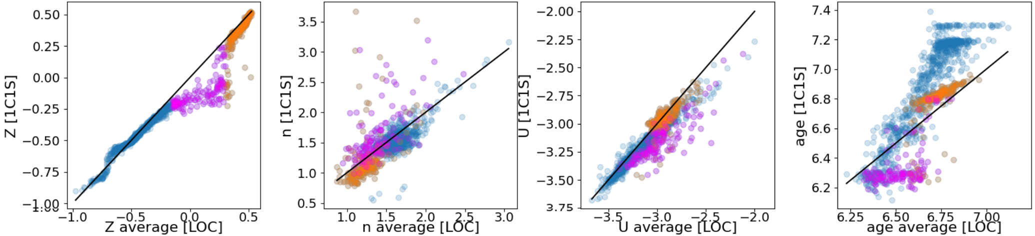
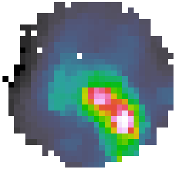
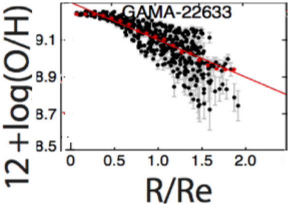
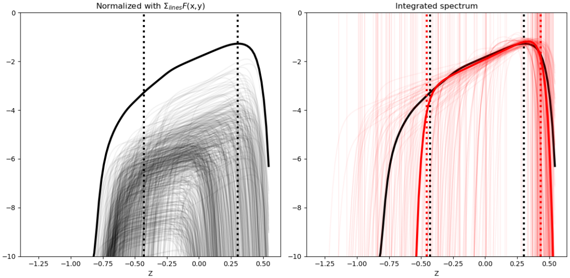


Illustration of ongoing works: recovering internal variations

- (Testing phase!) Each IFS pixel is a distribution, can we recover the internal variations from the integrated map?







Applications to high-z galaxies

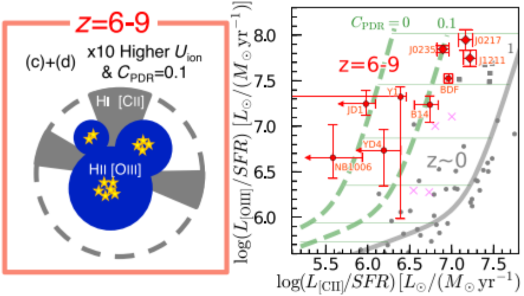


Fig.: *Considering HII region + PDR components in high-z galaxies to explain high (OIII)/(CII) observed with ALMA (Harikane+ 2020).*

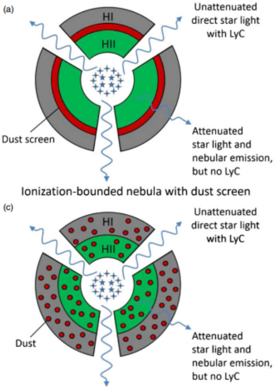


Fig.: *Considering different dust distributions for EoR galaxies (Zackrisson+ 2013).*

Caveats of combination of 1D models

- △Still 1D models: static, components don't talk to each other
- △Projection effects are difficult to handle
  - Better adapted to optically thin tracers and dust-poor ISM
- △Still parameterized models: potentially too wide allowed parameter space
  - Considered models combinations may be matching observations but may lead to parameter distributions that are unrealistic or not motivated/confirmed by self-consistent simulations
  - Unless priors are explicitly introduced

Outline

- General considerations and motivating questions
  - What are we trying to model?
  - Why are nearby galaxies useful
  - What physical processes to consider
  - What constraints
- Common diagnostics
- Modelling full galaxies, accounting for complexity of ISM and sources
  - Using 1D models
  - Evidence of mixing/smearing issues
  - Using >1D models & n x 1D models
  - Some results and ongoing works
- Statistical framework
  - Samplers
  - Model comparisons / decision tree
- Concluding remarks

Outline

- General considerations and motivating questions
  - What are we trying to model?
  - Why are nearby galaxies useful
  - What physical processes to consider
  - What constraints
- Common diagnostics
- Modelling full galaxies, accounting for complexity of ISM and sources
  - Using 1D models
  - Evidence of mixing/smearing issues
  - Using >1D models & n x 1D models
  - Some results and ongoing works
- Statistical framework
  - Samplers

Statistical framework

- How do we handle this many tracers and/or such complex combinations?

Deterministic methods

- Focus on selection of "best-fitting" parameters rather than on the uncertainties associated with these parameters, e.g., χ^2 method
  - Limited capabilities with interpolation, outliers, upper limits, more complex topology, confidence intervals, use of priors. . .

Probabilistic methods

- Bayesian inference with state-of-the-art posterior sampling techniques, such as the Markov Chain Monte-Carlo (MCMC) has become a standard practice
- Probabilistic approach also introduces priors, nuisance variables, and may allow a finer scan of the parameter space

Statistical framework

- How do we handle this many tracers and/or such complex combinations?

Deterministic methods

- Focus on selection of "best-fitting" parameters rather than on the uncertainties associated with these parameters, e.g., χ^2 method
  - Limited capabilities with interpolation, outliers, upper limits, more complex topology, confidence intervals, use of priors. . .

Probabilistic methods

- Bayesian inference with state-of-the-art posterior sampling techniques, such as the Markov Chain Monte-Carlo (MCMC) has become a standard practice
- Probabilistic approach also introduces priors, nuisance variables, and may allow a finer scan of the parameter space

Bayesian inference

- Single models too long to run at each iteration ⇒ fine grid and/or interpolation techniques

Inference on grids including pre-computed geometry

- `NebulaBayes` *(Thomas+ 2016)* (some caveats with line normalization hypothesis)
  - Agnostic to model grid used so desired topology can be pre-computed/tabulated in grid
  - "Brute force" Bayesian likelihood calculations
- `MULTIGRIS` *(Lebouteiller+ 2022)*: same as `NebulaBayes` but with a sampler
  - + RVs controlling combination of models and any nuisance variables (e.g., extinction, systematic uncertainties. . .)

Random walkers (i.e., one or a few "chains")

- High dimensionality is common ⇒ slow parameter space exploration
- ISM model grids are highly multi-modal, worse if combination of models. . .
  - Can get stuck in local likelihood maximum
  - Stochasticity: different solutions from different starting points
  - Difficult to probe the entire parameter space (⇒ difficult to compute marginal likelihood)

Bayesian inference

- Single models too long to run at each iteration ⇒ fine grid and/or interpolation techniques

Inference on grids including pre-computed geometry

- `NebulaBayes` *(Thomas+ 2016)* (some caveats with line normalization hypothesis)
  - Agnostic to model grid used so desired topology can be pre-computed/tabulated in grid
  - "Brute force" Bayesian likelihood calculations
- `MULTIGRIS` *(Lebouteiller+ 2022)*: same as `NebulaBayes` but with a sampler
  - + RVs controlling combination of models and any nuisance variables (e.g., extinction, systematic uncertainties. . .)

Random walkers (i.e., one or a few "chains")

- High dimensionality is common ⇒ slow parameter space exploration
- ISM model grids are highly multi-modal, worse if combination of models. . .
  - Can get stuck in local likelihood maximum
  - Stochasticity: different solutions from different starting points
  - Difficult to probe the entire parameter space (⇒ difficult to compute marginal likelihood)

General principle

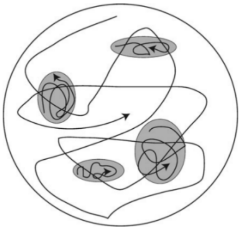
- $p(\theta|\mathbf{O},\mathcal{M}) = \frac{p(\mathbf{O}|\theta,\mathcal{M})p(\theta|\mathcal{M})}{p(\mathbf{O}|\mathcal{M})} \propto p(\mathbf{O}|\theta,\mathcal{M})p(\theta|\mathcal{M})$
- $posterior = \frac{likelihood \times prior}{marginalization}$ (for a given model)

Multi-modal posteriors

- Nested sampling techniques (e.g., BEAGLE)
- Sequential Monte-Carlo (e.g., MULTIGRIS): Markov kernels used to rejuvenate particle using IMH or HMC kernels
- Genetic algorithms. . .

Example: particle filtering techniques (e.g., SMC)

- Tempered likelihood *(Ching and Chen 2007, Mison+ 2013)*
  - Parallel runs varying the "temperature" of (many) particles through the index β
  - $p(\theta|\mathbf{O},\mathcal{M})_\beta \propto p(\mathbf{O}|\theta,\mathcal{M})^\beta p(\theta|\mathcal{M})$



General principle

- $p(\theta|\mathbf{O},\mathcal{M}) = \frac{p(\mathbf{O}|\theta,\mathcal{M})p(\theta|\mathcal{M})}{p(\mathbf{O}|\mathcal{M})} \propto p(\mathbf{O}|\theta,\mathcal{M})p(\theta|\mathcal{M})$
- $posterior = \frac{likelihood \times prior}{marginalization}$ (for a given model)

Multi-modal posteriors

- Nested sampling techniques (e.g., BEAGLE)
- Sequential Monte-Carlo (e.g., MULTIGRIS): Markov kernels used to rejuvenate particle using IMH or HMC kernels
- Genetic algorithms...

Example: particle filtering techniques (e.g., SMC)

- Tempered likelihood *(Ching and Chen 2007, Mison+ 2013)*
  - Parallel runs varying the "temperature" of (many) particles through the index β
  - $p(\theta|\mathbf{O},\mathcal{M})_\beta \propto p(\mathbf{O}|\theta,\mathcal{M})^\beta p(\theta|\mathcal{M})$

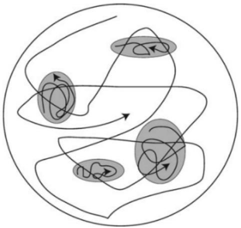
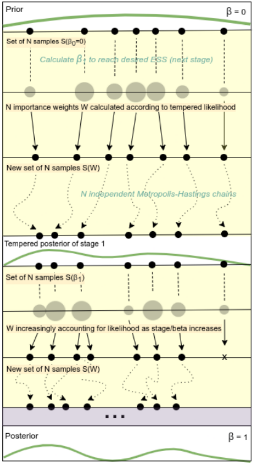


Illustration (SMC)



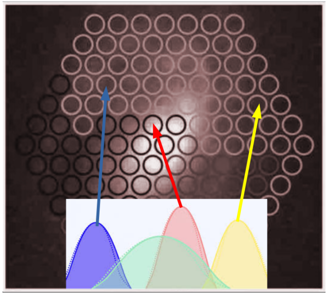
- Makes it possible to evaluate the entire parameter space if well sampled
  - Marginal likelihood can be estimated

Illustrations (see accompanying .odp file)

- No U-turn Sampler (Hamiltonian MC sampler)
- SMC

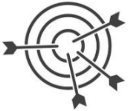
Hierarchical method

- Sample of galaxies
- Sample of galaxy regions
- Pixels in IFS observations $\Rightarrow$ spatial regularization / smoothing
- Still very recent for spectroscopic diagnostics



The power of statistics

- We usually don't know the geometry a priori: assumed geometry and infer parameters
- Usually assuming simple geometries when few tracers are available, and increase complexity with additional tracers (*"Don't use more parameters than tracers"*)
- BUT more complex (i.e., realistic) geometries can still be explored/evaluated!
  - Is it better to knowingly use an unrealistic geometry and infer inaccurate (possibly precise) measurements? Or use a realistic geometry and infer an accurate (possibly imprecise) measurement?
- My 2 cents: accept imprecision due to unknown/unconstrained geometry and in the future rely on hierarchical methods to gain as much precision as possible




Outline

- General considerations and motivating questions
  - What are we trying to model?
  - Why are nearby galaxies useful
  - What physical processes to consider
  - What constraints
- Common diagnostics
- Modelling full galaxies, accounting for complexity of ISM and sources
  - Using 1D models
  - Evidence of mixing/smearing issues
  - Using >1D models & n x 1D models
  - Some results and ongoing works
- Statistical framework
  - Samplers
  - Model comparisons / decision tree

Model comparison: evaluating the hypotheses

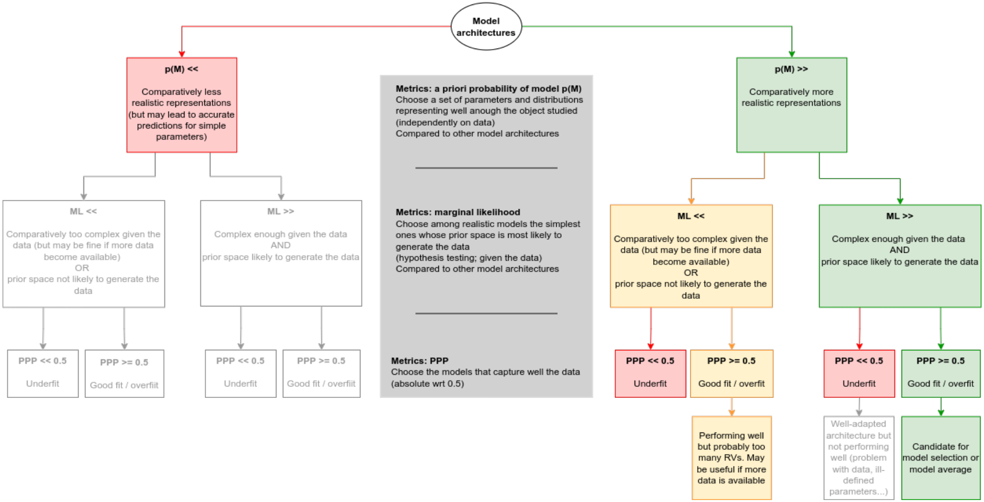
$$\begin{cases} R_{\text{bayes}} = \dfrac{p(\mathcal{M}_1|\vec{O})}{p(\mathcal{M}_2|\vec{O})} \\ R_{\text{bayes}} \propto \dfrac{p(\vec{O}|\mathcal{M}_1)}{p(\vec{O}|\mathcal{M}_2)} \dfrac{p(\mathcal{M}_1)}{p(\mathcal{M}_2)} \end{cases}$$

MARGINAL LIKELIHOOD

PRIOR PROBABILITY OF THE MODEL

- Marginal likelihood: hypothesis testing, i.e., how likely the prior space may generate the data
  - $p(\mathbf{O}|\mathcal{M}) = \int_\theta p(\mathbf{O}|\theta, \mathcal{M}) p(\theta|\mathcal{M}) d\theta$
  - (integration on the whole parameter space of likelihoods × priors on θ)
- Prior probability of the model: how likely the model is, independently on the data
  - i.e., how likely the architecture, choice of parameters, is
  - Quite arbitrary! For instance: a single 1D model has a "low" $p(\mathcal{M})$

Model decision tree



Implications

- A simple architecture (e.g., 2 components sharing the same radiation field) may produce great metrics, even marginal likelihood
  - But single parameter values are fine tuned to match observations, it doesn't make it realistic
- LOC models include thousands of models but linked through very few parameters
  - We may loose in some metrics because the architecture is less flexible than n-component models
  - But we gain in realism (i.e., prior probability of the model)

Outline

- General considerations and motivating questions
  - What are we trying to model?
  - Why are nearby galaxies useful
  - What physical processes to consider
  - What constraints
- Common diagnostics
- Modelling full galaxies, accounting for complexity of ISM and sources
  - Using 1D models
  - Evidence of mixing/smearing issues
  - Using >1D models & n x 1D models
  - Some results and ongoing works
- Statistical framework
  - Samplers
  - Model comparisons / decision tree
- Concluding remarks

Some prospective

Future modeling

- Short term: bridge the gap between models with complex geometries and simulations
  - Static nD models can be combined within complex, parameterized geometries but are not fully self-consistent ⇒ Needs calibration (IFS, simulations. . .)
  - Simulations rapidly improve the chemistry network, photoionization treatment etc. . .
- Long term: probabilistic methods applied to big data
  - Run dynamic simulations on the spot with specific parameters and/or make grids of simulations ⇒ ML techniques
  - Use full static 3D models with parameterized geometries / ingredients
- Dust and gas, orientation of models. . .
- Multi-wavelength resolved observations are essential

End of presentation

Modelling a full galaxy and its suite of lines

Geometry

- It may be apparently simpler to consider a single spectrum but we have to consider that the spectrum is the result of strong selection effects (a simple spectrum doesn't imply a simple model)
- Sources: is the galaxy dominated by a single or a couple of excitation mechanisms (e.g., AGN, HII galaxies) or by a "standard" distribution of HII regions?
- ISM: physical conditions are highly inhomogeneous but follow some physically-motivated distributions
- X-rays and/or PDRs complicate everything and require complex geometries

Model ingredients

- Stellar populations: need to be systematically explored and tested
- Cosmic rays, magnetic field: rely on poorly constrained prescriptions until new tracers become available and/or new knowledge of dependency with environment
- Turbulence: exists in Cloudy for line transfer purposes
- Shocks: high-spectral resolution is key until shocks are self-consistently integrated in photoionization codes

Modelling a full galaxy and its suite of lines

Geometry

- It may be apparently simpler to consider a single spectrum but we have to consider that the spectrum is the result of strong selection effects (a simple spectrum doesn't imply a simple model)
- Sources: is the galaxy dominated by a single or a couple of excitation mechanisms (e.g., AGN, HII galaxies) or by a "standard" distribution of HII regions?
- ISM: physical conditions are highly inhomogeneous but follow some physically-motivated distributions
- X-rays and/or PDRs complicate everything and require complex geometries

Model ingredients

- Stellar populations: need to be systematically explored and tested
- Cosmic rays, magnetic field: rely on poorly constrained prescriptions until new tracers become available and/or new knowledge of dependency with environment
- Turbulence: exists in Cloudy for line transfer purposes
- Shocks: high-spectral resolution is key until shocks are self-consistently integrated in photoionization codes

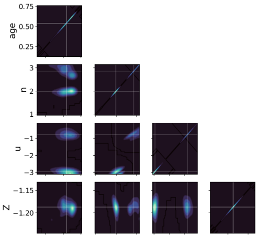
Illustration: MULTIGRIS *(Lebouteiller Ramambason 2022)*

- Grid of predicted fluxes (+ interpolation function)
- Model $\mathcal{M}$ ("architecture"): components, mixing weights, parameters θ, priors...
- Data d: observed emission lines / bands (incl. upper limits)

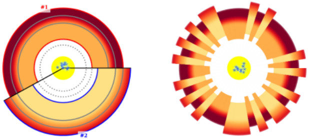
PyMC

- Sampling: draw from the likelihood with a given step algorithm
  - $\mathcal{L} = p(\mathbf{O}|\theta) = \prod_{i=0}^{N} \mathcal{N}(\mu = O_i, \sigma^2 = U_i^2)$

- Use Bayes theorem to obtain posterior probability density functions (PDFs)
  - $p(\theta|\mathbf{O}, \mathcal{M}) \propto \frac{p(\mathbf{O}|\theta,\mathcal{M})p(\theta|\mathcal{M})}{p(\mathbf{O}|\mathcal{M})}$



2D representation of a 3D object modeled with 1D code



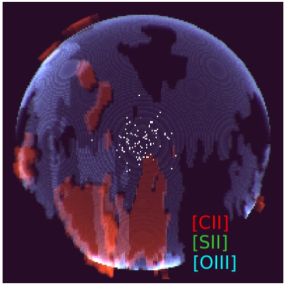
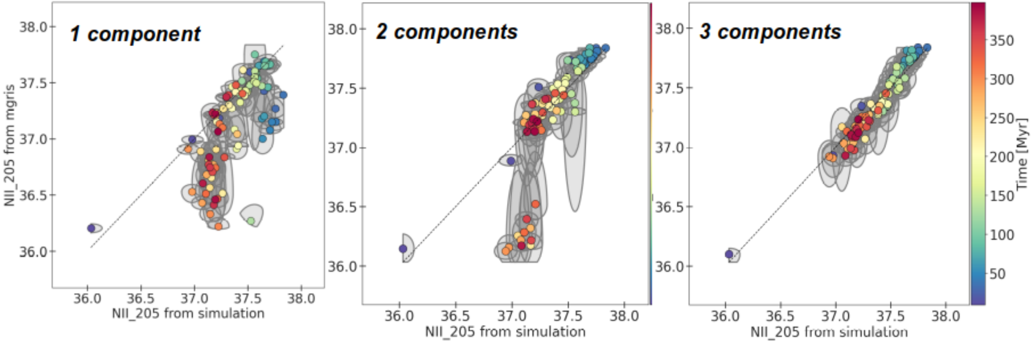


Illustration of ongoing works: comparison with simulations

- (Testing phase!) Calibrating combinations of 1D models to match PRISM simulations. Some lines are particularly difficult to reproduce with a single component.



Why is the probabilistic approach best

- BUT this doesn't mean that more complex (i.e., realistic) geometries shouldn't be explored/evaluated!
  - Even if few tracers used, that doesn't mean the real geometry is simple. It means we don't have enough constraints to constrain a complex geometry $\Rightarrow$ Are the parameters derived from a simple architecture meaningful? Depends on the parameter. . . !
  - For instance, we have only 3 lines (e.g., ALMA high-z) and we wanna know the mass of HI or $f_{\rm esc}$. Is it better to knowingly use an unrealistic geometry and infer inaccurate (possibly precise) measurements? Or use a realistic geometry and infer an accurate (possibly imprecise) measurement?
  - My 2 cents: accept imprecision due to unknown/unconstrained geometry and in the future rely on hierarchical methods to gain as much precision as possible

△ Terminology

- Model: physical nD model (typically RT) usually adapted to given physical object within a galaxy
- Model "architecture": pompous way to refer to a galaxy model, i.e., ways to consider galaxy components, i.e., how do we build a galaxy, i.e., model of models