Diagnosing the ISM of galaxies and energetic processes in a cosmological context

Ramos Padilla, Andres Felipe

DOI:
10.33612/diss.251115545

rijksuniversiteit
groningen

Diagnosing the ISM of galaxies and energetic processes in a cosmological context

Proefschrift

ter verkrijging van de graad van doctor aan de
Rijksuniversiteit Groningen
op gezag van de
rector magnificus prof. dr. C. Wijmenga
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

dinsdag 22 november 2022 om 12:45 uur

door

Andrés Felipe Ramos Padilla

geboren op 17 november 1991
te Bogotá, Colombia

Promotores
Prof. dr. S.C. Trager
Prof. dr. F.F.S van der Tak

Copromotor
Dr. L. Wang

Beoordelingscommissie
Prof. dr. F. Fraternali
Prof. dr. P. van der Werf
Prof. dr. R. Davé

You can't always get what you want
But if you try sometimes, well, you might find
You get what you need

The Rolling Stones - You can't always get what you want

Words are flowing out like endless rain into a paper cup
They slither wildly as they slip away across the Universe

The Beatles - Across the Universe

To my family. I owe them my way of being.
A mi familia. A ellos les debo mi forma de ser.

The research leading to this thesis was funded by SRON, The Netherlands Institute for Space Research

Front and back cover: Design by Andrés Felipe Ramos Padilla of a "space telescope" showing the ISM phases of distant (simulated) galaxies. Simulated galaxy images created with the help of Py-Sphviewer. Modified image from colliding simulated galaxies from NASA/Goddard Space Flight Center; Advanced Visualization Laboratory, National Center for Supercomputing Applications; and B. Robertson, L. Hernquist with recolouring and additional layouts. Original AGN image (NGC 5495) by ESA/Hubble & NASA, J. Greene (R. Colombari) with recolouring. Original image of binary code from Vecteezy.com.

Printed by: ...

Contents

Blackbird singing in the dead of night
Take these broken wings and learn to fly

The Beatles - Blackbird

Caw! Caw! Kraa! for you as well

Andrés looking at a crow

1 Introduction

Our desire to gaze into the dark sky and imagine how big the Universe may be has led us to develop technological advances to take a closer look at some of those bright, fuzzy spots we see on a starry night. Some of those fuzzy spots are galaxies that are far away from us outside our own galaxy, the Milky Way. By observing these distant galaxies we hope to be able to understand some of the more complex processes in the Universe and take our imagination and knowledge to higher levels.

The composition of the gas in galaxies and its physical processes (e.g. thermodynamics) play an important role in our understanding of the formation and evolution of galaxies. Most of our knowledge of these gas processes comes from observations of the Milky Way and its closest companion galaxies. However, not all extragalactic sources form and evolve in the same way as the Milky Way. Therefore, our interpretation of gas processes may be biased towards the local Universe. In this thesis, I focus on the effects that some physical parameters have on the gas properties in a cosmological context.

The interstellar medium (ISM) is the main point of interaction between the different gas components within a galaxy. The ISM contains information about the gas and its environment and can be disentangled in different phases depending on its characteristics. Fine-structure emission lines that trace the cooling and heating of the gas also trace these ISM phases. In this thesis, I diagnose the ISM of galaxies using far-infrared (FIR) emission lines to disentangle the ISM phases and analyse their dependence on other properties of the galaxies.

The energetic output of stars and supermassive black holes causes large differences

in the gas ISM within a galaxy. The effects of intense star formation (SF) activity or the so-called active galactic nuclei (AGN) are crucial to understanding the evolution of galaxies. These kinds of activities help to classify galaxies into different types that can reflect the evolution of gas in a galaxy. In this thesis, I examine the relationship between star-formation rates and fractional AGN contributions in galaxies with different activity types to understand the main-sequence of star-forming galaxies (see. Sect 1.2.3).

This introductory chapter provides the reader with the essential background to understand the contents of this thesis. I start by describing the main physical processes within the ISM. Next, I explain SF in a cosmological context and the different types of galaxies that we currently find in the Universe. In Section 1.4, I introduce the tools used to understand gas processes in galaxies in a cosmological context. Finally, I present an outline of the contents of this thesis.

1.1 Interstellar medium

The interstellar medium (ISM) mainly contains the gas and dust between the stars in a galaxy (besides other components such as cosmic rays and magnetic fields). The ISM plays a crucial role in the evolution of galaxies because it contains information of all previous generations of stars, as well as future generations (Tielens 2010). Dust in the ISM is mainly represented by grains with a range of sizes (0.01–0.1 μm) and temperatures (10–100 K) that emit at infrared (IR) wavelengths both in infrared bands like Polycyclic Aromatic Hydrocarbons (PAHs), and in continuum emission peaking at FIR wavelengths. On the other hand, the gas in the ISM exists in different thermal phases as well as density and ionisation regimes depending on the environmental conditions where they are found. These conditions define the ISM gas phases.

1.1.1 ISM gas phases

Initially, the ISM was defined in two thermally stable phases that coexist in pressure equilibrium: cold ($T< 300$ K) and warm ($T\sim10^4$ K) ISM (Field et al. 1969). Then, due to the effects of expanding supernova remnants (Cox & Smith 1974), it became clear that a third hot phase ($T> 10^5$ K) was needed. In addition to the incorporation of this hot phase, the warm phase was divided into two components: neutral and ionised (McKee & Ostriker 1977). Later works also incorporated molecular clouds as part of the ISM phases, however it is important to note that molecular clouds are self-gravitating (Larson 1981); therefore, they are not in pressure equilibrium with the other ISM phases (McKee 1990). In this thesis, I assume that only the inner cores of molecular clouds (their densest parts) are decouple from the ISM. In parallel, HII regions took an important role to understand the ionised regime around stars (Shields 1990). These HII regions expand towards or near the neutral clouds and affect the interface that defines the interaction between other ISM phases (i.e Photo-dissociation regions (PDRs), Hollenbach & Tielens 1999). To better understand the differences between these ISM phases, I present a schematic representation of the commonly assumed components of the ISM (Tielens 2010, inspired by Figs.1–3 of McKee & Ostriker (1977)) in Fig. 1.1 and describe its main characteristics:

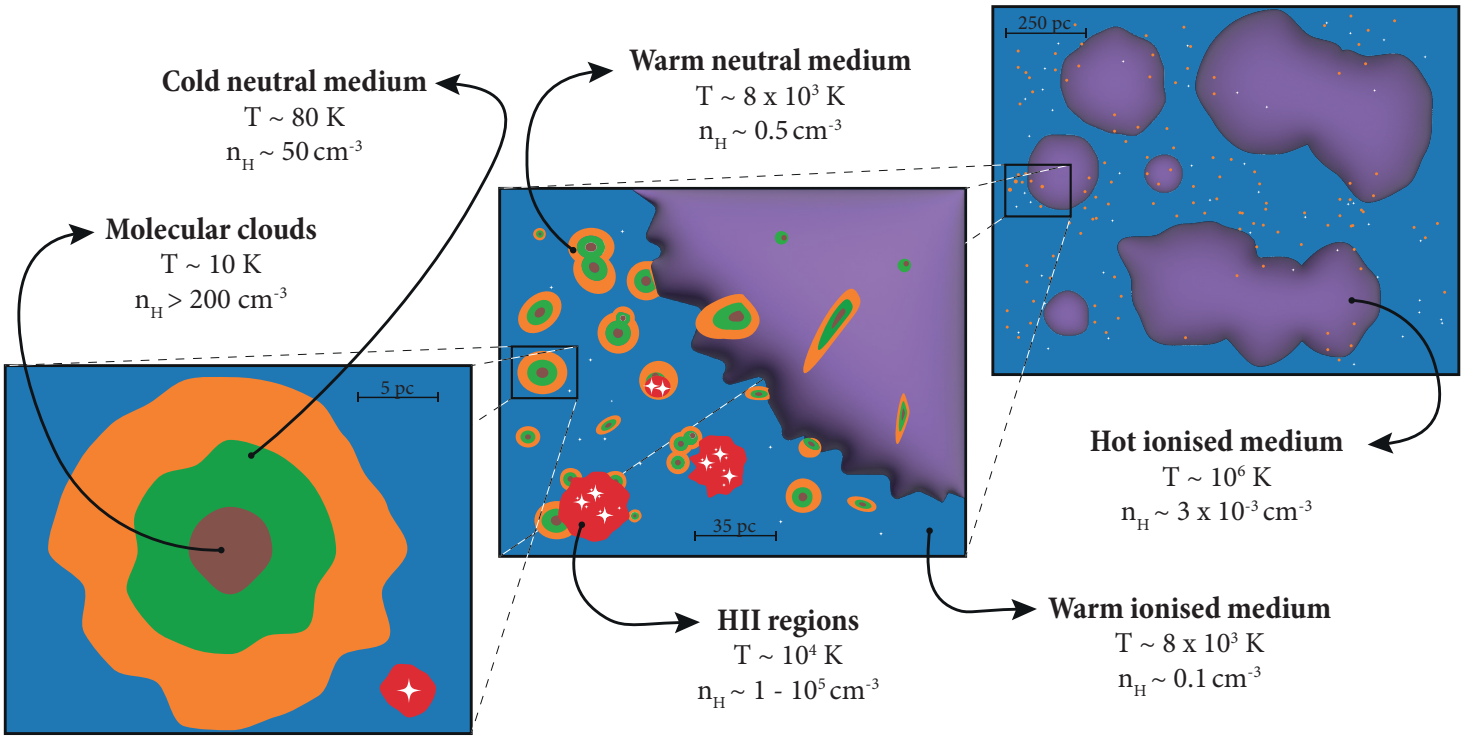

Figure 1.1 – *Schematic representation of the commonly assumed ISM phases in the literature. Inspired by Figs.1–3 of McKee & Ostriker (1977) with physical values from Tielens (2010). Supernova explosions heat the surrounding gas creating the hot ionised medium (HIM, purple) and, together with* HII *regions (red) located around young stars, displace the warm ionised medium (WIM, blue) and destroy or reshape neutral clouds while expanding. The edge of these neutral clouds is defined by the warm ionised medium (WNM, orange) and photo-dissociation regions (PDRs) when changing from ionised (WIM, HIM and* HII *regions) to neutral gas. Deep in the neutral clouds, we find the cold neutral medium (CNM, green) and in the inner core the molecular clouds (brown).*

- Molecular clouds: These are the inner core structures of neutral clouds. These clouds are described as the places where molecular hydrogen (H_2) should exist, but are typically traced by CO molecules at sub-mm wavelengths. Molecular clouds tend to have very low temperatures ($\sim$10 K) with densities high enough ($> 200\,\mathrm{cm}^{-3}$) to be gravitationally bound, and thus they are the birthplace for future generations of stars.

- Cold neutral medium (CNM): This phase of the ISM is in thermal and pressure equilibrium with its surroundings with a temperature of $\sim$80 K and densities of $50\,\mathrm{cm}^{-3}$. It is generally traced by far-infrared [C II] emission and radio HI absorption, and is the main component of neutral clouds.

- Warm neutral medium (WNM): A temperature of $\sim 8 \times 10^3$ K defines this phase of the warm ISM. It normally surrounds neutral clouds with a density of $\sim$0.5 cm^{-3}. As a boundary between regions, it is one of the most important phases of the ISM and is mainly found at the photo-dissociation regions (PDRs). It is traced by the 21 cm emission line and is also known as "Warm intercloud medium" in some cases.

- Warm ionised medium (WIM): This is the ionised component of the warm medium. This phase fills almost half of the volume of the ISM, with temperatures $\sim 8 \times 10^3$ K and densities lower than the WNM ($\sim$0.1 cm^{-3}). It is also known as diffuse ionised gas (DIG) and is traced by [N II] lines in the optical and infrared wavelengths.

- HII regions: These regions are defined by the ionising photon flux coming from O and B (young) stars. These regions have temperatures $\sim 10^4$ K with high densities in the range of 1–$10^5\,\mathrm{cm}^{-3}$, and are typically traced by Hα emission and radio continuum.

- Hot ionised medium (HIM): This ISM phase is created by gas heated via supernova explosions. It has the lowest densities ($\sim 3 \times 10^{-3}\,\mathrm{cm}^{-3}$) and highest temperatures ($\sim 10^6$ K) that any other ISM phase. It can be traced by soft X-ray emission and some absorption lines (e.g. O IV). It is also known as "hot coronal gas" because of its typical location outside of the plane of the Milky Way's disc.

The definition of these phases is only a guide to describe and characterise the different ISM environments. However, it is important to mention that there may be different variations in the definition and characteristic values from these phases (e.g. Abdullah et al. 2017). In this thesis, I use a simple definition of the ISM phases originally used by Olsen et al. (2015). Nevertheless, the physical processes that describe these ISM phases will be very similar.

1.1.2 Heating and cooling mechanisms*

The temperature of the different phases of the ISM depends on the heating and cooling mechanisms that allow them to be in a stable thermal equilibrium. This thermal equilibrium is the balance between the volumetric heating rate g and the

* In this subsection, I describe the mechanisms in a simple way as done by Ryden & Pogge (2021)

volumetric cooling rate l, which will define a heating gain Γ and a cooling loss L for a given single gas particle. Field et al. (1969) found that the heating gain Γ is independent of the gas temperature; therefore, it is possible to define an equilibrium density ($n_{\rm eq}$) in which heating gain and cooling loss balance each other

$$\begin{aligned} \Gamma &= L \\ \Gamma &= n_{\rm eq}(T)\Lambda(T), \end{aligned} \tag{1.1}$$

where $\Lambda(T)$ is defined as the cooling function. In other words, gas particles with gas density n are in thermal equilibrium if:

$$\begin{aligned} g &= l \\ n\Gamma &= nL \\ n\Gamma &= n^2\Lambda(T), \end{aligned} \tag{1.2}$$

Below I will describe the most common physical heating and cooling mechanism in this thermal equilibrium.

Heating

Heating mechanisms generally occur due to high-speed electron collisions that alter the kinetic energy (E) of the gas. Depending on where these electrons come from the heating mechanism will be different.

Cosmic ray heating: Cosmic rays from supernovae tend to have very high energies ($E{\sim}25\,\mathrm{MeV}$) and are mostly protons. When a cosmic ray proton collides with a hydrogen atom, it produces an electron with an average energy of $E{\sim}35\,\mathrm{eV}$,

$$p_{cr} + H \rightarrow p_{cr} + p + e^{-}, \tag{1.3}$$

this electron heats the atoms in its surroundings, leading to a cosmic ray heating gain which can be defined as

$$\Gamma_{cr} \approx 1 \times 10^{-27} \mathrm{erg\,s^{-1}} \left(\frac{\zeta_{cr}}{10^{-16}\mathrm{s^{-1}}} \right), \tag{1.4}$$

where ζ_{cr} is the primary cosmic ray ionisation rate, with a typical value of $\sim 10^{-16}\mathrm{s^{-1}}$ (e.g. Indriolo et al. 2015). This type of mechanism is the main heating contributor inside molecular clouds due to the energy range of cosmic rays.

Photoelectric heating: Some energetic UV photons ($> 5\,\mathrm{eV}$) can collide with small dust grains and, as a result, eject electrons, which is also known as the photoelectric effect. These electrons also heat the surroundings with an estimated heating gain of

$$\Gamma_{pe} \approx 2 \times 10^{-26} \mathrm{erg\,s^{-1}}. \tag{1.5}$$

This type of mechanism is the main heating contributor in the CNM and WNM, and to a small extent in the WIM.

Photoionisation heating: Other photons can ionise carbon atoms when they are in the energy range of $11.26\,\mathrm{eV} < h\nu < 13.60\,\mathrm{eV}$. In this photoionisation,

$$C + h\nu \rightarrow C^{+} + e^{-}, \tag{1.6}$$

the resulting electron has an energy in the range $E < 2.34\,\mathrm{eV}$. Assuming the average energy is 1 eV, the heating gain from the carbon photoionisation is

$$\Gamma_C \approx 5 \times 10^{-29} \mathrm{erg\,s^{-1}} \left(\frac{f_C}{10^{-3}}\right) \left(\frac{\zeta_{h\nu,C}}{3 \times 10^{-10} \mathrm{s^{-1}}}\right), \tag{1.7}$$

where f_C is the fraction of neutral carbon and $\zeta_{h\nu,C}$ is the photoionisation rate of carbon. The phases of the ISM that are heated by this mechanism are mainly the HII regions and the WIM.

Shock heating: When discontinuous perturbations drastically modify the surrounding environment, such as supernova explosions or collisions between molecular gas clouds or even galaxies, shocks are the main source of heating. These shocks add kinetic energy to the random particle motion, allowing temperatures in the ISM to rise. The phase where these shocks are most dominant is the HIM.

Cooling

Cooling is generally radiative through different types of photon emission. Therefore, the emitting phases of the ISM are, in essence, always trying to cool down.

Free-free emission cooling: When a charged particle, usually an electron, approaches another charged particle and is deflected, the moving particle loses energy and emits radiation in the form of photons. This "break radiation", better known as Bremsstrahlung (by the German expression), allows the cooling of ionised gas regions with very high temperatures (T$> 10^6$ K). This is an important cooling mechanism in the HIM and WIM.

Metal cooling: Between T$> 10^4$ K and T$< 10^6$ K we will find the maximum cooling efficiency. However, this maximum will depend on the metallicity of the gas, as different metal lines, crucial to the total cooling function, contribute differently between these temperatures (Gaetz & Salpeter 1983).

Hydrogen excitation line cooling: At temperatures between 10^4 K to 10^5 K, collisions between H atoms and electrons become very important for cooling. The transition from the first excited state of hydrogen (Lyman α, which is highly populated at a temperature of $T = 1.18 \times 10^5$ K) to the ground state ($E = 10.2$ eV) emits photons at 1216 Å which are very effective coolers. The reason for this is that even if a small fraction of H is excited, the large abundance of H is significant enough to compete with other cooling mechanisms. The cooling function for hydrogen is

$$\frac{\Lambda_{\mathrm{Ly}\alpha}}{10^{-27} \mathrm{erg\,cm^3\,s^{-1}}} \approx 6 \times 10^5 \left(\frac{x}{10^{-3}}\right) \left(\frac{T}{10^4\,\mathrm{K}}\right)^{-0.5} \exp\left(-\frac{1.18 \times 10^5\,\mathrm{K}}{T}\right), \tag{1.8}$$

where x is the fractional ionisation of the gas and T is the gas temperature. Hence, this Lyman α cooling is very important for the warm phases of the ISM (WIM and WNM).

Emission line cooling: Finally, the most important cooling mechanism occurs at temperatures below 10^4 K in the FIR wavelength regime. At these temperatures, almost all carbon is singly ionised (C II) and oxygen is neutral (O I). In the case of C II, the upper fine-structure level is populated at a temperature of $T{\sim}91.2$ K and when the electron falls to the lower level, it emits a photon at 158 μm. Then, the cooling function for C II from free electrons is

$$\frac{\Lambda^{e}_{\mathrm{C\,II}}}{10^{-27}\mathrm{erg\,cm^3\,s^{-1}}} \approx 3.1 \left(\frac{x}{10^{-3}}\right) \left(\frac{T}{100\,\mathrm{K}}\right)^{-0.5} \exp\left(-\frac{91.2\,\mathrm{K}}{T}\right), \tag{1.9}$$

when x is very low, the cooling function for C II comes from collisions with hydrogen atoms

$$\frac{\Lambda^{H}_{\mathrm{C\,II}}}{10^{-27}\mathrm{erg\,cm^3\,s^{-1}}} \approx 5.2 \left(\frac{T}{100\,\mathrm{K}}\right)^{0.13} \exp\left(-\frac{91.2\,\mathrm{K}}{T}\right). \tag{1.10}$$

Therefore, the cooling function for C II is $\Lambda_{\mathrm{C\,II}} = \Lambda^{e}_{\mathrm{C\,II}} + \Lambda^{H}_{\mathrm{C\,II}}$.

In the case of O I, the upper fine-structure level is populated at a temperature of $T{\sim}228$ K and when the electron falls to the lower level, it emits a photon at 63 μm. Then, the cooling function for O I is

$$\frac{\Lambda_{\mathrm{O\,I}}}{10^{-27}\mathrm{erg\,cm^3\,s^{-1}}} \approx 4.1 \left(\frac{T}{100\,\mathrm{K}}\right)^{0.42} \exp\left(-\frac{228\,\mathrm{K}}{T}\right). \tag{1.11}$$

These two cooling functions are very important for the cold phases of the ISM. $\Lambda_{\mathrm{C\,II}}$ will be dominant in the molecular clouds and most of the CNM, while $\Lambda_{\mathrm{O\,I}}$ will be dominant in the CNM up to the transition with the WNM.

To complement the explanation of these cooling processes and the total cooling function (i.e. $\Lambda_{\mathrm{Total}} = \Lambda_{\mathrm{Ly}\alpha} + \Lambda_{\mathrm{C\,II}} + \Lambda_{\mathrm{O\,I}}$), in Fig. 1.2 I show how the cooling functions change with respect to temperature and fractional ionisation of the gas. At temperatures below 500 K the $\Lambda_{\mathrm{C\,II}}$ is the most important cooling function, and depending on fractional ionisation, C II cooling coming from collisions with electrons or hydrogen atoms will be the most important. At values around 1000 K the $\Lambda_{\mathrm{O\,I}}$ takes over and then at $T{\sim}10^4$ K, $\Lambda_{\mathrm{Ly}\alpha}$ increases dramatically due to the abundance of hydrogen which also depends on x.

1.1.3 IR emission lines

Some of the cooling processes mentioned above can be traced by FIR line emissions. The most important FIR emission lines are [C II] at 158 μm and [O I] at 63 μm. However, in the IR there are other important lines that trace different phases of the ISM and at the same time give some insight into specific physical parameters, such as gas reservoirs and AGN activity. An example of some of these lines are the CO, [N II] and [O III] lines, which trace the molecular clouds, low-excitation ionised gas

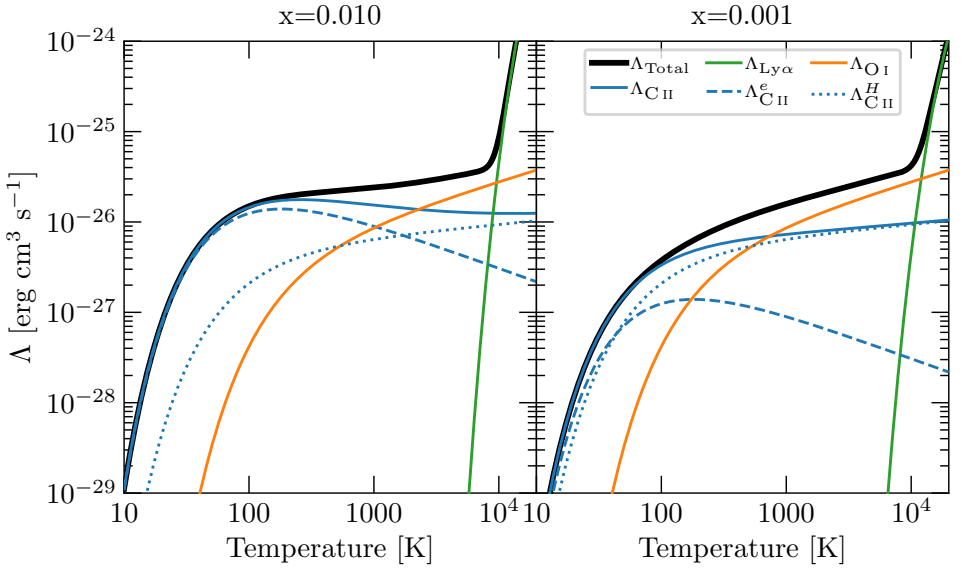

Figure 1.2 – *Cooling functions in the ISM for a high (left) and a low (right) value of the fractional ionisation, as described in Equations 1.8–1.11. Figure adapted from (Ryden & Pogge 2021).*

and high-excitation ionised gas, respectively (Lagache 2018; van der Tak et al. 2018, e.g.). In this thesis, I do not focus on molecular clouds, so I do not take into account the CO emission. I describe the eight most important FIR emission lines in detail in Chapter 3.

In addition to these FIR lines, at shorter wavelengths in the mid-IR (MIR), there are a couple of lines that can help trace not only the star-forming regions (HII regions) but also AGN activity. This AGN activity can be very important since it can heat the gas, breaking the thermal balance of the ISM phases (e.g. Herrera-Camus et al. 2018b). The set of Ne lines found in the 5–30 μm range has the advantage that SF and AGN processes can be observed and traced simultaneously from some space telescopes (Sturm et al. 2002; Ho & Keto 2007; Abel & Satyapal 2008). In this thesis, I use two of these lines [Ne II] and [Ne V], which I describe in detail in Chapter 5. Other lines such as [O IV] at 25.9 μm could also trace the AGN, but may also come from SF (e.g. Goulding & Alexander 2009; Spinoglio et al. 2017). In addition to these lines, in the MIR there are several IR features coming from PAHs that are useful for tracing ionisation states and grain sizes of the dust. These emissions are crucial in many studies in the Milky Way and the local Universe, where it is possible to observe them in detail. In this thesis, PAH features are only described in the context of the spectral energy distribution (SEDs, see Section 1.4.3) in Chapters 5 and 6. In Table 1.1, I mention the ISM phases that these MIR emissions can trace to highlight their relevance in ISM studies. I include other Ne lines along with their ionisation potentials and wavelengths.

The lines discussed in this section trace different phases of the ISM. These emission

Line Species	Wavelengths [μm]	IP [eV]	Main ISM phases
Mid-infrared emission lines			
[Ne VI]	7.65	126.21	Violent ISM[a]
[Ne II]	12.81	21.56	HII regions
[Ne V]	14.32 & 24.32	97.12	Violent ISM[a]
[Ne III]	15.55	40.96	WIM and HII regions
[O IV]	25.91	54.93	HII regions and violent ISM[a]
Far-infrared emission lines			
[O III]	51.81 & 88.36	35.12	WIM and HII regions
[N III]	57.34	29.60	WIM and HII regions
[O I]	63.18 & 145.53	···	CNM and WNM
[N II]	121.80 & 205.30	14.53	WIM and HII regions
[C II]	157.68	11.26	CNM, WNM, WIM and HII regions

Table 1.1 – *IR emission lines discussed in this thesis. Wavelength information and ionisation potential (IP) comes from Kramida et al. (2020) and Spinoglio et al. (2017).*

Notes: [a] Here we assign some of the IR emission lines to the violent ISM because these lines are present when the ISM is disrupted and equilibrium is not easily reachable (Tielens 2010). Generally, emission lines like [Ne V] can come from supernovae (shocks), Wolf-Rayet stars, or the narrow-line region (NLR, described in Sect. 1.3.2) of AGNs (e.g. Abel & Satyapal 2008).

lines cover a wide range of physical processes that will help us to decouple the ISM phases from the lines and are crucial to understand the heating and cooling processes of the ISM. These emission lines are also useful to examine the relationship between AGN and SF. Therefore, we can understand part of the physical processes of the star-formation cycle: from the cooling of the gas in molecular clouds and in the CNM, to the heating of gas in the surroundings of young stars (HII regions) and AGN, and the intermediate thermal and ionisation processes in the WIM and WNM.

1.2 Star formation in a cosmological context

The star-formation cycle is comprised of different stages (McKee & Ostriker 2007): First, the diffuse ISM phases condense into giant molecular clouds (GMC, also known as neutral clouds) due to gravitational instabilities of the gas. Then, the gas stratifies in temperature and ionisation gradients, where the inner core is very cold (CNM) and the outskirts are still warm (WNM). Molecular clouds form in the inner core, and when the turbulence and/or magnetic fields are very low (e.g. André et al. 2014; Ching et al. 2022), it allows the gas cloud to collapse to form a protostar. Some of these protostars evolve into bright young stars, where their surroundings become ionised via HII regions. These HII regions expand to the point where turbulence affects the remnants of the GMC by adding more turbulence to the gas. This process also occurs when the stars explode in supernovae, adding even more turbulence to the

gas. Both processes affect ISM regions that are generally warm (WIM and WNM). The gas then returns to its diffuse phase, allowing hot ionised gas to escape to hot regions (HIM). Finally, the cycle is complete when this diffuse gas begins to condense back into GMCs.

The topic I present in this section is how we track this star formation process and what we know about it in the context of the main-sequence of star-forming galaxies.

1.2.1 SFR tracers

One of the main problems to trace SF in the Universe is that the relevant timescales are orders of magnitude larger than the human lifetimes. We typically look at different galaxy populations at different cosmic epochs, rather than directly observing how SF in galaxies changes over time. In other words, we are only looking at a quasi-static snapshot of the Universe. On top of that, we still cannot resolve the stellar information that comes from outside our Milky Way and our closest neighbours. Fortunately, we have found a way to estimate how many stars are formed each year in galaxies away from us.

The first steps in understanding the total star-formation rate (SFR) in a galaxy were made by Schmidt (1959), who proposed a power-law relationship between the SFR density and the density of interstellar gas:

$$\rho_{\rm SFR} = \rho_{\rm gas}^{n}, \tag{1.12}$$

where n is the index (slope) of the power law. However, the tools to calibrate and estimate star formation from young stellar environments in galaxies were not ideal. Half a century ago, we were able to trace the SFR assuming synthesis stellar population models and translating them into optical colours, such as $U-V$ and $B-V$ (Tinsley 1968, 1972). A decade later, it was possible to obtain the SFR of galaxies using continuum fluxes in the IR and ultraviolet (UV) and integrated emission lines (e.g. Kennicutt 1998a, and references therein). This allows astronomers to define the Kennicutt–Schmidt law in terms of the surface densities ($\Sigma_{\rm SFR}$ and $\Sigma_{\rm gas}$) and calibrate a similar index value of Equation 1.12 to $N = 1.4$ (e.g. Kennicutt 1998b; Bacchini et al. 2019).

With the advent of multi-wavelength surveys, it is possible to estimate the relationship between different SFR tracers at different wavelengths using photometric and spectroscopic information. However, all SFR tracers depend on assumptions we make about the initial mass function (IMF). For example, the shape of the IMF can be described by different power-law functions (Salpeter 1955; Kroupa 2001; Chabrier 2003). In addition, it is also debated whether the IMF is universal (e.g. Cappellari et al. 2012; Smith 2020). Kennicutt & Evans (2012) show a list of the most commonly used SFR luminosity tracers: FUV, NUV, Hα, TIR, 24 μm, 70 μm, 1.4 GHz and 2–10 keV. Each of them has its own limitations. For example, some of them are affected by dust attenuation and sensitivity to certain ages in stellar populations. In recent years, other SFR tracers gained more interest to try to solve these limitations in other wavelength ranges as those noted above, such as the SFR calibrations of IR line emissions or low-frequency radio bands (e.g. [C II] and 150 MHz, De Looze et al. 2014; Wang et al. 2019a, respectively).

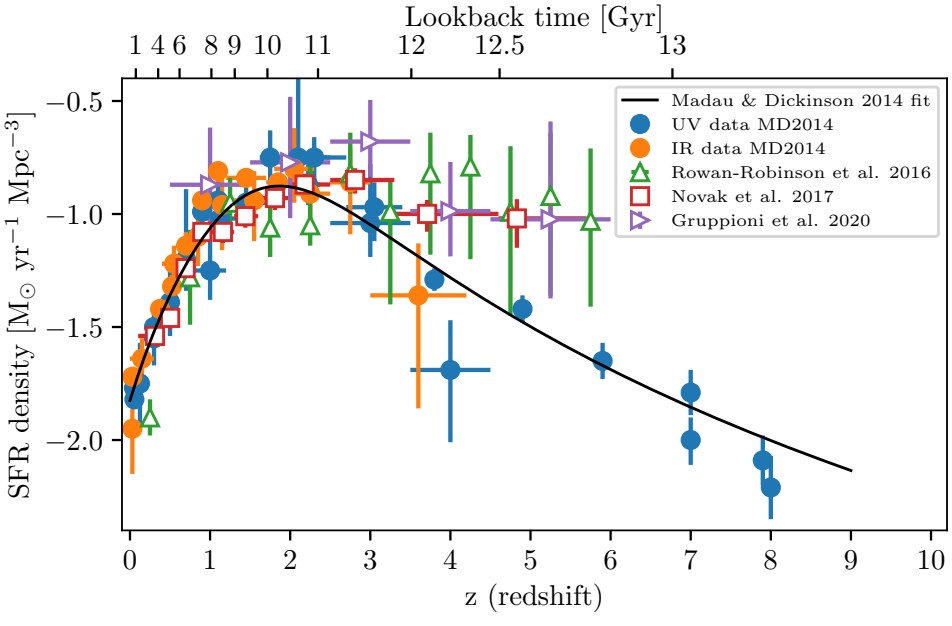

Figure 1.3 – *Star-formation rate density versus redshift with data from the compilation sample of Madau & Dickinson (2014). UV data points have been corrected for dust attenuation, while IR data hava not. I include data from See Madau & Dickinson (2014) Table 1 for the full list of references. I include data from Rowan-Robinson et al. (2016), Novak et al. (2017) and Gruppioni et al. (2020) to compare the discrepancy of the results in new IR surveys.*

Nonetheless, the knowledge gained from these SFR tracers allows us to understand how the SFR can be calculated at different cosmic times and, at the same time, to see how the SFR evolves with redshift.

1.2.2 SFR density evolution

For over a decade, the star-formation rate density (SFRD) has been shown to change with redshift, as reviewed by Madau & Dickinson (2014). They compiled different UV and IR surveys, corrected the UV data for dust attenuation, and fitted the data to show how SFRDs evolve in the Universe, as shown in Figure 1.3. From this figure, it was possible to observe that the SFRD increased from $z{\sim}8$ and reached its maximum point about 10 Gyr ago (near $z = 2$). Then, in the last 10 Gyr, the SFRD continuously decreased to its current value. This period of time where the SFRD peaked is also known as “cosmic noon”.

Since then, several studies have confirmed these findings and even cosmological simulations have been able to reproduce the SFRD (e.g. Davé et al. 2016; Pillepich et al. 2018). However, there are still some doubts about the behaviour of the SFRD at $z > 3$ due to observational constraints. New IR and radio studies have shown that the SFRD could be flat at $z > 3$ (e.g. Rowan-Robinson et al. 2016; Novak et al. 2017; Gruppioni et al. 2020), which disagrees with UV surveys even when correcting for

dust attenuation, as also show in Fig.1.3. This discrepancy with UV surveys may reside in the dust-obscured nature of galaxies (UV/optical dark galaxies) and/or the difficulties of UV surveys in estimating the contribution of star formation in highly embedded systems (Rowan-Robinson et al. 2016; Gruppioni et al. 2020). Although the answer is not yet clear, it is possible to define a transition point around $z = 4$–5 from unobscured to obscured SFRD estimations (Bouwens et al. 2020).

Fortunately, the SRFD is not the only empirical relationship we have to understand star formation. Thanks to the multi-wavelength studies and surveys of the last twenty years, it was possible to obtain a scaling relation between the stellar mass and the SFR of galaxies, also called the "main-sequence of star-forming galaxies".

1.2.3 The main-sequence of star-forming galaxies

The Sloan Digital Sky Survey (SDSS, York et al. 2000) opened the doors to study large samples of galaxies in a homogeneous way. Using SDSS data, Kauffmann et al. (2003b) found a bimodal distribution of galaxies in terms of stellar masses ($M_\star$) that separated the young and old stellar populations inside galaxies. Then, Brinchmann et al. (2004) found that these kinds of galaxies were also divided in terms of their specific star-formation rate ($sSFR = SFR/M_\star$). This bimodal distribution of galaxies was later described as "blue cloud" (young galaxies with high sSFR) and "red sequence" (old galaxies with low sSFR). From these results, it was clear that most of the star-forming galaxies (blue cloud) followed a clear (relatively narrow) relationship between the SFR and $M_\star$. This correlation was studied also in galaxies at higher redshifts ($z < 2.5$) in the following years (Noeske et al. 2007; Elbaz et al. 2007; Daddi et al. 2007) finding the same narrow relation, which was then called the "main-sequence of star-forming galaxies" (MS hereafter).

The properties of the galaxies that are on the MS relation and the MS shape were the subject of debate thereafter. Sample selection can influence the shape of the MS (Whitaker et al. 2012). Even though the MS relation between galaxies holds at different redshifts, the slope of the relation may change with redshift (Whitaker et al. 2014). In addition, many studies have reported a flattening of the MS towards high stellar masses. Therefore, the MS relation could be assumed to be linear or sublinear depending on the properties of the analysed galaxies (Pearson et al. 2018). For example, Speagle et al. (2014) compiled different MS literature samples to observe the redshift dependency of the MS. They assumed a simple power-law relation for all stellar masses to fit the evolution with redshift in the way (their Equation 28):

$$\log(\mathrm{SFR}) = ((0.84 - 0.026t)\log(\mathrm{M}_\star)) - (6.51 - (0.11t)) \tag{1.13}$$

where t is the age of the Universe in Gyr. However, other studies found that using a turn-over in the stellar mass improves the fit of the MS shape (Tomczak et al. 2016). Recent works support the idea of adding a mass turn-over, as this may indicate the morphology dependence of the galaxies in the MS (Leslie et al. 2020), and may give some insights into the role that the AGN can play in quenching the SFR (Leslie et al.

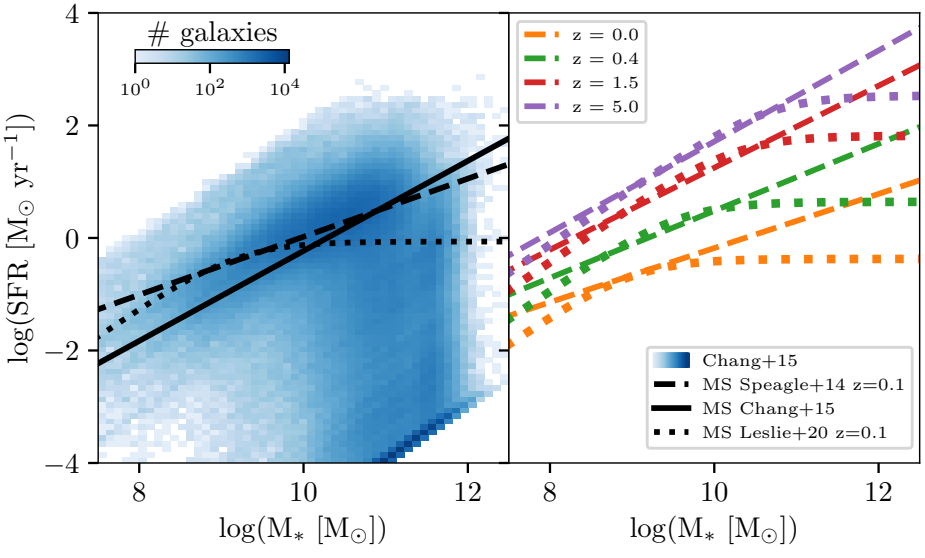

Figure 1.4 – *Main-sequence of star-forming galaxies for the sample of Chang et al. (2015) and the typical relations used to fit the data (Speagle et al. 2014; Leslie et al. 2020). The left panel shows the comparison at $z = 0.1$ of the observational data with the fitting functions, while the right panel shows the derived relation from Speagle et al. (2014) and Leslie et al. (2020) for different redshifts.*

2016). This turn-over mass can be described as (Equation 6, Leslie et al. 2020):

$$\begin{aligned} \mathrm{M_{turn}} &= \mathrm{M_0} - \alpha t \\ \log(\mathrm{SFR}) &= \mathrm{S_0} - \beta t - \log\left(1 + \left(\frac{10^{\mathrm{M_{turn}}}}{10^{M}}\right)\right) \end{aligned} \tag{1.14}$$

where $\mathrm{S_0} = 2.8$, $\mathrm{M_0} = 10.8$, and $\alpha = 0.23$ and $\beta = 0.13$ are the parameterisation constants for normalisation, mass and time respectively, with M the stellar mass in $\log(\mathrm{M_\odot})$ units.

In the left panel of Figure 1.4, I show observational data from a sample of almost a million galaxies with a mean redshift of $z = 0.1$ for which SFR and $\mathrm{M_\star}$ were calculated using SED modelling (Chang et al. 2015). I compare these data and their assumed MS with two common fits for the MS shape as described in Equations 1.13 and 1.14. This comparison shows that either of these fits could describe the observational data for the star-forming galaxies (region at the top of the 2d-histogram); however, this is not true for cases where quiescent galaxies (region at the bottom of the 2d-histogram) are present (e.g. Renzini & Peng 2015). In the right panel, I show the differences at different redshifts (right panel) of the two fitting functions. At almost all redshifts the agreement is good between Speagle et al. (2014) and Leslie et al. (2020) in $\log(\mathrm{M_\star}[\mathrm{M_\odot}]) < 10$. At $\log(\mathrm{M_\star}[\mathrm{M_\odot}])\sim12$ the differences can be $\sim$1 dex. Therefore, these differences may play a crucial role in the interpretation of the results for high-z galaxies. A more detailed explanation of the MS and its effects on understating star-formation in galaxies can be found in Förster Schreiber & Wuyts (2020) and Tacconi et al. (2020).

Now that we know how important the selection of galaxy samples can be for interpreting star-formation in the cosmological context, we can focus on the different types or classes of galaxies that we can find when studying star-formation at the ISM.

1.3 Classification of active galaxies

The first classifications of galaxies were based on their morphology and luminosity (e.g. Reynolds 1920; Hubble 1926). These efforts were the first steps to the well-known description of galaxies depending on their apparent shape and probable evolution between these types (Hubble 1936; de Vaucouleurs 1959). At the same time, it was also possible to classify galaxies according to their colours (de Vaucouleurs 1961) which modify these galaxy classifications continuously (see Sandage 2005, for a historical description). Later, these classifications changed from qualitative to more quantitative terms, which facilitated the analysis on large samples of galaxies (de Vaucouleurs 1977).

Spectral information from the stars in the galaxies was also used to classify galaxies (Morgan & Mayall 1957). This opened the door to the optical spectral classification of galaxies which was excellent in differentiating the main physical properties of galaxies (Baldwin et al. 1981), also known as the BPT diagram. In the BPT diagram is possible to classify galaxies based on four optical line ratios [O III]/Hβ, [N II]/Hα, [S II]/Hα and [O I]/Hα. These ratios divide galaxies into star-forming galaxies (SFG), AGN (Seyfert and LINERs) and composite (SFG and AGN) (Kewley et al. 2006). In this section, I describe these types of galaxies that are important for the content of this thesis. Although there are other ways to classify galaxies (e.g. Cid Fernandes et al. 2011; Kormendy & Bender 2012), I use this general description in the probable zoo of galaxy types (see Buta et al. 2015, for an example of the different types of galaxies only in morphological terms).

1.3.1 Star-forming and starburst galaxies

Star-forming galaxies (SFG) are galaxies that are actively forming stars. As I described in the previous section, these galaxies tend to be young and have blue colours with high sSFR (Brinchmann et al. 2004). These galaxies are the backbone of the MS. However, there is a set of SFG that have extreme SFRs, these galaxies are known as "starburst".

There are many ways to classify a galaxy as a starburst. Some types of starburst are classified as such due to the dominant emission from young stars such as the Wolf–Rayet and HII regions (Telles & Terlevich 1995; Leitherer et al. 1996). Other types of starburst are morphologically small and exhibit a relatively high SFR, such as the blue compact dwarf and green pea galaxies (Cardamone et al. 2009; Cairós et al. 2010). Alternatively, the easiest way to classify a galaxy as a starburst is to set a SFR value above which galaxies can be considered as extreme SFG (Rowan-Robinson et al. 2018). This way can present an advantage for infrared sources since it is possible to associate the IR luminosity in terms of SFR (Kennicutt 1998a; Kennicutt & Evans 2012). In that case, some of the extreme cases of luminous infrared galaxies (LIRGs), ultra-luminous and hyper-luminous infrared galaxies (ULIRGs and HyLIRGs, respectively)

can be considered as starburst. However, these IR sources may have a luminosity that is contaminated by the AGN (Brandl et al. 2006) or only found in the nuclear regions (Brandl et al. 2012).

Nowadays, the best way to identify a starburst galaxy is to use the MS as a reference point for SFG and then, depending on the distance to the MS, classify them as starburst. In other words, the starburst will be those who have the highest sSFR of the galaxies, typically falling more than an order of magnitude above the MS (Herrera-Camus et al. 2018b). However, there is evidence that some starburst could also be located inside the typical scatter of the MS (Elbaz et al. 2018), which may not be related with the self-regulated evolution of SFG galaxies (Tacchella et al. 2016).

1.3.2 Active galactic nuclei (AGN)

Galaxies classified as Active galactic nuclei (AGN) are a type of galaxy characterised by an extreme luminosity coming from the centre of the galaxy, which cannot be explained by the presence of stars (Peterson 1997). This extreme luminosity comes from accretion processes surrounding the supermassive black hole (SMBH) in the centre of the galaxies. Depending on the orientation with respect to the line of sight, galaxies are classified into different types, also called the AGN unified model (Antonucci 1993; Urry & Padovani 1995).

In Fig. 1.5, we show a diagram that represents our current understanding of the AGN unified model. In it, the SMBH is surrounded by an accretion disc, a hot X-ray corona and a broad line region (BLR). The BLR gets its name from the gas that revolves around the SMBH causing a Doppler shift in the emission lines, making them appear broad. Farther from the SMBH, a dusty torus surrounds these internal AGN structures and blocks their emission depending on the line of sight. Finally, on the outskirts, the narrow-line region (NLR) shows gas that is unaffected by the high velocities around the accretion disc and BLR. In some cases, some AGN galaxies tend to have a bipolar synchrotron jet, especially radio-loud galaxies.

Depending on the line of sight and the emission coming from different wavelengths, AGN galaxies will be classified into different types. Low and high excitation radio galaxies (LERG/HERG) include AGN galaxies with a bipolar jet and are divided into different types: narrow-line radio galaxy (NLRG), broad-line radio galaxy (BLRG), steep spectrum radio quasar (SSRQ) and Blazars (BL Lac, flat spectrum radio quasar (FSRQ), and optically violent variables (OVV)). Other AGN types are the quasi-stellar objects (QSO or Quasars, depending on their radio emission) where its main characteristic is the high accretion rates (or electromagnetic (EM) power). Finally, those galaxies that have low levels of radio emission and low accretion rates are known as Seyferts. Seyferts (as QSO) can be divided into Type I or Type II depending on whether the torus blocks the emission from the BLR. This is why Seyfert 2 (Seyfert Type II) galaxies are also classified as narrow emission-line galaxies (NELG).

The low-ionisation nuclear emission-line regions galaxies (LINERS, Heckman 1980), are commonly assumed as a type of AGN galaxies. However, they cannot be explained by the AGN unification model and are sometimes referred to as Seyfert 3 galaxies (e.g. Véron-Cetty & Véron 2010). LINERs tend to have properties of red sequence galaxies,

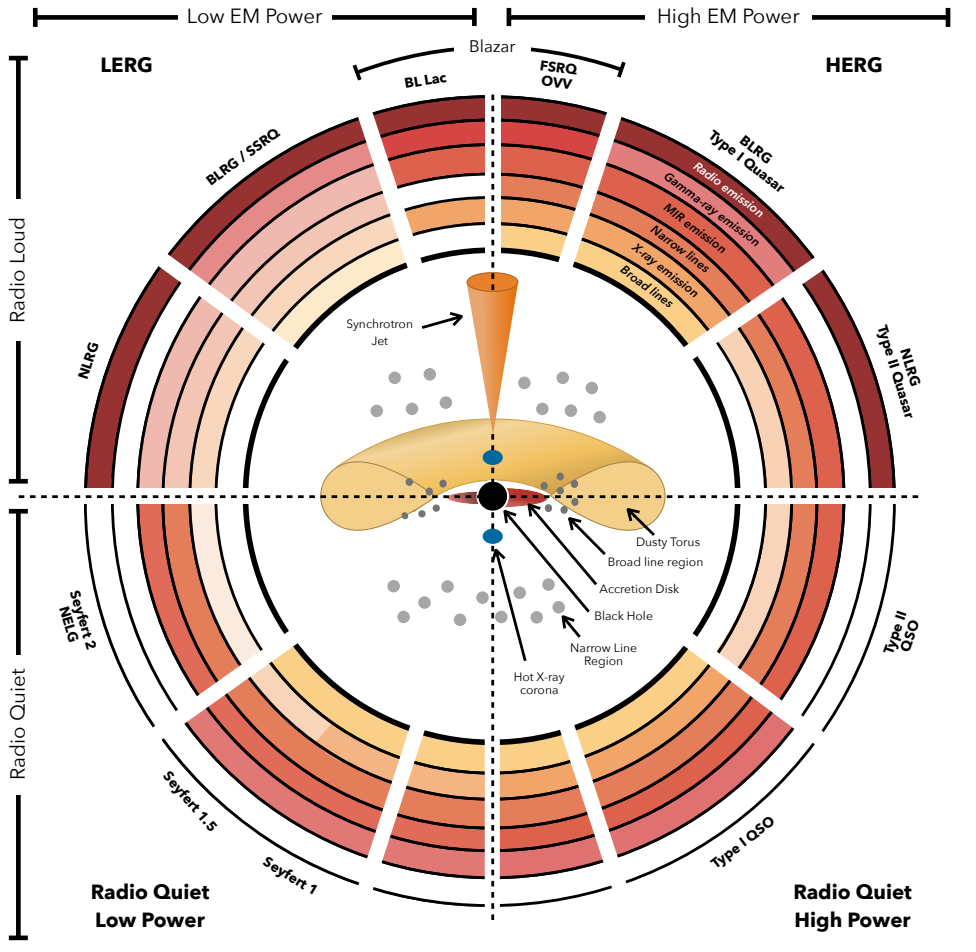

Figure 1.5 – *AGN unification diagram from Thorne et al. (2022). In the centre of the diagram, we find the structures that describe the AGN: the supermassive black hole (SMBH), the accretion disc, the corona, the broad- and narrow-line regions (BLR and NLR, respectively), and a dusty torus. Some of these galaxies tend to have a bipolar synchrotron jet that is visible in Radio Loud galaxies. The concentric rings show the emission at different wavelengths for the AGN types that exist depending on the line of sight. The transparency of the colours shows how strong the emission is for each type of galaxy. Finally, these types can be divided into quadrants in terms of the presence of the jet and its SMBH accretion rate (or electromagnetic (EM) power).*

with an old stellar population in massive galaxies and low SFRs (Singh et al. 2013). These galaxies may help to trace the warm gas, but it appears that they are not related with the AGN (Yan & Blanton 2012).

The AGN unified model has been the subject of debate due to the diversity of types, which are affected by selection bias, observational capabilities and internal variability in these galaxies (see Padovani et al. 2017, for a general overview of AGN at different wavelengths). However, recent results with high-resolution images seem to confirm the foundations of the unified model (Gámez Rosas et al. 2022). Even assuming the

unified model is the most realistic AGN scenario, a remaining open question is what role does the AGN play in the context of the MS and what influence does it have on the estimations of SFR and other ISM properties.

1.3.3 Composite SF-AGN galaxies

There are some galaxies that have both SFG and AGN properties. These galaxies may be intermediate steps between two evolutionary stages of galaxies and are identifiable on the BPT diagram using the [O III]/Hβ and [N II]/Hα ratios. The reason for using that diagram is that AGN contributions become important for $\log([\mathrm{N\,II}]/\mathrm{H}\alpha) > -0.5$, then this ratio is sensitive to low AGN contributions (Kewley et al. 2006).

These composite SF–AGN galaxies may result from interactions between galaxies, also known as mergers. These mergers are believed to be the triggering mechanism for ULIRGs (Sanders & Mirabel 1996). When gas inflows during a merging scenario, both star formation and nuclear activity are triggered (Di Matteo et al. 2005). This generates a cyclical evolutionary model in which mergers trigger both starburst and AGN as presented by Hopkins et al. (2006, see also Fig. 1.6). A piece of evidence comes from the results of Yuan et al. (2010). Yuan et al. (2010) used a sample of ULIRGs that were classified according to the BPT diagram of Kewley et al. (2006) into SFG, composites, and AGN, and into merging stages according to their nuclear separation and morphology. Their results show that most of the ULIRGs were composite SF–AGN galaxies. These composite galaxies tend to have an intermediate merger stage compared to SFG and AGN, which describe the early or late merging stages, respectively. These trends in SFG have also been observed in other works (Luo et al. 2014; Cibinel et al. 2019). Although there has been a long list of studies for or against Hopkins et al. (2006) model (see Ellison 2019, and the list therein), observational results at the local Universe tend to support it (Ellison et al. 2019).

To study those composite/merger galaxies, it will be ideal to separate the SF and AGN components of galaxies. This can be possible with photo-ionisation models (e.g. Kewley et al. 2019, and reference therein) or the distance from the SF branch in the BPT diagrams (Yuan et al. 2010). However, a more reliable method seems to be the use of Spectral Energy distributions (SED, see Sect. 1.4.3) to separate the AGN and the SF contributions (Ciesla et al. 2015). In Fig. 1.6 we show an example of how SED fitting tools can recover AGN fractional contributions to the IR and SFR from merger simulations. These simulations, originally presented by Lanz et al. (2014), show how the AGN and SF processes change with different interaction stages. The SED derived parameters agree with the simulations, although the AGN fractional contribution is not good below 15% (Dietrich et al. 2018), even with the new versions of the SED tools. What is most interesting is to see how SFR peaks a few Myr before the AGN fractional contribution, as expected from the Hopkins et al. (2006) model.

The problem with using these methods to understand the composite SF-AGN mergers galaxies is that it is impossible to have the detailed timescales of evolution in observations as in simulations. Furthermore, it is difficult to recover the star-formation histories (SFH) on these timescales, because it may be that galaxies lose the memory of their previous SFH in only 200 Myr (Caplar & Tacchella 2019). In addition, classifying interacting galaxies can be very complicated as it will depend mainly on

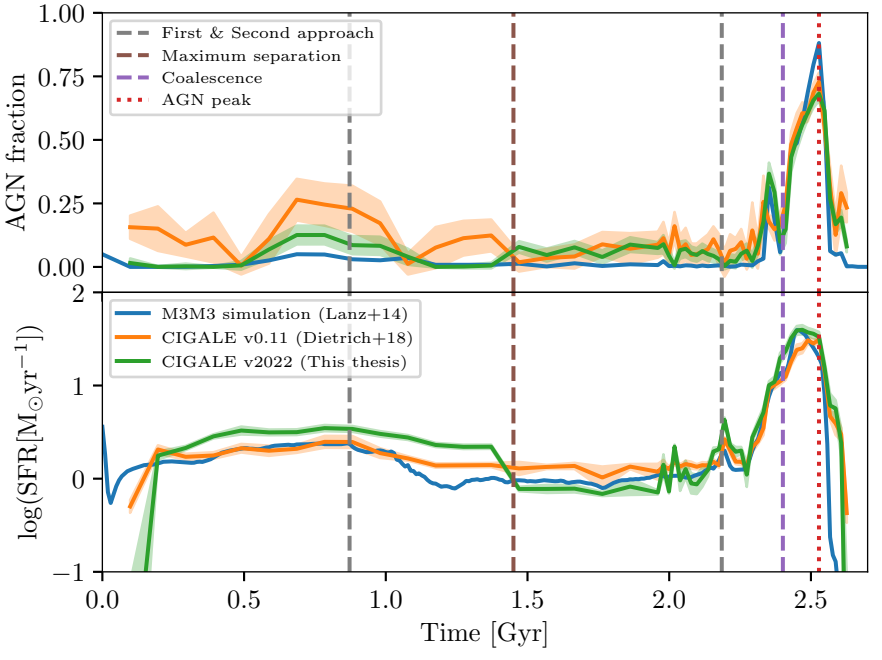

Figure 1.6 – *AGN fraction and SFR in a merger simulation (M3M3) from Lanz et al. (2014) compared with results from Dietrich et al. (2018) and a simple re-run with a new version of* CIGALE *(see Sect. 1.4.3). From these estimates is possible to see how SFR peaks a few Myr before the AGN.*

their morphological disturbances and not on the actual timescales. However, morphological disturbances can be used to identify mergers and reconstruct their SFH with the help of citizen science and computational advances in machine learning methods (Holincheck et al. 2016; Pearson et al. 2019a,b).

1.4 Tools to understand galaxy formation and evolution

Nowadays, there are different tools to understand how galaxies form and evolve. Most of these tools are computationally expensive, as they require us to take into account most of the theoretical knowledge we have about galaxies. Nonetheless, advances in different methods and techniques to apply physical assumptions in these tools are allowing us to get closer to the real scenarios expected in the Universe. This section describes the most important tools I use in this thesis.

1.4.1 Cosmological simulations

Even though discussion are ongoing on the physical nature of dark matter (e.g. Milgrom & Sanders 2003; Kroupa 2012; Chae et al. 2020), the currently accepted paradigm to understand our Universe comes from the Λ Cold Dark Matter model

(ΛCDM) in which dark energy, dark matter and baryonic matter build most of the Universe (Frenk & White 2012). Therefore, most cosmological simulations assume a ΛCDM Universe, where dark matter is the backbone of galaxy formation (Dayal & Ferrara 2018).

There are two kinds of ΛCDM cosmological simulations, Semi-analytical models (SAMs) and numerical simulations. In this thesis, I focus on the numerical type, since both types have reached a point where their qualitative results are very similar (Somerville & Davé 2015). Numerical simulations can be created mainly using two hydrodynamic methods: Lagrangian and Eulerian. In Lagrange's method, a set of particles carries the physical information (e.g. temperature, mass or metallicity) of the simulated environment (e.g. gas or stars) according to a weighted kernel that smooths the physical properties. Due to this "smoothing", this method is mainly known as smoothed particle hydrodynamics (SPH). In the Eulerian method, the physical fluid is constrained in cells and properties are computed at the fixed cell boundaries. In some cases, when these properties meet certain criteria (e.g. in mass or temperature), cells are further divided, a technique known as adaptive mesh refinement (AMR). SPH, AMR and combinations of them are the basis of many current cosmological simulations (Somerville & Davé 2015; Dayal & Ferrara 2018).

Currently, there is a large number of cosmological hydrodynamical simulations* which vary in sizes, mass resolution, number of particles and physical recipes to reproduce the Universe (an example of comparing these characteristics in cosmological simulations can be found in Fig.1 of Nelson et al. 2019). Among those simulations, I highlight popular (and open to the public) simulations such as EAGLE (Schaye et al. 2015; Crain et al. 2015), SIMBA (Davé et al. 2019) and IllustrisTNG-50 (Pillepich et al. 2019; Nelson et al. 2019). Of these simulations, IllustrisTNG-50 seems to be the most promising simulation in terms of resolution and computational effort, however, in this thesis we focus on EAGLE simulations, which we describe below.

EAGLE

The Evolution and Assembly of GaLaxies and their Environments (EAGLE) project is a set of simulations created with the aim of understanding how galaxies form and evolve. EAGLE simulations were run using a modified version of GADGET-3 (last described by Springel 2005). This modified version of GADGET-3 adopts an SPH pressure-entropy parameterisation following Hopkins (2013). In the subgrid physics of the simulations, different physical processes are included, listed in the following chapters, such as SF and AGN feedback (Schaye & Dalla Vecchia 2008; Dalla Vecchia & Schaye 2012).

The set of EAGLE simulations spans a range of box sizes between 12 and 100 cMpc (comoving Mpc), with intermediate and high mass resolutions ($1.81 \times 10^6\,M_{gas}$ and $2.26 \times 10^5\,M_{gas}$, respectively). Global information on the galaxies, merger trees and apertures can be found in the EAGLE database (McAlpine et al. 2016), while particle data can be found on their webpage† (The EAGLE team 2017). In total, EAGLE

* A list of cosmological hydrodynamical simulations of the Universe can be found at https://www.tng-project.org/dev707/data/landscape/

† http://icc.dur.ac.uk/Eagle/

contains seven fiducial simulations (Schaye et al. 2015) and 17 simulations where subgrid parameters have been modified to explore the parameter space for the SF feedback and IMF (Crain et al. 2015; Barber et al. 2018).

One of the reasons to use EAGLE in this thesis is that it successfully reproduces some of the observational results in terms of cosmic SFRD, sSFR and galaxies' colours (Schaye et al. 2015; Furlong et al. 2015; Trayford et al. 2015; Katsianis et al. 2017). An example of a disk galaxy simulated within the high-resolution model in EAGLE is shown in Fig. 1.7. It is possible to notice the difference in the spatial distribution of SPH particles of gas and stars for face-on and edge-on views. The information available in these SPH is what defines the ISM conditions of the galaxy. Unfortunately, EAGLE does not resolve the cold ISM; therefore we use a radiative transfer tool to estimate the conditions of the ISM separately.

1.4.2 Radiative transfer models

The transport of electromagnetic radiation through different environments in the Universe is very complex. Photon scattering, absorption, and re-emission do not allow us to solve radiative transfer paths with simple models. Therefore, complex numerical models are required to estimate the radiation in idealised environments to be used as an approximation of the real environments. To calculate the radiation emitted by the ISM in galaxies, and especially the atomic species in the FIR used in this thesis, we require the use of these numerical models of radiative transfer. Table 1.2 presents a list of the different numerical models and software that can be used for this purpose, with their main characteristics and in alphabetical order. This list contains software that has been discussed in Olsen et al. (2018b) and that I am aware of.

As we can see in Table 1.2, some of these tools are no longer under development (i.e. they are obsolete) or are not open to the public, which limits their usefulness. In terms of geometry, some of them are designed to work only on 1-dimension (e.g. Cloudy and DESPOTIC), 3-dimensions (e.g. ART^2 and RADMC-3D) or both (e.g. LOC). Furthermore, some of the codes were not designed for a general (astrophysical) purpose and are limited by their chemical networks or density regimes (Olsen et al. 2018b). To quantify the popularity of these codes, I show in the last column of Table 1.2 the reach of the referenced works (in the second column). I present the average number of readings per year that these models have in the NASA Astrophysics Data System Bibliographic Services (ADS) in the last five years (from 2017 to 2021). The comparison of these numbers can show the importance and attractiveness that these tools can have for doing research. However, these numbers are for illustrative comparisons only and should not be taken as an indicator of research impact.

In this thesis, I decided to use one of the radiative transfer computational tools to estimate line emission. We are inclined to use Cloudy in the following chapters due to the community support around it and the different environments we can create using this tool. In the next paragraph, I will briefly explain some of the details of Cloudy.

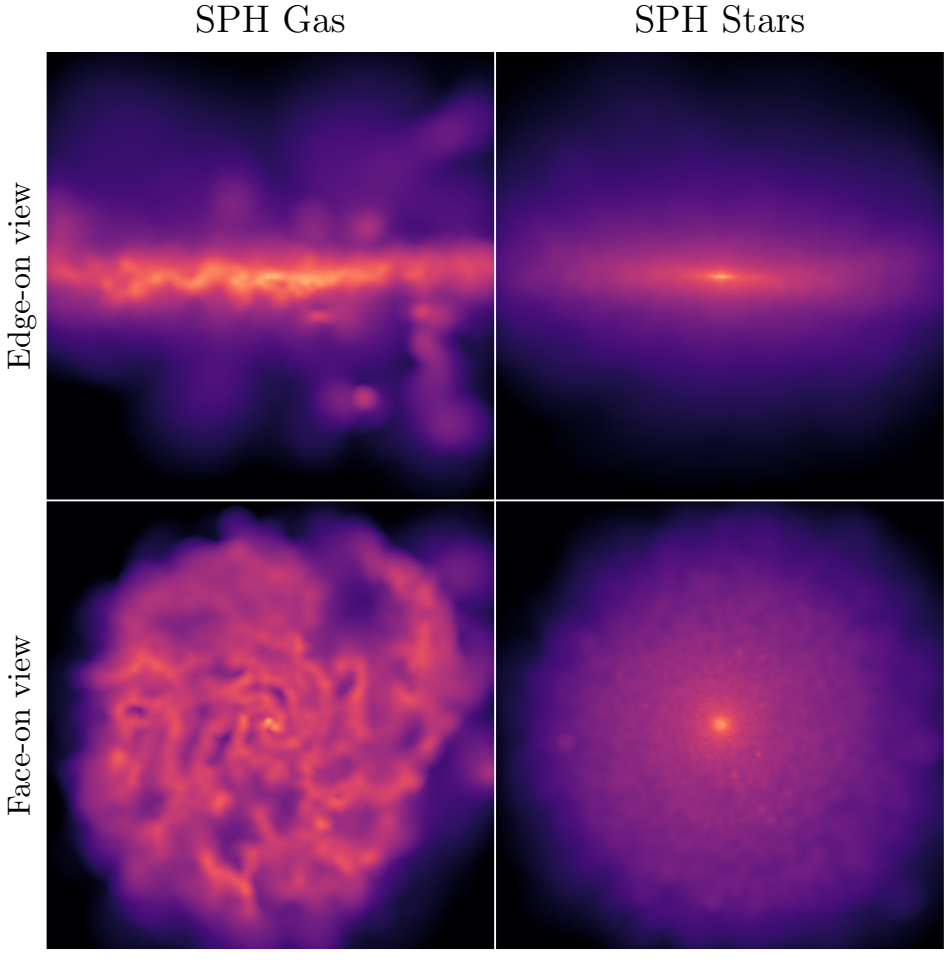

Figure 1.7 – *Example of a galaxy created by* EAGLE *in terms of the SPH gas and stars particles. Upper panels show the edge-on view of the disc galaxy while bottom panels show the face-on view. The size of the galaxy is ∼30 kpc. The colour scale for the SPH mass distribution has been arbitrarily chosen to highlight the structure of the galaxy. Image created with the help of Py-Sphviewer (Benitez-Llambay 2015).*

Name	References	Public code	Latest stable version[a]	Programming Language	ADS reach	In development
ART[2]	[1–3]	No	No release	FORTRAN[b]	28 – 53[c]	Yes
CLOUDY	[4–6]	In Gitlab (1)	v17.03	C	339 – 835	Yes
DESPOTIC	[7]	In Bitbucket (2)	2022	Python	107	Yes
LIME	[8]	In Github (3)	v1.9.5	C	175	No
LOC	[9]	In Github (4)	2022	Python and C	22[c]	Yes
MAIHEM	[10–12]	No	No release	FORTRAN	40 – 49	Yes?
MOLLIE	[13–15]	No	2013	FORTRAN and C	24 – 58	No
POLARIS	[16–17]	In Github (5)	V4.04	C++	60 – 150[c]	Yes
RADEX	[18]	In personal webpage (6)	2017	FORTRAN (Python[d])	341	Yes
RADMC-3D	[19]	In Github (7)	v2.0	FORTRAN and Python	333	Yes
RATRAN	[20]	In personal webpage (8)	2016	FORTRAN	112	No
SKIRT	[21–23]	In Github (9)	v9.0	C++	38 – 99[c]	Yes

Table 1.2 – *List of available software used in astrophysics to estimate line emissions.*

Notes: [a] When there is no stable version we cite the year of the last changes. [b] From the original Montecarlo code, but not clear in the references. [c] References have less than 5 years. [d] RADEX has two Python wrappers currently available in Github (`https://github.com/uclchem/SpectralRadex` and `https://github.com/keflavich/pyradex/`).
References: [1–3]: Li et al. (2008, 2020b); Yajima et al. (2012), [4 –6]: Ferland et al. (1998, 2013, 2017), [7]: Krumholz (2014), [8]: Brinch & Hogerheijde (2010), [9]: Juvela (2020), [10–12]: Gray et al. (2015); Gray & Scannapieco (2016, 2017), [13–15]: Keto (1990); Keto et al. (2004); Keto & Rybicki (2010), [16–17]: Reissl et al. (2016, 2019), [18]: van der Tak et al. (2007), [19]: Dullemond et al. (2012), [20]: Hogerheijde & van der Tak (2000), [21–23]: Baes et al. (2003); Camps & Baes (2015, 2020).
Public code links: (1): `https://gitlab.nublado.org/cloudy/cloudy/-/wikis/home`, (2): `https://bitbucket.org/krumholz/despotic`, (3): `https://github.com/lime-rt/lime` , (4): `https://github.com/mjuvela/LOC`, (5): `https://github.com/polaris-MCRT/POLARIS`, (6): `https://personal.sron.nl/~vdtak/radex/index.shtml`, (7): `https://github.com/dullemond/radmc3d-2.0`, (8): `https://personal.sron.nl/~vdtak/ratran/frames.html`, and (9): `https://github.com/SKIRT/SKIRT9`.

Cloudy

Cloudy is a non-local thermodynamic equilibrium (NLTE) spectral synthesis and plasma simulation code for ISM conditions (Ferland et al. 2017). With the help of Cloudy, it is possible to predict the thermal, ionisation and chemical structure of a cloud defined by the user. There are three main components that need to be defined: the geometry of the cloud, the intensity of the radiation field hitting the cloud and the atomic composition of the cloud. In terms of geometry, Cloudy assumes a 1D plane-parallel or spherical cloud spherical with a constant or variable density. It is possible to define an open geometry where the radiation is external to the cloud or a closed geometry where the radiation comes from the centre of the cloud. The size (radius) of the clouds can be defined manually, as well as the depth. In terms of the atomic composition, the abundance of most elements can be specified. The software lets these elements react into molecules (e.g. C and O in CO) which can also be added together with dust grains in the computations. Ionisation and recombination mechanisms are also computed for the species in the software. Finally, in terms of the radiation field, it is possible to describe its spectral shape and intensity. This radiation field is not unique as it can be a combination (sum) of different radiation fields. It is also possible to define a stellar atmosphere to be used within Cloudy. In this thesis, we use Starburst99 so that the stellar evolution models are consistent with other Cloudy look-up tables that we use. It is important to mention that Cloudy does not perform exact radiative transfer. Instead, it uses the escape probability method to predict the intensity of the emission lines.

1.4.3 Spectral energy distributions

Some of the results of the radiative transfer models can be used to describe spectral energy distributions (SEDs). These SEDs take into account the contributions of emission processes that occur at different wavelengths and are reflected in spectral features and photometric bands. For example, in a SED model, it is possible to add the contributions of star formation histories (SFH), stellar templates, dust attenuation, dust emission, initial mass function (IMF), nebular emission lines, metallicities and AGN templates. With these contributions, it is feasible to estimate the physical parameters of the galaxies such as the SFR, the dust luminosity, the stellar mass and the contribution of the AGN to the IR emission, among others. Therefore, SED fitting tools can be very useful to retrieve information from galaxies where the wavelength coverage is enough to have reliable estimates.

There are different codes/software that can be used to combine the stellar, dust and AGN components to fit the SED in galaxies between UV and FIR wavelengths. Thorne et al. (2021) presented a list of the most popular SED fitting codes with their characteristics and templates used. For the purpose of this thesis, we focus on the contribution of the AGN in the IR and SFR. Unfortunately, only three of the models presented by Thorne et al. (2021) have a dedicated AGN module: CIGALE (Noll et al. 2009; Boquien et al. 2019), ProSpect (Leja et al. 2017; Johnson et al. 2021) and Prospector (Robotham et al. 2020). Of those SED tools, we decided to use CIGALE because it allows us to work with the different AGN templates and the nebular models that come from Cloudy.

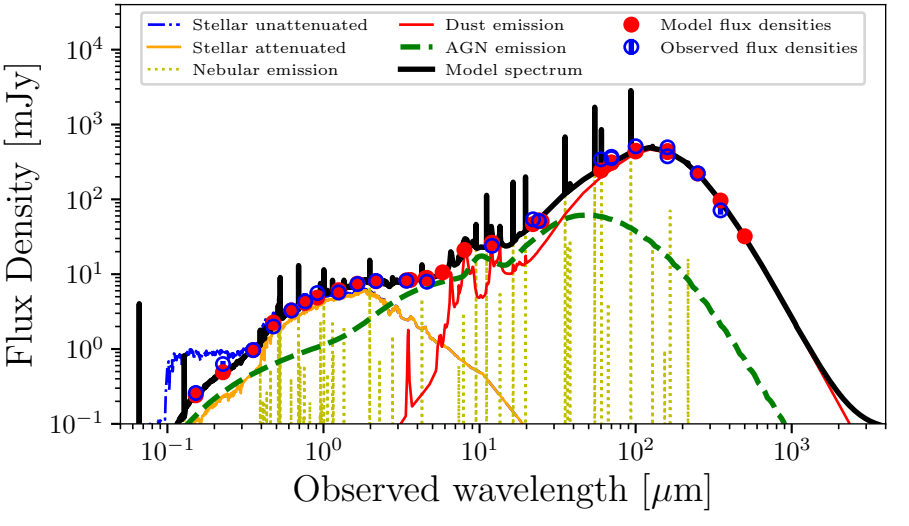

Figure 1.8 – *Example of the SED fitting by* CIGALE *for the galaxy Mrk 662. This plot contains the model spectrum (black line) that is created by the contribution of: nebular emission (gold dotted lines), attenuated (orange) and non-attenuated stellar emission (blue dot-dashed), dust emission (red solid), and AGN emission (green dashed). The red dots are the best model flux densities and the blue squares mark the observed flux densities with 1σ error bars.*

CIGALE

CIGALE (Boquien et al. 2019) is a SED tool that has been designed to follow three principles: modularity, clarity and efficiency. In terms of modularity, the physical components (i.e. the SFH, dust emission, etc.) work independently, as well as the input, computation, analysis and output phases. In terms of clarity, the code is written to be easy to understand and develop. And in terms of efficiency, the code optimises memory and power usage. These principles allow CIGALE to be used by a wide astronomy community since the modules are written in Python under a Free licence.

An example of the SED fitting done in CIGALE, also presented in Chapter 6, is shown in Fig. 1.8. Here, we see how the model spectrum created by user-defined contributions from stellar, dust, nebular lines and AGN emissions fits well with the observed photometric data of a galaxy. The final model spectrum is used to derive the physical parameters and, at the same time, estimate the model flux densities. Those estimated flux densities are used to compare with observational data and can be used as expected values when observations are not available. Currently, the last version of CIGALE covers the range between X-ray and radio wavelengths (Yang et al. 2022a), showing the potential that this tool has as we show in this thesis.

1.5 This thesis

The main goal of this PhD thesis is to understand the gas processes of galaxies in the cosmological context of star formation. In this thesis, I address the following questions:

- **How to trace the effects of the ISM gas properties as a function of cosmic time?**
- **How to diagnose and decouple the ISM phases in galaxies using FIR emission lines?**
- **What is the AGN fractional contribution in active galaxies (mergers, AGN and classical starburst) in the main-sequence of star-forming galaxies?**

To address these questions, we implement a theoretical and computational approach using the tools described in this introduction. Chapters 2–4 focuses on the ISM. I post-process a set of cosmological simulations from EAGLE and compute FIR line emission in the simulated galaxies with the help of CLOUDY. I implement a simple model of the ISM and its phases based on knowledge of the [C II] luminosity in the local Universe (Chapter 2). I then expand the model to several FIR lines and compare the luminosity estimates with the physical properties of the simulated galaxies, such as radiation field and pressure (Chapter 3). I present a web app that can be used to give estimates of galaxy physical properties based on observed line luminosities. We also include predictions on line observations (e.g. the number of sources as a function of redshift) of future ground and space IR telescopes (Chapter 4).

Chapters 5 and 6 focuses on SED analysis of active galaxies. I fit the SED from a set of merger, AGN and starburst galaxies in CIGALE to estimate their physical parameters (e.g. SFR and stellar mass) using photometric observational data in the range between UV and FIR wavelengths. I examine how these parameters change with the AGN contribution in the context of the main-sequence of star-forming galaxies (Chapter 5). Finally, I study how tools like CIGALE can be used to classify AGN galaxies (Seyferts) and at the same time help to understand the AGN unified model (Chapter 6).

1.5.1 Outline

Most of the work presented in this thesis has been published in peer-reviewed journals and/or is available to the public. Here I give a brief description of the contents of each of the chapters.

Chapter 2 (Ramos Padilla et al. 2021): In this chapter, I use the [C II] line at 158 μm as a benchmark of gas cooling in galaxies in the local Universe. I post-process EAGLE simulations to predict the luminosity of [C II] with the help of CLOUDY lookup tables. I implement a model that estimates the luminosity coming from three phases of the ISM: dense molecular gas, neutral atomic gas and

diffuse ionised gas. These phases are used to compare the observational luminosity data with the model in terms of the SFR–luminosity relationship. I examine the dependence of the different ISM phases in terms of the total star-formation rate and metallicity. This chapter presents the first estimates of $L_{[\mathrm{C\,II}]}$ in a cosmological simulation for local Universe galaxies.

Chapter 3 (Ramos Padilla et al. 2022): In this chapter, I extend the model described in Chapter 2 by adding HII regions as a new ISM phase. I also estimate the line luminosity for seven additional FIR lines: [O I] at 63 and 145 μm, [N II] at 122 and 205 μm, [O III] at 52 and 88 μm, and [N III] at 57 μm. Estimates are made in the local Universe ($z = 0$) and at different redshifts up to $z = 6$. I present the SFR–luminosity relationship and the fractional contribution of the different ISM phases for each of the FIR emission lines. Finally, I compare the results in terms of diagnostic diagrams between the observations and the ISM mode estimates. The observations come from a collection of reported work where at least one of the FIR line luminosities is available. This chapter presents all the observational and simulated data used for Chapter 4.

Chapter 4 (This thesis): In this chapter, I introduce a web app called DiagISM. DiagISM uses the information of the ISM model presented in Chapter 3 to estimate physical parameters from observations of FIR line luminosities. Within DiagISM it is possible to select between two multi-layer perceptron models to estimate physical parameters such as SFR, interstellar radiation field and metallicity, using a user-friendly environment. In addition, I present some number counts assuming a likely configuration for a dedicated FIR space telescope. Both the web app and the number of counts will be important for planning future observations.

Chapter 5 (Ramos Padilla et al. 2020): In this chapter, I collect several samples of galaxies at different interaction stages to understand how these types of galaxies have enhanced star formation and what means for the main-sequence of star-forming galaxies. I also compare interacting galaxies with samples of classical starburst and dominant AGN galaxies for control. All samples have good coverage of photometric bands between UV and FIR wavelengths. I reduce spectroscopic and photometric data from raw data to be used in the SED analysis. Using CIGALE, I retrieve the physical conditions of all galaxies (i.e. SFR, AGN fractional contribution, stellar mass, dust luminosity, etc.) to assess the impact of the AGN on the main-sequence of star-forming galaxies. Results from this modelling are compared with other literature diagnostics to validate the estimations. This chapter shows the importance of taking into account the AGN contribution in SED models.

Chapter 6 (Ramos Padilla et al. 2022): In this chapter, we focus on the validity of classifying AGN galaxies depending on their line-of-sight (viewing angle) as predicted by the unified model of AGN. We evaluate the importance of the viewing angle in the classification of AGN galaxies using the AGN SED models available in CIGALE. We use a sample of Seyfert galaxies with data available in public astronomical databases so that the analysis is reproducible. We use machine learning ensemble methods to verify the classifications from the databases and see how these classifications are reflected in the inferred physical parameters of galaxies (e.g in terms of the viewing angle and SFR). This chapter shows that the observed

and intrinsic AGN disc luminosity (i.e. accretion rate power) can be used to draw evolutionary paths between Type-1 and Type-2 AGN.

Chapter 7: In this chapter, I provide the conclusions of the thesis on three topics: The ISM conditions at $z \leq 6$, star-forming and AGN galaxies, and the future prospects of observational data, modelling and open science.

1.5.2 Data availability

From each of the chapters presented in this thesis (except for Chapter 7), I provide the links to the data involved to create the figures or the final outputs of the research papers.

Chapter 1: The data and codes of the figures presented in this chapter can be found in the Zenodo and GitHub repositories at https://github.com/aframosp/PhDThesisIntro. The figures are presented in Jupyter Notebooks to be easily reproduced.

Chapter 2: Python code describing the ISM model (also for Chapter 3) is available in the Github repository https://github.com/aframosp/DiagISM, while estimated data from EAGLE simulated galaxies are in a Zenodo repository at https://doi.org/10.5281/zenodo.6638024.

Chapter 3: Estimated data from EAGLE simulated galaxies and the collection of observational samples are in a Zenodo repository at https://doi.org/10.5281/zenodo.6134280 (2022).

Chapter 4: The web app described and used in this chapter can be accessed at https://aframosp-diagism.streamlitapp.com. The code and data to reproduce the results of this chapter are in the Github repository at https://doi.org/10.5281/zenodo.6705031.

Chapter 5: Photometry and estimated data from CIGALE on the physical parameters of galaxies described in the chapter are located in Vizier at https://cdsarc.cds.unistra.fr/viz-bin/cat/J/MNRAS/499/4325.

Chapter 6: The data and code to reproduce the results of this chapter are in the Github repository at https://github.com/aframosp/AGNView (2021a). Estimated data from CIGALE on the physical parameters of galaxies described in the chapter are located in Vizier at https://cdsarc.cds.unistra.fr/viz-bin/cat/J/MNRAS/510/687. The complementary data with all the estimates made with CIGALE are in a Zenodo repository at https://doi.org/10.5281/zenodo.5221764 (2021b).

1

Every human being has a basic instinct to help each other out. It might not seem that way sometimes, but it's true.

Andy Weir - The martian

Food! There is free food!

Andrés during any event

2

Diagnosing the interstellar medium of galaxies with far-infrared emission lines: The [C II] 158 μm line at z$\sim$0

A. F. Ramos Padilla, L. Wang, S. Ploeckinger, F. F. S. van der Tak and S. C. Trager

Published in Astronomy & Astrophysics, Vol. 645, A133, 18 pp.

Highlights

- The [C II] line at 158 μm is a robust indicator of gas cooling in galaxies in the local Universe.
- We post-process EAGLE simulations to predict the luminosity of [C II].
- We implement a model that estimates the luminosity that comes from three phases of the ISM: dense molecular gas, neutral atomic gas and diffuse ionised gas.
- This model shows a dependence on the fractional contribution of the different ISM phases to the $L_{[\mathrm{C\,II}]}$ of the total star-formation rate and metallicity.
- Star-formation regulation and metallicity could be responsible for the variations observed in $L_{[\mathrm{C\,II}]}$ at high infrared luminosities.

Abstract

Context: Atomic fine structure lines have been detected in the local Universe and at high redshifts over the past decades. The [C II] emission line at 158 μm is an important observable as it provides constraints on the interstellar medium (ISM) cooling processes.
Aims: We develop a physically motivated framework to simulate the production of far-infrared line emission from galaxies in a cosmological context. This first paper sets out our methodology and describes its first application: simulating the [C II] 158 μm line emission in the local Universe.
Methods: We combine the output from EAGLE cosmological hydrodynamical simulations with a multi-phase model of the ISM. Gas particles are divided into three phases: dense molecular gas, neutral atomic gas, and diffuse ionised gas (DIG). We estimate the [C II] line emission from the three phases using a set of Cloudy cooling tables.
Results: Our results agree with previous findings regarding the contribution of these three ISM phases to the [C II] emission. Our model shows good agreement with the observed $L_{[\mathrm{C\,II}]}$–star formation rate (SFR) relation in the local Universe within 0.4 dex scatter.
Conclusions: The fractional contribution to the [C II] line from different ISM phases depends on the total SFR and metallicity. The neutral gas phase dominates the [C II] emission in galaxies with SFR ~ 0.01–$1\,\mathrm{M_\odot\,yr^{-1}}$, but the ionised phase dominates at lower SFRs. Galaxies above solar metallicity exhibit lower $L_{[\mathrm{C\,II}]}$/SFR ratios for the neutral phase. In comparison, the $L_{[\mathrm{C\,II}]}$/SFR ratio in the DIG is stable when metallicity varies. We suggest that the reduced size of the neutral clouds, caused by increased SFRs, is the likely cause for the $L_{[\mathrm{C\,II}]}$ deficit at high infrared luminosities, although EAGLE simulations do not reach these luminosities at $z = 0$.

Keywords: Galaxies: evolution, formation – ISM: lines and bands, clouds, evolution – methods: numerical

2.1 Introduction

Over the past decades, significant efforts to develop theoretical models have helped improve our knowledge of how galaxies form and evolve. Implementations of hydrodynamical simulations (e.g. Vogelsberger et al. 2014; Hopkins et al. 2014; Schaye et al. 2015; Pillepich et al. 2018) and semi-analytic models (e.g. Somerville & Primack 1999; Cole et al. 2000; Croton et al. 2006) have provided fundamental insights. Most models can now roughly reproduce the observed specific star formation rate (SFR) and mimic the scenario where more massive galaxies formed their stars earlier than lower mass galaxies, an effect known as 'downsizing' (Cowie et al. 1996; Pérez-González et al. 2008; Haines et al. 2017). These implementations agree, within a factor of three, with physical properties of galaxies and, in recent years, have begun to converge regarding physical interpretations (Somerville & Davé 2015).

One critical question, tackled by several works in recent years (see e.g. Nagamine et al. 2006; Vallini et al. 2013, 2015; Katz et al. 2017, 2019; Pallottini et al. 2017a,b; Olsen et al. 2015, 2017; Lupi & Bovino 2020; Lupi et al. 2020; Moriwaki et al. 2018; Arata et al. 2020), is how to include the cooling processes of the interstellar medium (ISM) in galactic theoretical models beginning from early epochs. Unfortunately, there is still no single model capable of reproducing all the observational data across all of

cosmic history. The balance of gas heating and cooling in the ISM environment (which affects physical properties such as temperature and density) is central to theoretical models. Owing to observational constraints, our ISM knowledge mainly comes from our Galaxy and its nearby cosmic neighbourhood, where cold atomic clouds (cold neutral medium, CNM), diffuse warm neutral and ionised emission (WNM and WIM), and H II regions are distinguishable (e.g. Hollenbach & Tielens 1999; Wolfire et al. 1995, 2003; Kaufman et al. 1999).

Far-infrared (FIR) emission line observations of nearby galaxies that trace these phases of the ISM (Díaz-Santos et al. 2017; Herrera-Camus et al. 2018b) are an important tool for understanding ISM cooling processes and their relation with the SFR, especially in cool gas where permitted lines of hydrogen cannot be excited ($< 10^4$ K). However, the current resolution of (most) instruments limits the capability to provide spatially resolved line observations of high-redshift galaxies as in nearby galaxies. There are a few exceptions: Gas an dust clouds can be observed at high redshifts using high spatial resolution with ALMA (e.g. Oteo et al. 2017) or using gravitational lensing (Dessauges-Zavadsky et al. 2019). Extrapolating observations from the local Universe is not a good option as the ISM phases are likely to be different at earlier times. For example, gas-phase metallicities change with redshift, which will have an impact on the ISM phases (Bialy & Sternberg 2019). The growing body of FIR line observations at high-redshift has so far mainly been interpreted based on empirical relations. For example, the $L_{[\mathrm{C\,II}]}$–SFR relations obtained by De Looze et al. (2014) are used to interpret high-redshift galaxies (Inoue et al. 2016; Pentericci et al. 2016; Knudsen et al. 2016). However, over the last few years, new high redshift observations have begun to use emission models to interpret the observations of these lines (e.g. Maiolino et al. 2015; Carniani et al. 2017; Bakx et al. 2020; Béthermin et al. 2020).

Hydrodynamical simulations represent one of the most promising methods to avoid this extrapolation from the local Universe to high redshifts. These simulations can predict the interplay between dark matter and baryons in the large-scale structure and the final properties of galaxies (Dayal & Ferrara 2018). However, this method is computationally expensive if all the relevant physics are considered. Limitations in spatial resolutions and simulation techniques have to be taken into account in the sub-grid physics in different box sizes (see Dayal & Ferrara 2018, their Table 1). Fortunately, zoom-in techniques are starting to bridge the gap between individual stellar and galactic scales, which help to model the general ISM (Somerville & Davé 2015). Therefore, hydrodynamical simulations can help us to predict emission lines at different cosmic epochs and characterise the ISM at different cosmic times (Pallottini et al. 2017a).

The [C II] 158 μm emission line is one of the brightest emission lines in the FIR. Its luminosity is equivalent to values around 1% of the FIR luminosity of galaxies (e.g. Stacey et al. 1991; Brauher et al. 2008). It is an easily observed line that traces various phases of the ISM where gas is exposed to energies above the carbon ionisation potential (11.3 eV compared to 13.6 eV for hydrogen). [C II] can be considered as a robust cooling line (in the range of 20-8 000 K) of the ISM, acting as a thermostat (Gong et al. 2012; Goldsmith et al. 2012; Olsen et al. 2015).

Nevertheless, [C II] is difficult to interpret, as it arises from diverse environments:

CNM, WNM and WIM. In addition, at higher FIR luminosities, the luminosity of [C II] over the FIR luminosity increases at a lower rate, an effect known as the '[C II] deficit' (e.g. Díaz-Santos et al. 2017, who also observed deficits in other emission lines). These problems do not limit the capacity of [C II] for tracing SFR in local luminous infrared galaxies (e.g. Malhotra et al. 2001; Stacey et al. 2010; Díaz-Santos et al. 2013; Herrera-Camus et al. 2015; Díaz-Santos et al. 2017), but a complete understanding of the origin of the deficit is necessary in order to use this line as a SFR indicator in high-redshift galaxies (De Looze et al. 2011; Gullberg et al. 2015; Spilker et al. 2016). Currently, dependencies on metallicities and radiation fields (radiative feedback) seem to be the most probable regulators of the [C II] line luminosity in theoretical (e.g. Malhotra et al. 2001; Muñoz & Oh 2016; Narayanan & Krumholz 2017; Vallini et al. 2017; Ferrara et al. 2019) and observational (Stacey et al. 2010; Díaz-Santos et al. 2017; Herrera-Camus et al. 2018b) studies.

The goal of this paper is to present a model for the ISM [C II] line emission and show its applications in the local Universe ($z = 0$). We implement this model of FIR line emission to comprehend the ISM physical conditions in galaxies with line properties. We use $z = 0$ as a benchmark, as locally we have the best observational constraints that cover a diverse range of galaxies in different environments. Our first target is the [C II] 158 μm line at $z = 0$, for which we assume contributions from the atomic, molecular, and ionised ISM phases to obtain a model of the [C II] emission in galaxies. We model the emission of [C II] by post-processing hydrodynamical simulations from the Evolution and Assembly of GaLaxies and their Environments (EAGLE) project (Schaye et al. 2015; Crain et al. 2015) with a physically motivated model of the ISM. We use Cloudy (Ferland et al. 1998, 2013, 2017) cooling tables (Ploeckinger & Schaye 2020) to predict line emission to help us constrain different ISM phases in galactic environments. Throughout this paper, we assume the ΛCDM cosmology from Planck results (Planck Collaboration et al. 2014b) ($\Omega = 0.307$, Ω_Λ=0.693, $H_0 = 67.7$ km s^{-1} Mpc^{-1} and $\sigma_8 = 0.8288$).

In this paper, we first give an overview of the methods we use to predict the emission lines (Sect. 2.2). Then, we verify that our results agree with observations of local galaxies in terms of physical parameters and scaling relations (Sect. 2.3). After that, we discuss the difference between our findings and other papers. (Sect. 2.4), and finally, we present our conclusions (Sect. 2.5).

2.2 Methodology

In the next sections, we describe the sets of simulations we use (Sect. 2.2.1), the model to characterise the multi-phase structure of the ISM (Sect. 2.2.2), and the estimation of line luminosity using Cloudy cooling tables (2.2.3).

2.2.1 The EAGLE simulations

EAGLE (Schaye et al. 2015; Crain et al. 2015) consists of several cosmological hydrodynamical simulations, run in an N-body smoothed particle hydrodynamics (SPH) code. Briefly, EAGLE adopts a pressure-entropy parameterisation using the description of Hopkins (2013). The simulations include radiative cooling and photo-electric

Table 2.1 – EAGLE *simulations used in this work. The top simulation is the only high-resolution simulation used, while the other two are intermediate-resolution simulations with different box-sizes. The box-size (L) and maximum softening length are presented in comoving and proper distances (cMpc and pkpc). The last column shows the number of galaxies used in this work for a given simulation.*

Name in Schaye et al. (2015)	L (cMpc)	# particles	Gas mass ($M_\odot$)	Max. Softening (pkpc)	# galaxies
RECAL-L0025N0752	25	752^3	2.26×10^5	0.35	415
REF-L0025N0376	25	376^3	1.81×10^6	0.7	202
REF-L0100N1504	100	$1\,504^3$	1.81×10^6	0.7	$5\,000^a$

Notes. $^{(a)}$ We selected the top 5 000 galaxies in terms of gas mass.

heating (Wiersma et al. 2009a), star formation (Schaye & Dalla Vecchia 2008), stellar evolution and mass loss (Wiersma et al. 2009b), black hole growth (Springel et al. 2005; Rosas-Guevara et al. 2015), and feedback from star formation and AGNs (Dalla Vecchia & Schaye 2012).

This work uses three simulations of different sizes and resolutions from EAGLE: REF-L0025N0376, REF-L0100N1504, and RECAL-L0025N0752. Table 2.1 presents their main characteristics. We use these simulations to compare the implemented model when using different box-sizes and resolutions. Variations in terms of box-size are analysed when comparing the REF-L0100N1504 and REF-L0025N0376 simulations. Variations in terms of resolution are analysed when comparing the REF-L0025N0376 and RECAL-L0025N0752. EAGLE simulations calibrate the physical parameters of the sub-grid routines to reproduce the observed galaxy stellar mass functions at $z \approx 0$ (GSMF; Schaye et al. 2015). The sub-grid parameters of the high-resolution, small-box simulation RECAL-L0025N0752 were re-calibrated to account for the increased resolution (Schaye et al. 2015; Crain et al. 2015). REF-L0100N1504 and RECAL-L0025N0752 are similar in terms of 'weak convergence', which means numerical results converge in different simulations after re-calibrating the sub-grid parameters (Furlong et al. 2015; Schaye et al. 2015).

We selected our sample of galaxies from the EAGLE database (McAlpine et al. 2016) similarly to the study of Thob et al. (2019). We focussed on 'central' galaxies, sub-halos containing the particle with the lowest value of the gravitational potential, instead of 'satellite' galaxies, to avoid environmental influence on the morphology and kinematics of the galaxies. We used galaxies with at least 300 star particles within 30 pkpc (proper kpc) from the centre of the potential. For REF-L0100N1504, the top 5 000 most-gas-massive galaxies are selected to cover the range of gas mass that REF-L0025N0376 cannot cover. Simulated galaxies are retrieved from the SPH data (The EAGLE team 2017) by using FoF (Friends-of-Friends) and SUBFIND algorithms (Springel et al. 2001; Dolag et al. 2009) in the dark matter halos. In the last column of Table 2.1, we present the total number of galaxies used to calculate line luminosities with our model in each of the simulations.

Table 2.2 – *Physical properties used from the* EAGLE *data for stars and gas particles.*

Property	Symbol	Description [Units]
Gas properties		
Density	ρ	Co-moving density [g cm^{-3}]
Smoothed element abundance hydrogen	X_H	SPH weighted hydrogen abundance
Smoothed element abundance carbon	X_C	SPH weighted carbon abundance
Entropy	E	Particle entropy [Pressure × Density$^{\gamma}$]
Mass	$m_{\rm SPH}$	Particle mass [g]
Smoothed metallicity	Z	SPH kernel weighted metallicity
Smoothing length	h	Co-moving SPH smoothing kernel [cm]
Star formation rate	SFR	Instantaneous star formation rate [M$_\odot$/yr]
Temperature	$T_{\rm SPH}$	Particle temperature (from the internal energy) [K]
Star properties		
Star formation time	t_*	Time when a star particle was born
Stellar mass	m_*	Current particle mass [g]
Smoothed metallicity	Z_*	Co-moving SPH smoothing metallicity

2.2.2 The multi-phase ISM model

In this section, we describe the model for the multi-phase ISM using gas and star SPH particles in terms of the simulation properties (Sect. 2.2.2) and three different ISM environments: dense molecular clouds (Sect. 2.2.2), neutral atomic gas (Sect. 2.2.2), and diffuse ionised gas (DIG, Sect. 2.2.2).

Gas and star particle properties

We selected SPH particles inside a volume of radius 30 pkpc to ensure that the derived parameters in our modelling are similar to those in the EAGLE database* for each galaxy. In Table 2.2, we present the symbol and the properties from the particle data that we use for the modelling. In addition, we obtained the IDs for the particles and their position to study their spatial distribution.

We retrieved physical parameters from the gas in each of the SPH particles before dividing these SPH into neutral and ionised gas phases. We calculated the total hydrogen number density as

$$n(\mathrm{H}) = \frac{\rho X_H}{m_H}, \tag{2.1}$$

with m_H the hydrogen mass, ρ the density and X_H the SPH weighted hydrogen abundance. EAGLE imposes a pressure floor expressed as a polytropic equation-of-state as $P \geq P_{\rm lim}\,(\rho/\rho_{\rm lim})^{\gamma_{\rm lim}}$, where $\gamma_{\rm lim} = 4/3$, $P_{\rm lim}$, and $\rho_{\rm lim}$ are constants ($\sim 1.2 \times 10^{-13}$Ba and $\sim 2.23 \times 10^{-25}$g cm^{-3}, Schaye & Dalla Vecchia 2008; The EAGLE team 2017). A temperature threshold is then obtained from $P_{\rm lim} = \rho_{\rm lim} T_{\rm lim}/(\mu m_H)$, where $\mu \approx 1.23$ is the mean molecular weight of the neutral gas. For gas particles limited by the pressure floor, the temperature $T_{\rm lim}$ is an effective temperature, including unresolved processes in addition to the thermal temperature. We therefore constrain $T_{\rm SPH}$ at $n(\mathrm{H}) > 0.1\,\mathrm{cm}^{-3}$ to 8 000K, which is typical of the warm ISM (WNM and WIM).

We followed Rahmati et al. (2013) to calculate the fraction of neutral hydrogen $\eta =$

* http://virgodb.dur.ac.uk/

$n(H_I)/n(\mathrm{H})$ given by the ionisation equilibrium $n(\mathrm{H\,I})\Gamma_{\mathrm{Tot}} = \alpha_A n_e n(\mathrm{H\,II})$ as

$$\Gamma_{\mathrm{Tot}} = \frac{\alpha_A \left(1-\eta\right)^2 n(\mathrm{H})}{\eta}, \tag{2.2}$$

where α_A is the Case A recombination rate, Γ_{Tot} is the total ionisation rate, and n_e, $n(\mathrm{H\,I})$, and $n(\mathrm{H\,II})$ are the number densities of electrons, neutral hydrogen and ionised hydrogen, respectively. To solve this equilibrium, we used α_A from Hui & Gnedin (1997),

$$\alpha_A = 1.269 \times 10^{-13} \frac{\lambda^{1.503}}{\left[1+\left(\frac{\lambda}{0.522}\right)^{0.47}\right]^{1.923}} \mathrm{cm^3 s^{-1}}, \tag{2.3}$$

where $\lambda = 2T_{\mathrm{TR}}/T$, with $T_{\mathrm{TR}} = 157\,807\mathrm{K}$, is the ionisation threshold for H I, and T is the gas temperature. Γ_{Tot} can be defined as

$$\Gamma_{\mathrm{Tot}} = \Gamma_{\mathrm{Phot}} + \Gamma_{\mathrm{Col}}, \tag{2.4}$$

where Γ_{Phot} is the photo-ionisation rate, and Γ_{Col} is the collisional ionisation rate. Γ_{Col} was calculated following Theuns et al. (1998) as

$$\Gamma_{\mathrm{Col}} = \Lambda_T (1-\eta) n(\mathrm{H}), \tag{2.5}$$

where

$$\Lambda_T = 1.17 \times 10^{-10} \frac{T^{1/2} \exp\left(-157\,809/T\right)}{1+\sqrt{T/10^5}} \mathrm{cm^3 s^{-1}}. \tag{2.6}$$

Rearranging Eq. 2.2 with Eqs. 2.3-2.6 as

$$\begin{aligned} A &= \alpha_A + \Lambda_T \\ B &= 2\alpha_A + \frac{\Gamma_{\mathrm{Phot}}}{n(\mathrm{H})} + \Lambda_T \\ C &= \alpha_A \\ \eta &= \frac{B - \sqrt{B^2 - 4AC}}{2A}, \end{aligned} \tag{2.7}$$

we obtained the neutral fraction η for a given density $n(\mathrm{H})$, temperature T and photo-ionisation rate. With η, we calculated the total neutral mass associated with the neutral phase as $m_{\mathrm{neutral}} = \eta \times m_{\mathrm{SPH}}$ and at the same time we defined the total ionised gas mass as $m_{\mathrm{ionised}} = m_{\mathrm{SPH}} - m_{\mathrm{neutral}}$.

We took the background interstellar radiation field (ISRF) over the SPH particle due to star formation into account as another important parameter. For this, we assumed that gas in the disk is self-gravitating and obeys the Kennicutt-Schmidt (KS) star-formation law (Schaye & Dalla Vecchia 2008). We calculated the Jeans length as

$$L_J = \frac{c_s}{\sqrt{G\rho}} \text{ with} \tag{2.8}$$

$$c_s = \sqrt{\frac{\gamma P_{\mathrm{Tot}}}{\rho}}, \tag{2.9}$$

where c_s is the effective sound speed (Schaye & Dalla Vecchia 2008; Schaye 2001), and $\gamma = 5/3$ is the adiabatic index (different from the polytropic index $\gamma_{\rm lim}$). The total pressure for the SPH particle (entropy-weighted, The EAGLE team 2017) is defined as

$$P_{\rm Tot} = E\rho^{\gamma}, \tag{2.10}$$

where E is the entropy. Then the SFR density is

$$\rho_{\rm SFR} = A\rho \left(1{\rm M}_{\odot}/{\rm pc}^2\right)^{-n} \left(\frac{\gamma}{G} f_g P_{\rm Tot}\right)^{(n-1)/2}, \tag{2.11}$$

where the KS law exponent is $n = 1.4$, G is the gravitational constant, the gas fraction f_g is assumed to be unity and the absolute star-formation efficiency is $A = 1.515 \times 10^{-4}\,{\rm M}_{\odot}\,{\rm yr}^{-1}\,{\rm kpc}^{-2}$ (Schaye & Dalla Vecchia 2008; The EAGLE team 2017). The SFR surface density is defined from Eqs. 2.8 and 2.11 as $\Sigma_{\rm SFR} = \rho_{\rm SFR} L_J$. Following Lagos et al. (2015), we determined the background ISRF in units of the Habing radiation field (Habing 1968, $G_0 = 1.6 \times 10^{-3}\,{\rm erg\,cm^{-2}\,s^{-1}}$) as

$$G'_{0,{\rm bkg}} = \frac{\Sigma_{\rm SFR}}{\Sigma_{\rm SFR,MW}}, \tag{2.12}$$

where $\Sigma_{\rm SFR,MW} = 0.001{\rm M}_{\odot}\,{\rm yr}^{-1}\,{\rm kpc}^{-2}$ is the value of the SFR surface density in the solar neighbourhood (Bonatto & Bica 2011).

We calculated the ISRF coming directly from the local stars close to the gas particles (inside the smoothing length h of the gas particle) following the procedure described by Olsen et al. (2017). We used the information from `starburst99` (Leitherer et al. 2014) stellar models with a mass of $10^4{\rm M}_{\odot}$ to obtain a grid of models at different metallicities* and ages† for the star particles. A Kroupa initial mass function (IMF) was adopted from `starburst99` although the EAGLE simulations used a different IMF (Chabrier); the expected differences in the FUV luminosities ($L_{\rm FUV}$) are negligible ($< 2\%$) in these range of parameters (Olsen et al. 2017). From these stellar models, we obtained the respective $L_{\rm FUV}$ (calculated as the average luminosity between 912 and 2066 Å, $6-13.6$ eV) for each metallicity and age. Then we interpolated these values to estimate the $L_{\rm FUV}$ for each of the star particles in the galaxy.

The local radiation field is then defined as

$$\frac{G_{0,{\rm loc}}}{{\rm erg\,cm^{-2}s^{-1}}} = \sum_{|r_{\rm gas} - r_{*,i}| < h} \frac{L_{\rm FUV,i}}{4\pi |r_{\rm gas} - r_{*,i}|^2} \frac{m_{*,i}}{10^4{\rm M}_{\odot}}, \tag{2.13}$$

where $|r_{\rm gas} - r_{*,i}|$ is the distance between the gas and star particles, $m_{*,i}$ is the stellar mass and $L_{\rm FUV,i}$ the value of the FUV luminosity for each star particle. We then defined the total ISRF hitting the outer part of the neutral cloud as

$$G_{0,{\rm cloud}} = G_{0,{\rm bkg}} + G_{0,{\rm loc}}, \tag{2.14}$$

* We adopted the Geneva stellar models (Schaller et al. 1992) with standard mass loss.

† The star formation time t_* is used to calculate the age of the stars (i.e. lookback time).

where we have normalised the value of $G'_{0,\rm bkg}$ by $G_{0,\rm MW}$ (i.e. $G'_{0,\rm bkg} = G_{0,\rm bkg}/G_{0,\rm MW}$), using $G_{0,\rm MW} = 9.6 \times 10^{-4}$ erg cm^{-2} s^{-1} (Seon et al. 2011). Typical values for $G_{0,\rm cloud}$ are in the range ~ 0.1–10^8. The low ranges (~ 0.1) are regions where the gas particles are only affected by $G_{0,\rm bkg}$, while at high ranges ($\sim 10^8$) $G_{0,\rm loc}$ is dominant as the stars particles are very close to the gas particles.

Neutral clouds

Following the work of Olsen et al. (2015), we identified two phases in the environments of the neutral gas clouds that we analyse in this work, the dense molecular gas and the neutral atomic gas. To determine the properties of these phases, the neutral mass $m_{\rm neutral}$ is divided into the neutral clouds by sampling the mass function as observed in the Galactic disk and Local Group:

$$\frac{dN}{dm_{\rm cloud}} \propto m_{\rm cloud}^{-\beta}, \tag{2.15}$$

with $\beta = 1.8$ (Blitz et al. 2007). We applied a lower cut on the mass of $10^4\,{\rm M}_\odot$ and an upper cut of $10^6\,{\rm M}_\odot$ (Narayanan et al. 2008a,b). The remaining gas below the lower limit of $10^4\,{\rm M}_\odot$ is discarded and so it goes to the ionised mass ($m_{\rm ionised}$); in general this is below 1% of the mass. The neutral clouds were randomly distributed within $0.5h$, with the radial displacement scaling inversely with $m_{\rm cloud}$, to conserve the mass distribution in the galaxy. Changing this radial distribution limit ($0.5h$) affects the final luminosities minimally.

We used the entropy to obtain the external pressure (Eq. 2.10) and to calculate the radius of the neutral cloud ($R_{\rm cloud}$). We took the relative contributions by the cosmic-ray (CR) and magnetic pressure into account, where $\alpha_0 = 0.4$ and $\beta_0 = 0.25$ following Elmegreen (1989):

$$P_{\rm ext} = \frac{P_{\rm Tot}}{1 + \alpha_0 + \beta_0}. \tag{2.16}$$

$R_{\rm cloud}$ is then obtained following the virial theorem from a pressure normalised mass-size relation as (Field et al. 2011; Hughes et al. 2013; Faesi et al. 2018)

$$\frac{R_{\rm cloud}}{\rm pc} = \left(\frac{P_{\rm ext}/k_B}{10^4 {\rm cm}^{-3}{\rm K}}\right)^{-1/4} \left(\frac{m_{\rm cloud}}{290 {\rm M}_\odot}\right)^{1/2}, \tag{2.17}$$

resulting in sizes of $R_{\rm cloud} \approx 1$–$300\,{\rm pc}$ (Sect. 2.2.4). For the density inside the neutral cloud, we used a gas distribution described by a Plummer profile:

$$n({\rm H})(R) = \frac{3 m_{\rm cloud}}{4\pi R_p^3} \left(1 + \frac{R^2}{R_p^2}\right)^{-5/2}, \tag{2.18}$$

with $R_p = 0.1 R_{\rm cloud}$. Adopting this density profile ensures a finite central density. In addition, Popping et al. (2019) showed that this profile is better at reproducing the [C II] luminosity in galaxies from $z = 0$ to $z = 6$ compared to other distributions (power law, logotropic and constant density profiles). With these values, we described the structure for the neutral clouds. In addition, the neutral clouds inherit some physical parameters (e.g. Z, $G'_{0,\rm bkg}$, X_C among others) from the SPH particle.

The [C II] emission arises from within the neutral clouds except from the inner core region, which is shielded from FUV radiation. We calculated the radius where the abundances of C and C^+ are equal ($R_{\rm CI}$) and assumed that all the atoms inside this radius are neutral, so the emission of [C II] in that region is zero. We followed the steps described in Olsen et al. (2015), who used the approach of Röllig et al. (2006) with the following reactions for the formation and destruction of C^+:

$$\mathrm{C} + \gamma \longrightarrow \mathrm{C}^+ + e^- \tag{2.19}$$
$$\mathrm{C}^+ + e^- \longrightarrow \mathrm{C} + \gamma \tag{2.20}$$
$$\mathrm{C}^+ + \mathrm{H}_2 \longrightarrow \mathrm{CH}_2^+ + \gamma. \tag{2.21}$$

We solved $R_{\rm CI}$ with the equation

$$5.13 \times 10^{-10} s^{-1} \chi G_{0,\rm cloud} = n(\mathrm{H})(R_{\rm CI})\,[a_C X_C + 0.5 k_C] \tag{2.22}$$

with

$$\chi = \int_1^\infty \frac{\exp\left(-\mu \xi_{\rm FUV} A_V(R_{\rm CI})\right)}{\mu^2} d\mu, \tag{2.23}$$

where the left-hand side of Eq. 2.22 is C^+ formation (Eq. 2.19) and the right-hand side is C^+ destruction (Eqs. 2.20 and 2.21). We used the same constants as Olsen et al. (2015) for the recombination and radiative association rate coefficients: $a_C = 3 \times 10^{-11} \mathrm{cm^3 s^{-1}}$ and $k_C = 8 \times 10^{-16} \mathrm{cm^3 s^{-1}}$ (Pelupessy & Papadopoulos 2009). The normalisation constant ($5.13 \times 10^{-10} s^{-1}$) comes from the ionisation rate in the photo-dissociation region (PDR) model by Röllig et al. (2006). In the integral, the isotropic radiation field is accounted by $\mu = 1/\cos\theta$, where θ is the angle between the Poynting vector and the normal direction. The visual extinction correspondingly is defined as $A_V(R_{\rm CI}) = 0.724 \sigma_{\rm dust} Z' \langle n(\mathrm{H}) \rangle R_{\rm cloud} \ln\left(R_{\rm cloud}/R_{\rm CI}\right)$ (Pelupessy & Papadopoulos 2009), where $\sigma_{\rm dust} = 4.9 \times 10^{-22} \mathrm{cm^2}$ is the FUV dust absorption cross section (Mezger et al. 1982), Z' is the mean metallicity of the galaxy and $\langle n(\mathrm{H}) \rangle$ is the average density inside the neutral cloud. The difference in the opacity between visual and FUV light is set to $\xi_{FUV} = 3.02$ (Röllig et al. 2006). The carbon abundance relative to hydrogen X_C comes from the carbon mass fraction of the parent SPH particle. For simplicity, Röllig et al. (2006) assumed that the density of electrons is similar to that of carbon to obtain Eq. 2.22.

As these calculations are computationally expensive, we created a grid of four variables to solve for $R_{\rm CI}$, using the following ranges: $4 \leq \log(m_{\rm cloud}[M_\odot]) \leq 6$, $18.5 \leq \log(R_{\rm cloud}[cm]) \leq 21$, $-1.5 \leq \log(G_{0,\rm cloud}) \leq 8$ and $-7 \leq \log(X_C) \leq -1.5$, all in steps of 0.125 dex. With this grid of ∼1.2 million points, we obtained a solution for $R_{\rm CI}$ for each neutral cloud.

To differentiate neutral atomic gas and dense molecular gas, we needed to define a radius at which molecular hydrogen dominates. We calculated the molecular H_2 fraction following Gnedin & Kravtsov (2011) as

$$f_{\mathrm{H}_2} = \left(1 + \frac{\tilde{\Sigma}}{\Sigma_{HI+H_2}}\right)^{-2}, \tag{2.24}$$

where

$$\tilde{\Sigma} = 20 \mathrm{M}_{\odot} pc^{-2} \times \frac{\Lambda^{4/7}}{D_{\mathrm{MW}}} \frac{1}{\sqrt{1 + U_{\mathrm{MW}} D_{\mathrm{MW}}^2}}, \tag{2.25}$$

$$\Lambda = \ln\left[1 + g D_{\mathrm{MW}}^{3/7} \left(U_{\mathrm{MW}}/15\right)^{4/7}\right], \tag{2.26}$$

$$g = \frac{1 + \alpha s + s^2}{1 + s}, \tag{2.27}$$

$$s = \frac{0.04}{D_* + D_{\mathrm{MW}}}, \tag{2.28}$$

$$\alpha = 5 \frac{U_{\mathrm{MW}}/2}{1 + \left(U_{\mathrm{MW}}/2\right)^2}, \tag{2.29}$$

$$D_* = 1.5 \times 10^{-3} \ln\left[1 + \left(3 U_{\mathrm{MW}}\right)^{1.7}\right], \tag{2.30}$$

the metallicity of gas in solar units is $D_{\mathrm{MW}} = Z_{\mathrm{gas}}/Z_{\odot}$, and the local UV background is $U_{\mathrm{MW}} = \mathrm{SFR}/\mathrm{SFR}_{\mathrm{MW}}$. Then the molecular fraction can be used for determining the radius at which the transition from atomic to molecular H occurs ($R_{\mathrm{H_2}}$) in a Plummer profile (Eq. 2.18):

$$f_{\mathrm{H_2}} = \left(\frac{R_{\mathrm{H_2}}}{R_{\mathrm{cloud}}}\right)^3 \left(\frac{R_{\mathrm{p}}^2 + R_{\mathrm{cloud}}^2}{R_{\mathrm{p}}^2 + R_{\mathrm{H_2}}^2}\right)^{3/2}. \tag{2.31}$$

This information helped us to calculate the density and temperature at $R_{\mathrm{H_2}}$ in the dense molecular gas and neutral atomic gas regions.

Diffuse ionised gas

For the diffuse ionised gas (DIG), we followed a slightly different approach than that used by Olsen et al. (2015, 2017), where they assumed a DIG cloud with a radius equal to the smoothing length (h). We wanted to avoid dependency on the simulation's resolution and overproduction of DIG in SPH particles with very large h ($\gtrsim$ 5 kpc in 100 Mpc boxes, Sect. 2.2.4). We instead calibrated our models with the estimated luminosity of [N II] lines at 122 and 205 μm emitted almost entirely from the ionised medium. By comparing these lines it is possible to deduce the fractions of the ionised gas (Croxall et al. 2017). For example [N II] at 122 μm is associated with the DIG rather than with a compact H II region (Cormier et al. 2012).

To calibrate the DIG fraction in the [C II] line, we created a distribution of radii of the DIG (R_{DIG}) assuming an isothermal sphere with uniform density. We assumed that the distribution function behaves as a smoothed broken power law for R_{DIG}:

$$p(R_{\mathrm{DIG}}) = \left(\frac{R_{\mathrm{DIG}}}{R_b}\right)^{-\alpha_1} \left\{\frac{1}{2}\left[1 + \left(\frac{R_{\mathrm{DIG}}}{R_b}\right)^{1/\Delta}\right]\right\}^{(\alpha_1 - \alpha_2)\Delta}, \tag{2.32}$$

Table 2.3 – *Grids for the calibration of Eq. 2.32 used in this work. The resulting total number of grid points is 4131.*

Parameter	Unit	Min.	Max.	Interval
$\log R_b$	[kpc]	−0.8	0.5	0.026
α_1	–	−5.0	−1.0	0.5
α_2	–	1.0	5.0	0.5

where R_b is the break radius, α_1 is the power law index for $R_{\rm DIG} \ll R_b$, α_2 is the power law index for $R_{\rm DIG} \gg R_b$ and Δ is the smoothness parameter. We created a grid of different values as described in Table 2.3 for the parameters in Eq. 2.32. We fixed the Δ parameter to 0.1 to achieve a distribution with a smooth peak. We estimated the line luminosity for the given grid of values in a random sample (20%) of galaxies from Ref-L0100N1504. We only used Ref-L0100N1504 for the DIG calibration as the small simulation boxes do not contain enough galaxies to compare with the observational sample.

We used observational data from Fernández-Ontiveros et al. (2016) to calibrate the contribution of the DIG. We used the $L_{\rm [N\,II]122}$–SFR, $L_{\rm [N\,II]205}$–SFR and $L_{\rm [N\,II]122}/L_{\rm [N\,II]205}$–SFR relations to compare the observational and simulated samples. We considered only cases with a SFR inside the values that the selected simulation could achieve ($-2.8 < \log({\rm SFR}) < 1.7$), for a total of 69 observed galaxies. The two datasets (observational and simulated) were binned by SFR into identical bins. We calculated the mean and standard deviation in each bin to compare them using a chi-square (χ^2) test. We noticed that χ^2 values were not significantly affected by the α parameters, so we selected $\alpha_1 = -2.0$ and $\alpha_2 = 1.0$, assuming that the DIG size is larger than the neutral clouds. We found that the R_b value that minimises the χ^2 is $R_b \approx 900$ pc.

In Fig. 2.1 we present the $L_{\rm [N\,II]}$–SFR relations at $z = 0$ from the simulations used in this work compared with the sample of galaxies with [N II] used in the DIG calibration from Fernández-Ontiveros et al. (2016). We recover the same trend of the observed [N II] line luminosities with Ref-L0100N1504. The DIG calibration shows a good agreement for simulations with the same resolutions (Ref-L0100N1504 and Ref-L0025N0376) but Recal-L0025N0752 is off by less than 1 dex, which is related to the re-calibration (see discussion in Sect. 2.4.4).

We randomly sampled the DIG radii using Eq. (2.32) for each galaxy. We used the assigned DIG radii and the ionised mass ($m_{\rm ionised}$) to calculate the density of this ISM phase for each SPH particle. These lead us to densities ranging from around 10^{-6} to 3 cm^{-3} for the DIG (due to the limits in Cloudy cooling tables, Table 2.4), with a peak at 10^{-2}cm^{-3} (see results in Sect. 2.2.4).

2.2.3 Line luminosity prediction

We predicted line luminosities with the help of Cloudy v17.01 (Ferland et al. 2017), using a set of cooling tables presented by Ploeckinger & Schaye (2020). Cloudy is a 1D spectral synthesis code that predicts atomic and molecular line intensities in different environments using an escape probability formalism. Here we describe briefly the important aspects of the cooling tables used.

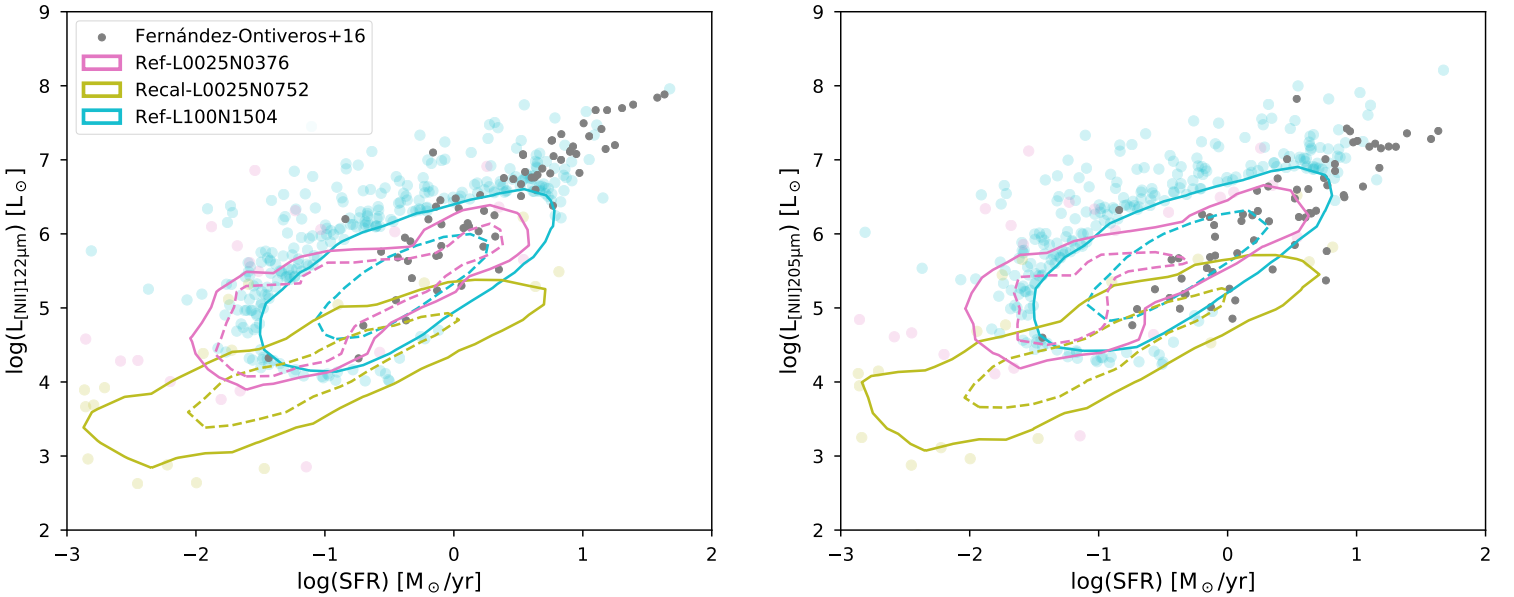

Figure 2.1 – *$L_{[\mathrm{N\,II}]122}$–SFR (left) and $L_{[\mathrm{N\,II}]205}$–SFR (right) relations for the three simulations used (*Ref-L0025N0376, Recal-L0025N0752, *and* Ref-L0100N1504*) and the observational sample used for the DIG calibration from Fernández-Ontiveros et al. (2016) (grey dots). The smoothed contours show where 90% (solid) and 50% (dashed) of the galaxies from the respective simulations lie. The remaining 10% of galaxies are represented as dots using the same colours.*

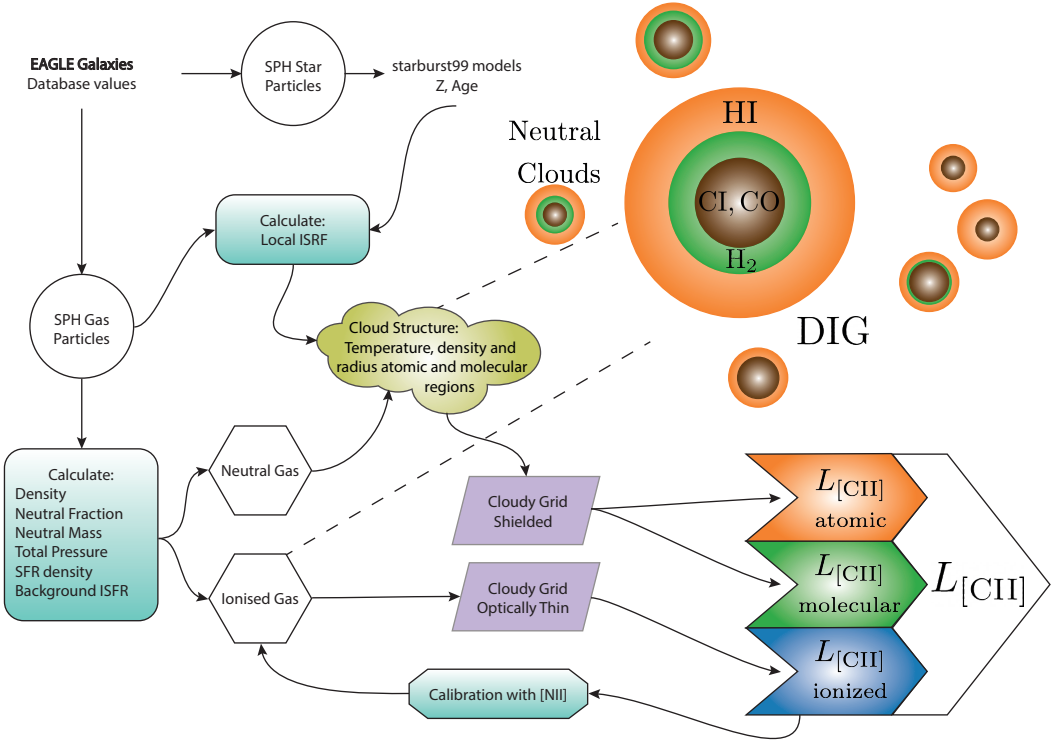

Figure 2.2 – *Flowchart of the sub-grid procedures applied to the SPH simulation to simulate line emission in post-processing. For each* EAGLE *simulation, we obtain the galaxy information from the database and we use the star and gas particle data to implement the ISM model. Next, we calculate physical properties in each phase to obtain the neutral clouds and DIG structures. We use* Cloudy *cooling tables (Ploeckinger & Schaye 2020) to get the line luminosity for [C II]. We first calculate $L_{[C\,II],DIG}$ and then calibrate the ionised gas using the predicted luminosities of the [N II] lines. We then obtain the total line luminosity from the luminosities of the different ISM phases. The dashed lines connect the gas environment to the ingredients of the model.*

An equally spaced grid was used in the dimensions redshift z, gas temperature $\log T$, metallicity $\log Z$, and gas density $\log n(\mathrm{H})$, as presented in Table 2.4. In this work, we only made use of the redshift bin at $z = 0$. The importance of using the redshift as a parameter resides on the non-negligible effect on the line emissivity from the cosmic microwave background (CMB). The spin temperature of the [C II] emission is similar to the CMB temperature at high redshift in low density environments (WNM), affecting the [C II] emission. In contrast, at high densities (CNM) the spin temperatures are larger, so the [C II] emission is not affected (Vallini et al. 2015; Pallottini et al. 2017a; Lagache et al. 2018; Olsen et al. 2018a). However, for this study at $z = 0$, the CMB is not important.

Cloudy is used to propagate the incident radiation in a plane-parallel gas until the column density N_{sh} is reached. For gas with temperatures of $T \leq 10^3\,\mathrm{K}$ the shielding column N_{sh} is assumed to be equal to one half of the Jeans column density

$$\log N_J[\mathrm{cm}^{-2}] = 19.44 + 0.5 \times \big(\log n(\mathrm{H})[\mathrm{cm}^{-3}] + \log T[\mathrm{K}]\big) \tag{2.33}$$

to model the shielding from the edge to the centre of a self-gravitating gas cloud with an extent equal to the Jeans length λ_{J}. For higher temperatures, the column density of the self-shielding gas asymptotically approaches that of the optically thin gas and a maximum column density of $N_{\mathrm{max}} = 10^{24}\,\mathrm{cm}^{-2}$ and a maximum length scale of $l_{\mathrm{max}} = 100\,\mathrm{kpc}$ are imposed.

For the radiation field, redshift-dependent contributions from the CMB and UV background (Faucher-Giguère 2020) are applied. In addition, the ISRF ('table ism' in Cloudy) and cosmic rays (CR), scaled to solar neighbourhood values, are added depending on the column density. Solar abundances are assumed, with the abundance ratios modified by dust depletion following Jenkins (2009). Primordial abundances are also calculated (when $\log Z/Z_\odot = -50$) using Planck Collaboration et al. (2016) primordial values for helium (not used in this work).

The 'Orion' grain distribution (from Cloudy) is used to take into account other physical effects from dust (e.g. photoelectric heating and charge and collisional processes) by assuming a dust-to-gas ratio dependent on the metallicity and column density (assuming $\log N_{\mathrm{H}}[\mathrm{cm}^{-2}] = 20.56$ from the gas surface density in the solar neighbourhood, Lagos et al. 2015). Furthermore, polycyclic aromatic hydrocarbons (PAHs) are included but quantum heating is disabled (no 'qheat'). The large (i.e. more detailed) H2 model in Cloudy is used and the CR photo-dissociation rate is re-scaled to match the UMIST database* values (see also Shaw et al. 2020).

The two cooling tables we used in this work, ideal for cosmological simulations, have the ISRF and CR rate values reduced by 1 dex relative to the MW values (Black 1987; Indriolo et al. 2007) to better match the observations of the transition between atomic and molecular hydrogen (Ploeckinger & Schaye 2020). In other words, physical parameters decrease: the SFR surface density Σ_{SFR} decreases from $10^{-3}\,\mathrm{M}_\odot\,\mathrm{yr}^{-1}\,\mathrm{kpc}^{-2}$ to $10^{-4}\,\mathrm{M}_\odot\,\mathrm{yr}^{-1}\,\mathrm{kpc}^{-2}$, and the CR hydrogen ionisation rate $\log\zeta$ decreases from $-15.7\,\mathrm{s}^{-1}$ to $-16.7\,\mathrm{s}^{-1}$. The first table includes self-shielding and the second table is

* udfa.net

Table 2.4 – *Sampling of gas properties in the* Cloudy *grid used in this work. The resulting number of grid points is 610 for the shielded and optically thin calculations.*

Parameter	Unit	Min.	Max.	Interval
$\log Z$	$[Z_\odot]$	−4.0	0.5	0.5
$\log n(\mathrm{H})$	$[\mathrm{cm}^{-3}]$	−6.0	6.0	0.2

the optically thin counterpart (1-zone unattenuated cloud) of the first table*. The latter provides us with the emissivities of the DIG, approximating this phase as optically thin. We sample and limit the grids to the values presented in Table 2.4.

We used thermal equilibrium temperatures, where cooling is equal to heating, which depend mainly on two variables $n(\mathrm{H})$ and Z. The temperatures of the gas phases were then obtained with a linear interpolation of the grid with density and metallicity (as other parameters like ISRF or CR depend on these) for the neutral atomic gas, dense molecular gas and DIG phases. This assumption is probably incorrect for the DIG, as ionised regions could be outside thermal equilibrium, because DIG gas can have higher temperatures, but temperature values $\approx 10^4$ K are commonly assumed when modelling ionised regions (e.g. Haffner et al. 2009; Vandenbroucke & Wood 2019; Ferrara et al. 2019; Ploeckinger & Schaye 2020). For atomic and molecular gas, thermal equilibrium is commonly used in PDR models (Tielens & Hollenbach 1985; Hollenbach & Tielens 1999).

The [C II] line luminosity is then computed from the integral of the [C II] emissivity, $\Lambda_{[\mathrm{C\,II}]}$, interpolated from the Cloudy cooling table as described above, through the equation

$$L_{[\mathrm{C\,II}]} = \int \Lambda_{[\mathrm{C\,II}]} dV = 4\pi \int_{R_1}^{R_2} \Lambda_{[\mathrm{C\,II}]} R^2 dR. \tag{2.34}$$

The limits of the integral are defined by the three components of the ISM model, DIG, atomic gas and molecular gas; therefore, $R_1 = 0$ and $R_2 = R_{\mathrm{DIG}}$ for the DIG, $R_1 = R_{\mathrm{H_2}}$ and $R_2 = R_{\mathrm{cloud}}$ for the atomic gas and $R_1 = R_{\mathrm{CI}}$ and $R_2 = R_{\mathrm{H_2}}$ for the molecular gas. Ten shells are used to estimate the luminosities for the atomic gas and molecular gas where we assumed that the density follows a Plummer profile (Eq. 2.18). In the case of the DIG, one shell is used because we assume a homogeneous density distribution.

2.2.4 Summary and verification

To summarise, we illustrate the path from the EAGLE simulations (Sect. 2.2.1) to the total luminosity of [C II] (Sect. 2.2.3) in Fig. 2.2. First, we selected a sample of galaxies from the EAGLE database and retrieved the gas and star particle data of the sample. Second, we calculated physical properties such as density and neutral fraction for the gas particles. Local ISRF was calculated with `starburst99` models derived from star particles, while background ISFR came from the SFR surface density. The

* The optically thick and optically thin tables are the UVB_dust1_CR1_G1_shield1 and UVB_dust1_CR1_G1_shield0 in Ploeckinger & Schaye (2020), respectively.

neutral gas mass and the ISRF were used to create the neutral cloud structures inside galaxies, while the ionised gas mass is used to create the DIG. After calibration with the [N II] lines for the DIG, we obtained the luminosity of [C II] line for each phase using CLOUDY cooling tables (Ploeckinger & Schaye 2020). Finally, we calculated the total [C II] luminosity in a galaxy with the contributions from the three ISM phases combined.

We plot the distribution of the sizes for the neutral clouds and DIG (described in Sect. 2.2.2) in Fig. 2.3. The values for $R_{\rm H_2}$ show that the molecular regions are in general smaller than $R_{\rm CI}$. As the neutral clouds are modelled with a Plummer profile (Eq. 2.31), the molecular fraction will be restricted to the centre of the neutral clouds at around one pc, so the emission coming from these regions will be less than the atomic gas emission in most cases. In cases where $R_{\rm H_2}$ is very small ($f_{\rm H_2}$ near zero and $R_{\rm CI} > R_{\rm H_2}$), the atomic phase will dominate between $R_{\rm cloud}$ and $R_{\rm CI}$. For the case when the shielding is efficient ($f_{\rm H_2} \gg 0$), the atomic gas goes from $R_{\rm cloud}$ to $R_{\rm H_2}$ and the molecular goes from $R_{\rm H_2}$ to $R_{\rm CI}$. Therefore, $R_{\rm CI}$ is the dominant internal bound for the atomic region in most cases. We sketch these radii in the different neutral clouds in Fig. 2.2.

Typical sizes for these neutral clouds ($\sim 0.5 - 200$ pc) agree with observed values in the MW (e.g. Murray 2011; Miville-Deschênes et al. 2017) and in the local Universe (e.g. Bolatto et al. 2008), as presented in the upper panel of Fig. 2.3. In the case of the DIG, we compared two different cases, using the smoothing length h or Eq. 2.32. When using h, the DIG sizes can go up to ~ 20 kpc, which is the size of some galaxies in the simulations, while with $R_{\rm DIG}$ from Eq. 2.32, sizes are between 100 pc and 10 kpc with a peak at around 0.9 kpc. Again, in Fig. 2.2 we sketch, as an example, the DIG as a volumetric region around the neutral clouds.

In Fig. 2.4 we show the densities and temperatures obtained with the gas particles for the different ISM phases, as well as the initial SPH gas and density from the simulation. Most of the initial gas density fills the area above 10^4K and below ~ 0.1 cm^{-3}, while the remaining initial gas is distributed along the equation-of-state threshold imposed by EAGLE. The initial gas distribution in the temperature-density plane shows how important it is to implement a physically motivated model for the different ISM phases where EAGLE is incapable of reaching. With our model, the DIG density runs from the lowest density (10^{-6}) to ~ 3 cm^{-3} with a peak around 0.01 cm^{-3}, the atomic density runs from 10^{-1} to $10^{3.5}$ cm^3 with a peak at 1 cm^{-3}, and the molecular gas density runs from 10 cm^{-3} to 10^6 cm^{-3} with a peak at 10^5 cm^{-3}. For the temperatures, the DIG range is between $10^{3.2}$ K and $10^{4.9}$ K, with a peak around 10^4 K, in the atomic the temperatures vary from 10^1 K and 10^4 K with two peaks around 60 K and 5000 K, and the molecular gas temperature is constrained to a small region between 10 to 300 K due to the H_2 heating processes.

Comparing Fig. 2.4 with recent simulations of individual local-like galaxies (MW and M51) from Tress et al. (2020a,b), we find similar locations for the atomic gas phases and thermal stable regimes ($T \sim 100$ K and $T \sim 10^4$ K). Molecular regions are located in the same regime ($T \sim 20$ K and $n({\rm H}) > 10^2$ cm^{-3}) as well as diffuse ionised gas ($T \sim 10^4$ K and $n({\rm H}) < 1$ cm^{-3}). The only difference is on the step transition we have between the thermal stable regimes. With our CLOUDY grids, the

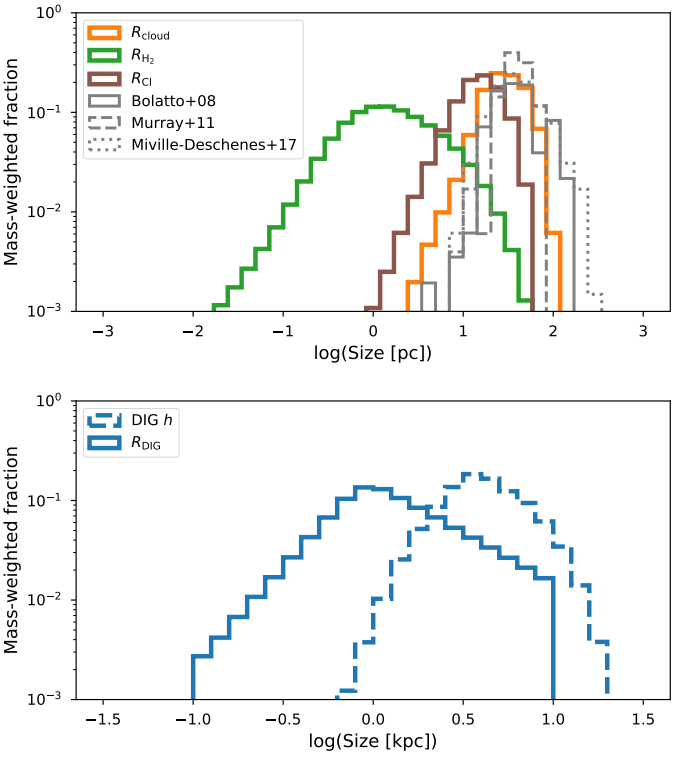

Figure 2.3 – *Mean mass-weighted distribution of the sizes of the ISM phases for* RECAL-L0025N0752. *Upper panel: Radii defining the limits of the phases inside the neutral clouds:* $R_{\rm cloud}$, $R_{\rm H_2}$ *and* $R_{\rm CI}$. *We compare with sizes of MW clouds (solid and dashed grey lines, Murray 2011; Miville-Deschênes et al. 2017) and local Universe clouds (Bolatto et al. 2008, dotted grey lines). Lower Panel: Assumed mean mass-weighted distribution of the radii of the DIG using the smoothing length* h *(dashed) or Eq. 2.32 (solid) to define* $R_{\rm DIG}$.

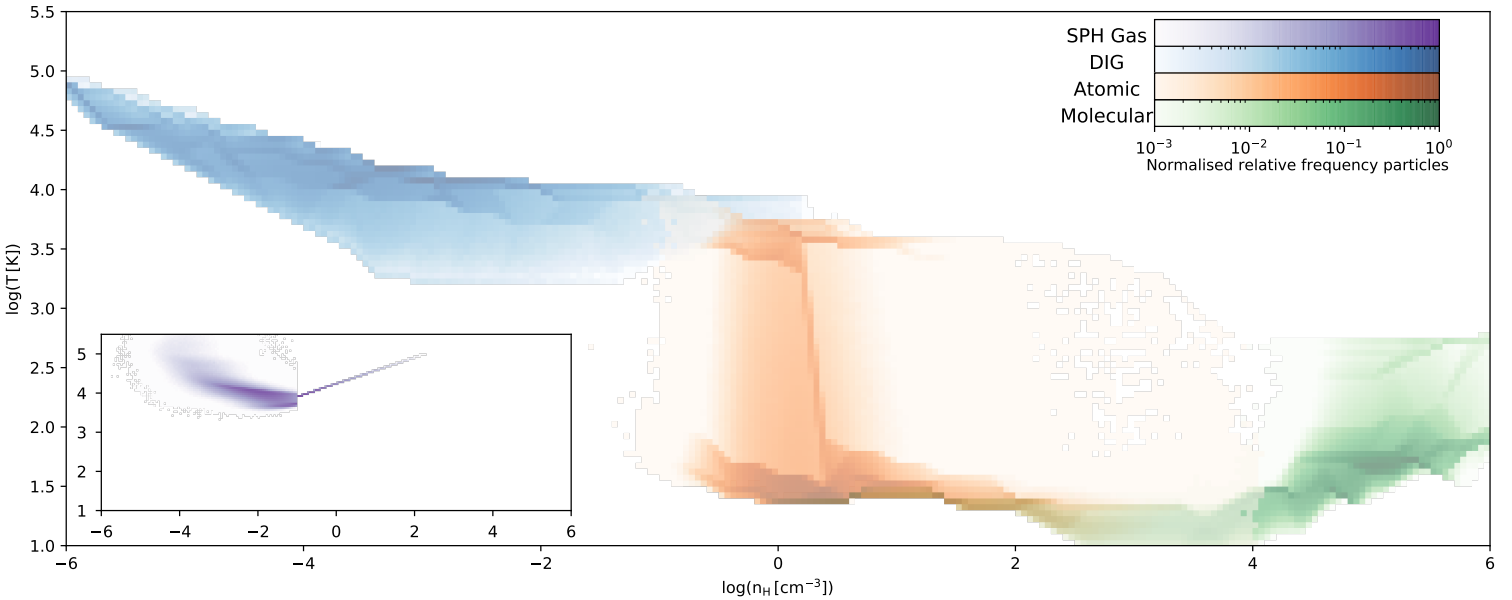

Figure 2.4 – *Temperature vs density for the three ISM phases in the* RECAL-L0025N0752 *galaxies. Each phase is represented with a normalised relative frequency between 0.1% and 100%. We observe that each phase is located in a specific region of the temperature vs density plane. The DIG (blue) is characterised by high temperatures ($\log T[\mathrm{K}] \gtrsim 3.5$) and low densities ($\log \mathrm{n(H)[cm^{-3}}] \lesssim -1$), atomic gas (orange) by intermediate densities ($-1 \lesssim \log \mathrm{n(H)[cm^{-3}}] \lesssim 3$) with a bridge between high temperatures ($\log T[\mathrm{K}] \approx 3.5$) and low temperatures ($\log T[\mathrm{K}] \approx 1.4$) and molecular gas (green) by high densities ($\log \mathrm{n(H)[cm^{-3}}] \gtrsim 3$) and a range of very low temperatures ($\log T[\mathrm{K}] \approx 1.1$) to intermediate temperatures ($\log T[\mathrm{K}] \approx 2.5$) due to H_2 heating processes (Bialy & Sternberg 2019). The temperature-density relation of the original SPH gas particles (inset plot) is shown in purple to demonstrate the need to use a more detailed ISM model when predicting FIR line strengths from these simulations.*

temperatures change from 10^4 K to 100 K over less than one order of magnitude in density, while Tress et al. (2020a,b) results show a smooth transition that covers two orders of magnitude in density. This transition depends on the assumed metallicity and radiation field (Bialy & Sternberg 2019). Therefore, the difference between our results and Tress et al. (2020a,b) may be related to different implementations in terms of numerical methods and chemical evolution, affecting the ISRF and metallicity.

2.3 Results

2.3.1 Contributions of ISM phases to [C II] emission

Various studies report that the atomic phase dominates the fractional contribution of $L_{[\mathrm{C\,II}]}$ in the MW and the local Universe (Pineda et al. 2014; Cormier et al. 2019; Abdullah & Tielens 2020). For example, Cormier et al. (2019) calculate the physical parameters of the Herschel Dwarf Galaxy Survey (HDGS) sample with a mix of two models, for low and high ionisations. They assume that only the dense H II model (high ionisation) produces an atomic gas phase. In terms of their atomic gas fraction there is an average increase in their sample from ~ 0.2 to ~ 0.5 when $\log(\mathrm{SFR})\,[\mathrm{M_\odot yr^{-1}}]$ changes from ~ -3 to ~ 0. Pineda et al. (2014) calculated the ISM contributions in the MW galactic plane where molecular, atomic, and ionised gas, contribute to 25%, 55%, and 20% of the [C II] luminosity, respectively. However, the contribution of the DIG phase is difficult to estimate accurately. For example, the DIG phase can be more vertically extended in disk galaxies such as the MW, which can cause a difference in the estimated DIG contribution from 4% to 20% as noted by Pineda et al. (2013, 2014). Finally, Abdullah & Tielens (2020) studied the Orion-Eridanus super-bubble and found that the surfaces of molecular clouds, mostly atomic gas, are the main contributors (80%) to the [C II] emission.

We separate the contributions from different ISM phases to $L_{[\mathrm{C\,II}]}$ as a function of SFR. We present the median fractional contribution (the luminosity contribution of a given phase to the total luminosity of the galaxy) of these ISM phases in Fig. 2.5 for the three simulations we have used. These results show the fractional contribution of each of the ISM phases presented in Fig. 2.5 varies with the simulation used and the total SFR of the galaxy.

We see a general agreement in terms of qualitative trends among the three simulations in Fig. 2.5. In all simulations, the fractional contributions from the neutral phases increase with increasing SFR, especially for the molecular phase. On the other hand, the DIG fractional contribution decreases with increasing SFR and then flattens, or increases again for Ref-L0100N1504. However, the three simulations also exhibit some differences in quantitative terms. The relatively small differences between the two intermediate simulations, Ref-L0025N0376 and Ref-L0100N1504, are due to sampling noise as the two simulations only differ in size. In particular, the larger box simulation provides a better sampling of the brighter and rarer galaxy population with higher SFRs. The differences between the intermediate and high-resolution simulations are much larger and are due to the GSMF re-calibration, which changes the feedback parameters in the simulation (Schaye et al. 2015). This re-calibration leads to different intrinsic characteristics for the galaxies between the simulations:

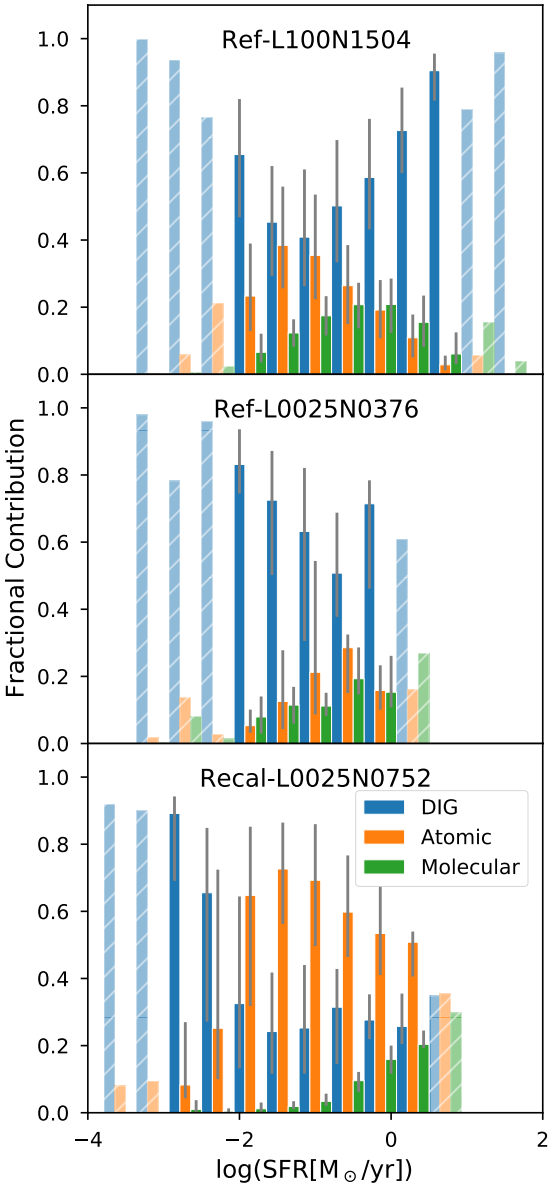

Figure 2.5 – *Fractional contributions of the ISM phases to the total $L_{[\mathrm{C\,II}]}$ for* Ref-L0100N1504 *(top panel),* Ref-L0025N0376 *(middle panel), and* Recal-L0025N0752 *(bottom panel) simulations. Shaded, white-diagonal-striped bars indicate bins with fewer than ten galaxies. The grey lines show the range between the 25th and 75th percentiles. Colours are the same as in Fig. 2.4.*

The stellar feedback produces more outflows of metal-enriched material in Recal-L0025N0752 (compared to Ref-L0100N1504 and Ref-L0025N0376), reducing the total metallicity in these galaxies (Schaye et al. 2015; De Rossi et al. 2017). In

addition, the effect of AGN feedback is important for changing the metallicities of high-mass galaxies, as pointed out by De Rossi et al. (2017). Changes in metallicity are responsible for the differences in the $L_{[C\,II]}$ estimates between these simulations.

We can compare the fractional contributions in RECAL-L0025N0752 with observational studies where most of the [C II] comes from the atomic phase instead of the DIG. In RECAL-L0025N0752 the atomic phase becomes increasingly more important with increasing SFR like in Cormier et al. (2019), with smaller contributions from the dense molecular gas and the DIG. However, in this work we do not model the H II regions, so our results are not entirely comparable with Cormier et al. (2019). Assuming a $\log(\mathrm{SFR})\,[\mathrm{M_\odot yr^{-1}}] \sim 0$ for the MW (Robitaille & Whitney 2010), the ISM contributions from RECAL-L0025N0752 at this SFR are 14^{+4}_{-2}%, 55^{+13}_{-13}% and 25^{+12}_{-3}% for the molecular, atomic and ionised gas, similar to the fractions presented by Pineda et al. (2014). Similar results are found in nearby dwarf galaxies (Fahrion et al. 2017; Cormier et al. 2019) and spiral galaxies (Abdullah et al. 2017). Thus, results from RECAL-L0025N0752 agree well with observations in the local Universe.

2.3.2 The [C II]–SFR relation

In this section, we compare the well-known relation between $L_{[C\,II]}$ and SFR in galaxies (De Looze et al. 2014) with that obtained with our model. As we mentioned before (Sect. 2.1), [C II] luminosity can be used as a SFR indicator (e.g. Stacey et al. 1991, 2010; Malhotra et al. 2001). However, at higher IR luminosities the [C II] luminosities are lower than expected, this effect is known as the '[C II] deficit'. Various reasons has been proposed (Muñoz & Oh 2016; Narayanan & Krumholz 2017; Smith et al. 2017; Díaz-Santos et al. 2017; Ferrara et al. 2019) pointing mainly to local conditions of the interstellar gas as the drivers of this deficit, such as the intensity of the radiation field, metallicity or gas density. However, the impact from other physical parameters such as dust temperatures (Pineda et al. 2018) or the AGN (Herrera-Camus et al. 2018b) are still not clear. In Fig. 2.6 we see the [C II] deficit when we compare the observational sample with the power law derived by De Looze et al. (2014): at higher SFR $\gtrsim 10\,\mathrm{M_\odot\,yr^{-1}}$ the slope with $L_{[C\,II]}$ is less steep than at low SFR $\lesssim 10\,\mathrm{M_\odot\,yr^{-1}}$ (FIR luminosities $\lesssim 10^{11}\mathrm{L_\odot}$). The $L_{[C\,II]}$–SFR relation is complex and a simple power law (like the one implemented by De Looze et al. (2014)) cannot describe it (Ferrara et al. 2019). This trend is also observed in other FIR lines as presented by Díaz-Santos et al. (2017).

The $L_{[C\,II]}$–SFR relation we inferred from the ISM model is presented in Fig. 2.6 and compared with selected samples of observed galaxies in the local Universe. We briefly describe the observational samples that we compare to in this paper in Appendix 2.5. The multi-phase ISM model we have implemented in this work reproduces the observed galaxy distribution in the $L_{[C\,II]}$–SFR relation from De Looze et al. (2014, entire sample) and Pineda et al. (2014). In the range $-3.5 < \log(\mathrm{SFR})\,[\mathrm{M_\odot yr^{-1}}] < 3$, 77% of RECAL-L0025N0752 galaxies are inside the 1σ dispersion of the De Looze et al. (2014) relation. For the intermediate-resolution simulations (REF-L0025N0376 and REF-L0100N1504), only 47–60% of galaxies are inside the 1σ dispersion. If we compare with the 2σ dispersion around the De Looze et al. (2014) relation, we find 83, 95 and 96% of the galaxies inside the dispersion for REF-L0025N0376, REF-

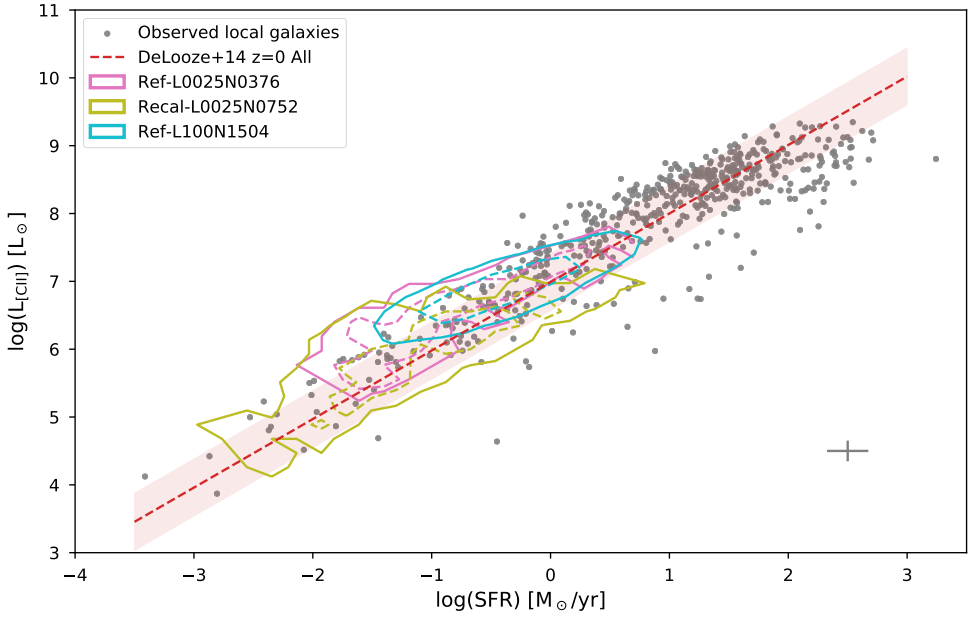

Figure 2.6 – *$L_{[\mathrm{C\,II}]}$–SFR relation for the three simulations used in this work (*Ref-L0025N0376, Recal-L0025N0752, *and* Ref-L0100N1504*) presented as contour maps (pink, olive and cyan, respectively) and an observational sample of local galaxies (grey dots, briefly described in Appendix 2.5). The contours shows where 90% (solid) and 50% (dashed) of the galaxies of the respective simulations fall in the relation. The sample of local galaxies is a compilation of different surveys containing main sequence galaxies (Brauher et al. 2008, ISO Compendium), starburst galaxies (Díaz-Santos et al. 2013, 2017, GOALS), dwarf galaxies (Cormier et al. 2015, 2019, HDGS), star-forming, AGN and LIRG galaxies (Herrera-Camus et al. 2018a,b, SHINING), dusty main sequence galaxies (Hughes et al. 2017, VALES) and intermediate-stellar-mass galaxies (Accurso et al. 2017, xCOLD GASS). We present the mean error from the observational samples in the bottom-right corner of the plot. The dashed red line is the power law derived by De Looze et al. (2014) for the relation at $z = 0$, with the shaded red region representing the 1σ scatter.*

L0100N1504 and Recal-L0025N0752, respectively. The dispersion (1σ) of the simulations (around 0.4 dex) is comparable to the De Looze et al. (2014) relation (0.42 dex). The typical statistical uncertainty in the observed SFR ($L_{[\mathrm{C\,II}]}$) is around 10% (5%) but can go up to 40% (35%) in some cases (e.g. Cormier et al. 2015; Villanueva et al. 2017; Díaz-Santos et al. 2017).

The simulations underestimate the abundance of star-forming galaxies ($1-10\,\mathrm{M_\odot\,yr^{-1}}$) in the local Universe due to the AGN feedback implementations used by EAGLE (Katsianis et al. 2017). In addition, the relatively small volume and physical prescriptions of the simulations (e.g. the IMF, lack of starburst galaxies and sub-grid physics) limit the comparison with luminous IR galaxies (e.g. (U)LIRGs) with SFR above $10\,\mathrm{M_\odot\,yr^{-1}}$ (Wang et al. 2019b). Thus, the results of this work are mainly for galaxies with SFRs below $10\,\mathrm{M_\odot\,yr^{-1}}$. However, caution is need when comparing theoretical results with observational measurements as there could be important systematic differences. For example, the SFRs in EAGLE are computed from the KS law (Schaye & Dalla Vecchia 2008), while observational measurements of SFRs are

normally based on IR luminosity.

Overall, Figs. 2.5 and 2.6 imply that in our simulated galaxies, the dominant phase of $L_{[\mathrm{C\,II}]}$ depends on the star-formation activity of the galaxy and the impact on each of the ISM phases can explain the shape of the $L_{[\mathrm{C\,II}]}$–SFR relation (see discussions in Sects. 2.4.1 and 2.4.2). Although the simulations do not go to SFR $\sim 100\,\mathrm{M_\odot\,yr^{-1}}$, we can study the variations and trends of the physical parameters, which can provide some insights into the [C II] deficit (Sect. 2.4.2).

2.3.3 Pressure and metallicity dependence

To understand the origin of the $L_{[\mathrm{C\,II}]}$–SFR relation, it is important to study how this relation varies as a function of other physical parameters. In this section, we study its dependence on pressure and metallicity.

We plot the $L_{[\mathrm{C\,II}]}$/SFR ratio as a function of pressure in Fig. 2.7. We combine the three simulations. We bin the pressure and $L_{[\mathrm{C\,II}]}$/SFR values with bin sizes determined using Knuth's rule (Knuth 2006), which selects the simplest model that best describes the data by maximising the posterior probability for the number of bins. We colour-code the hexagonal bins by SFR in the range $-2.5 < \log(\mathrm{SFR})\,[\mathrm{M_\odot\,yr^{-1}}] < 1$. We obtain similar results as Popping et al. (2019): At fixed SFR, the ratio of $L_{[\mathrm{C\,II}]}$/SFR decreases with increasing pressure. This trend is expected as the mass-size relation of the neutral clouds in the model (Eq. 2.17) depends on the pressure. At high pressure, the cloud is smaller and the emission region from the [C II] neutral phase decreases. We discuss the neutral cloud size dependency in Sect. 2.4.2.

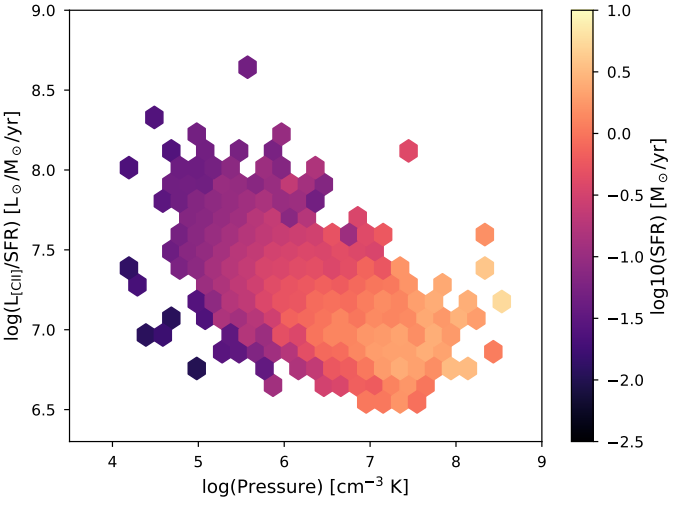

Figure 2.7 – *The $L_{[\mathrm{C\,II}]}$/SFR ratio as a function of pressure, colour-coded by SFR. We present the mean value of SFR for each bin for the galaxies in the three simulations assuming they represent the same population. Pressure is anti-correlated with the $L_{[\mathrm{C\,II}]}$/SFR ratio, in agreement with the result presented by Popping et al. (2019). We have confirmed that the trends do not depend on the specific simulation used.*

We used observational metallicity $(12 + \log(\mathrm{O/H}))$ values from the HDGS (Cormier

et al. 2019), xCOLD GASS (Accurso et al. 2017) and VALES samples (taken from Zanella et al. 2018; Hughes et al. 2017). We converted $L_{\rm IR}$ to SFR as described by Kennicutt & Evans (2012). As presented by Kewley & Ellison (2008), there is a large difference in the type of $12 + \log(\rm O/H)$ calibration used, in some cases up to 0.7 dex in $\log(\rm O/H)$. For example, VALES and xCOLD GASS samples use the calibration from Pettini & Pagel (2004), while the HDGS sample uses a compilation of metallicities (Rémy-Ruyer et al. 2013). In this work, we do not calibrate the observations to a single metallicity calibration. In EAGLE, gas metallicity is defined as the fraction of the gas mass in elements heavier than helium. We use a solar metallicity of $12 + \log(\rm O/H) = 8.69$ or $Z = 0.0134$ (Asplund et al. 2009) to compare EAGLE and observational values.

We show the variation of the $L_{[\rm C\,II]}$/SFR ratio with metallicity for the three simulations in Fig. 2.8. There is no clear trend in the observational data, as we only have a few data points. The $L_{[\rm C\,II]}$/SFR ratio in the simulations show a strong dependence on the mean gas metallicity values. Galaxies in Ref-L0025N0376 and Ref-L0100N1504 are located at higher values from the $L_{[\rm C\,II]}$/SFR ratio compared to Recal-L0025N0752 (as in Fig. 2.6), due to a higher contribution of the DIG phase (see discussion in Sect. 2.4.4). There is a decrease in the $L_{[\rm C\,II]}$/SFR ratio from low metallicities up to $3\,Z_\odot$ in galaxies in Recal-L0025N0752. The total number of observed galaxies and simulated galaxies in the low-metallicity regime (below $\sim 0.2\,Z_\odot$) is not enough to establish a clear trend between metallicity and $L_{[\rm C\,II]}$/SFR as presented by Vallini et al. (2015) and Lupi & Bovino (2020), where there is an increase in the $L_{[\rm C\,II]}$/SFR ratio with metallicity.

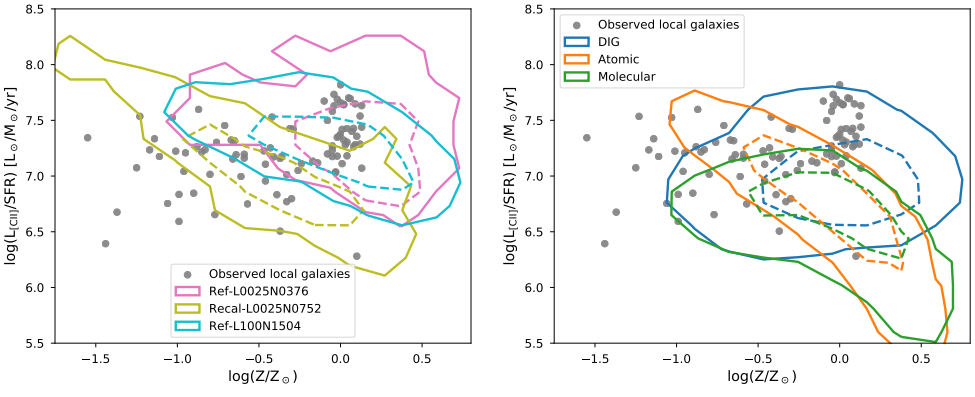

Figure 2.8 – *The $L_{[\rm C\,II]}$/SFR ratio as a function of metallicity. Left panel: Simulations compared with observed galaxies. Contours show where 90% (solid) and 50% (dashed) of the galaxies of the respective simulation lie. Right panel: The different ISM phases (DIG, atomic and molecular) in* Ref-L0100N1504. *For the atomic phase, the $L_{[\rm C\,II]}$/SFR increases with decreasing metallicity. For the molecular phase, it also decreases with increasing metallicity but plateaus at $\sim 1\,Z_\odot$. For the DIG phase, it remains more or less constant.*

For metallicities higher than solar, we see a clear reduction in the $L_{[\rm C\,II]}$/SFR ratio in Ref-L0100N1504; therefore, we plot in the right panel of Fig. 2.8 the different ISM phases in Ref-L0100N1504 as a function of metallicity. For the atomic phase, the $L_{[\rm C\,II]}$/SFR ratio decreases with increasing metallicity. For the molecular

phase, the ratio is more or less flat below solar metallicity, and then decreases with increasing metallicity above solar metallicity. On the other hand, the DIG is flat in the $L_{[\mathrm{C\,II}]}$/SFR ratio at all metallicities. The decrease in the $L_{[\mathrm{C\,II}]}$/SFR ratio at high metallicities can make a significant impact on the observed total $L_{[\mathrm{C\,II}]}$ when the atomic and molecular phases dominate $L_{[\mathrm{C\,II}]}$. This probably results from higher radiation fields in metal-rich galaxies, which reduce the sizes of the neutral clouds, as proposed by Narayanan & Krumholz (2017). We find that galaxies with high metallicity have average neutral cloud sizes $\sim 30\%$ smaller than metal-poor galaxies. We discuss this further in Sect 2.4.

2.3.4 Spatial distributions

Finally, we check the spatial distribution of the ISM fractional contributions inside the galaxies. We calculate the distance between each phase to the centre of the potential in each galaxy. We scale the distance using the half-mass radius (R_{50}) for all galaxies in Ref-L0100N1504 from the EAGLE Database. We present the fractional contribution of the ISM phases against the distance from the centre in units of R_{50} in Fig. 2.9. These radial distributions of the fractional contributions to $L_{[\mathrm{C\,II}]}$ show that atomic gas dominates the luminosity contribution at the centres of the galaxies. The molecular phase never dominates, but the contribution peaks at $\mathrm{r}/R_{50} \approx 0.5$, while in the outskirts the DIG dominates.

The atomic and molecular phases are confined to the galaxy centres compared with the DIG. Atomic gas always dominates below R_{50}, while the DIG median is nearly always lower than the molecular contribution within $\sim R_{50}$. The DIG could be significant when $L_{[\mathrm{C\,II}]}$ is measured in observations of unresolved galaxies. As we show in Fig. 2.5, the contribution of each phase also depends on the SFR of the galaxy. At high SFR, the molecular fraction increases while the atomic contribution decreases. In the same way, the DIG dominates at low SFR, which will lead to a different radial distribution in less actively star-forming galaxies. However, we expect a similar trend where the fractional contributions of atomic and molecular gas peak at the centres of the galaxies and the DIG dominates in the outskirts.

2.4 Discussion

2.4.1 The role of the DIG in the [C II] emission of galaxies

One of the main conclusions we can draw from the ISM model implemented in this work is the role that the DIG plays in the production of $L_{[\mathrm{C\,II}]}$. We show that the fractional contribution of each major ISM phase depends on the total SFR of the galaxy and the simulation used in Fig. 2.5. Intermediate-resolution simulations seem to overestimate the DIG contribution, while Recal-L0025N0752 is consistent with previous studies in the local Universe (e.g. Pineda et al. 2014; Narayanan & Krumholz 2017; Croxall et al. 2017; Cormier et al. 2019; Lupi & Bovino 2020), where most of the [C II] emission arises from neutral phases (atomic and molecular). Figure 2.5 shows that when Recal-L0025N0752 is used, the neutral phases dominate the $L_{[\mathrm{C\,II}]}$ fractional contribution, from around 60% at SFR $\approx 10^{-1}\,\mathrm{M_\odot\,yr^{-1}}$ to 80% at SFR $\approx 10\,\mathrm{M_\odot\,yr^{-1}}$. Similar results are obtained when we look only at the inner parts of

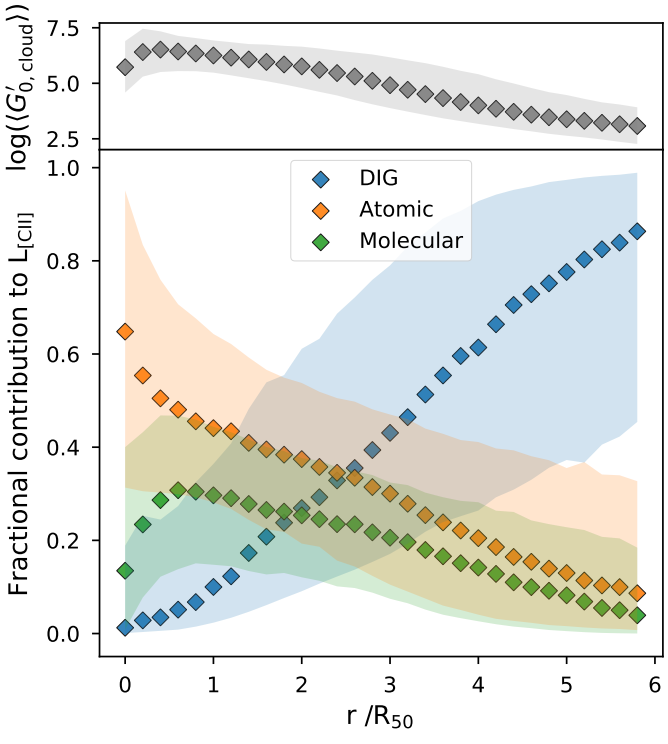

Figure 2.9 – *Median radial distribution (diamonds) of the radiation field (grey) in the neutral clouds (upper panel) and the fractional ISM phase contributions (lower panel) for galaxies in* Ref-L0100N1504. *There is a peak of neutral atomic gas contribution in the centre, the DIG contribution dominates beyond three times the half-mass radius (R_{50}) of the galaxies, and the dense molecular gas contribution peaks around $0.5R_{50}$ and then decreases slowly. The peak of the radiation field is similar to the one of the dense molecular gas. Shaded colour shows the coverage between the 25th and 75th percentiles.*

galaxies, inside R_{50} (Fig. 2.9). In galaxy centres, the total atomic contribution is $\sim 70\%$ and the molecular contribution is $\sim 20\%$, while at R_{50} the atomic contribution is $\sim 45\%$ and the molecular contribution is $\sim 30\%$.

The DIG is the dominant phase at metallicities above solar ($\sim 70\%$ of $L_{[\mathrm{C\,II}]}$), at $R_{50} \gtrsim 3$, and in the low-SFR regime ($< 10^{-2}\,\mathrm{M_\odot\,yr^{-1}}$) in all three simulations. The $L_{[\mathrm{C\,II}],\mathrm{DIG}}$/SFR ratio is more or less constant as a function of metallicity for the DIG phase (Fig. 2.8), in contrast to the atomic and molecular contributions, where the ratio is reduced drastically above solar metallicity. Below $1/3\ Z_\odot$ the atomic gas is the dominant component in all three simulations (up to 90% at $Z/Z_\odot = 0.03$), while the contribution of the dense molecular gas peaks near to solar metallicity. Thus the DIG is the dominant contributor to $L_{[\mathrm{C\,II}]}$ in high-metallicity galaxies. These results support the trend presented by other works (Croxall et al. 2017; Cormier et al. 2019), where the fractional contribution of the DIG increases at high metallicities.

The radial position where the fractional contribution of the DIG is high corresponds

to the outskirts of galaxies (Fig. 2.9). This agrees visually with the expected DIG location from Erroz-Ferrer et al. (2019, see their Appendix A), who estimated the location of the DIG regions, following a threshold value for the Hα emission (diffuse Hα emission), mainly in the outskirts of local star-forming galaxies. Although they do not compare Hα with [C II] emission, it is interesting to note the resemblance with another SFR tracer. This suggests that it could be possible to measure the DIG contribution in the outskirts of galaxies by measuring the [C II] flux, just like Hα emission. Croxall et al. (2017) and Sutter et al. (2019) found fractional contributions of DIG below 40% in resolved regions of galaxies, which coincide with our results inside the half-mass radii of galaxies (2–6 kpc). However, the results of Croxall et al. (2017) and Sutter et al. (2019) come mostly from regions within $0.25R_{25}$*, limiting the interpretation of these studies on the spatial distribution of the DIG phase.

In an analytical study, Popping et al. (2019) assume a constant density for diffuse gas. When Popping et al. (2019) reduce the density of the diffuse gas, the total [C II] luminosity is more affected (up to 0.5 dex with respect to their fiducial model) at $\log(\mathrm{SFR})\,[\mathrm{M_\odot\,yr^{-1}}] < 0$ than at $\log(\mathrm{SFR})\,[\mathrm{M_\odot\,yr^{-1}}] > 0$ (their Fig. 15). This result is consistent with the trends found in our work. When the DIG dominates at low SFRs, even small changes in the DIG composition can affect the total [C II] luminosity. However, the assumption of a constant density for diffuse gas and other differences in the ISM models, limits a direct comparison of the DIG dependency on $L_{[\mathrm{C\,II}]}$.

Our results show that, in metal-rich galaxies and where the SFR is low, the DIG could play a dominant role in producing the $L_{[\mathrm{C\,II}]}$. Ignoring these contributions at different SFRs can introduce a bias in the line emission estimations (Croxall et al. 2017). Detailed resolved observations of local galaxies and their outskirts will be essential to improve the calibration of the DIG fraction. At the same time, observations of galaxies with low SFR will also be important.

2.4.2 Variations in the $L_{[\mathrm{C\,II}]}$–SFR relation

As mentioned in Sects. 2.1 and 2.3.2, one problem of using $L_{[\mathrm{C\,II}]}$ as a SFR indicator is that the $L_{[\mathrm{C\,II}]}$–SFR relation is complex. A single power law cannot describe the relation accurately and variations are present due to changes in contribution from different ISM phases (Sects. 2.3.1 and 2.4.1) and the [C II] deficit, where the slope of $L_{[\mathrm{C\,II}]}$ with respect to FIR luminosity is less steep at high infrared luminosities.

A natural mechanism for the [C II] deficit was presented by Narayanan & Krumholz (2017), where the increase in the molecular fraction of the gas reduces the efficiency of [C II] emission due to the shrinking size of atomic gas in galaxies with high SFR. In Fig. 2.10, we present the average radius of the atomic region (R_{atomic}) in the simulated galaxies with respect to the total SFR, colour-coded by the average strength of the radiation field incident on the neutral cloud ($G'_{0,\mathrm{cloud}}$). The plot shows a gradual decrease in the effective atomic radius with SFR, which can be related to $G'_{0,\mathrm{cloud}}$. R_{atomic} shrinks because the sizes of the neutral clouds are reduced with increasing SFR, while $R_{\mathrm{H_2}}$ increases. This explains the trends observed in Fig. 2.5, where the molecular fractional contribution rises after the peak contribution of the atomic gas.

* 1/8 of the D_{25} standard diameter, the B-band isophotal radius at 25 mag arcsec^{-2} (de Vaucouleurs et al. 1991; Paturel et al. 1991; Prugniel & Heraudeau 1998).

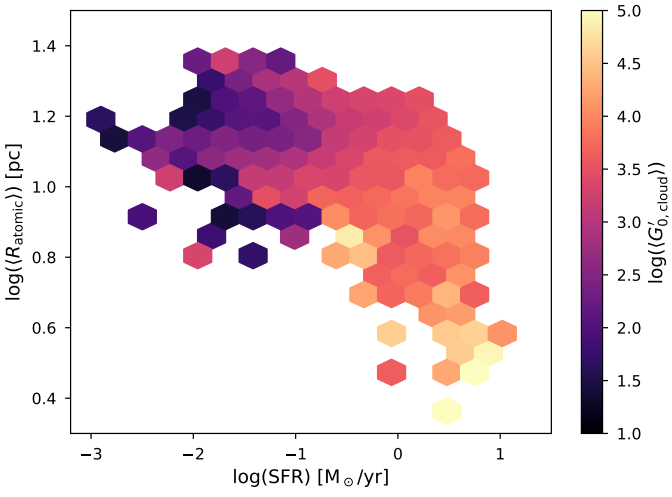

Figure 2.10 – *Average radius of the atomic region ($R_{\rm atomic}$) in all simulations (assuming they represent the same population) with respect to SFR, colour-coded by G'_0 (Habing units). We binned the $R_{\rm atomic}$ values to bin sizes of ~ 0.05 dex and the SFR to sizes of ~ 0.2 dex. This plot supports the idea proposed by Narayanan & Krumholz (2017) that the sizes of the atomic regions (emitting [C II]) shrink with increasing SFR, which can explain the [C II] deficit. At a given SFR, the radius decreases with increasing G'_0.*

This result is not surprising as the SFR in the EAGLE simulations is determined by the pressure in the galaxy, and at the same time, pressure constrains the size of the neutral cloud, so these results reflect those of Narayanan & Krumholz (2017). Unfortunately, the evolution of $R_{\rm atomic}$ cannot be tested at higher SFR for the local Universe in EAGLE (Katsianis et al. 2017), but $R_{\rm atomic}$ is expected to decrease for systems with higher SFR.

Another claim related to the variations in the $L_{\rm [C\,II]}$–SFR relation is that $L_{\rm [C\,II]}$ may not be a robust SFR indicator when intense radiation fields are present, such as in starburst galaxies (Herrera-Camus et al. 2018b; Ferrara et al. 2019). We test this hypothesis, following Herrera-Camus et al. (2018b), by calculating the specific star formation rate (sSFR = SFR/M$_\star$) for the galaxies in the simulations and normalising with the value derived for the main-sequence (MS) sSFR from Speagle et al. (2014). Figure 2.11 shows the $L_{\rm [C\,II]}$/SFR ratio with respect to ΔMS. This result is similar to that presented by Herrera-Camus et al. (2018b, their Fig. 6) but lacking the starburst outliers ($\Delta{\rm MS} \gtrsim 20$), which are not reproduced by EAGLE. The $R_{\rm atomic}$ reduction and the decrease in $L_{\rm [C\,II]}$/SFR ratio at higher ΔMS show that the strength of the radiation field can be a major factor in the observed variations in the $L_{\rm [C\,II]}$–SFR relation.

Díaz-Santos et al. (2017) show that less prominent [C II] deficits are related to higher [C II] fractional contributions coming from the neutral atomic phase. In Sutter et al. (2019) the main contributor to $L_{\rm [C\,II]}$ is found to be the neutral phase, which shows a less prominent [C II] deficit compared to the ionised phase. These results agree with those presented in this work. When the fractional contribution to $L_{\rm [C\,II]}$ from

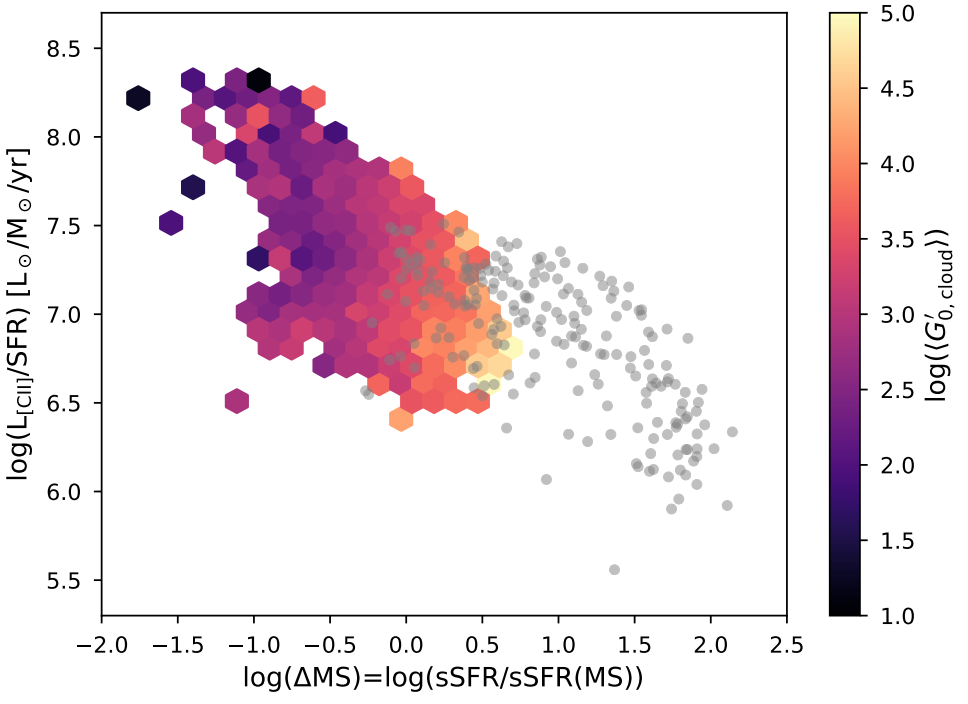

Figure 2.11 – *$L_{[C\,II]}$/SFR ratio as a function of ΔMS for the three simulations (assuming they represent the same population), colour-coded by G'_0 (Habing units). We binned the $L_{[C\,II]}$/SFR ΔMS values to sizes of ∼ 0.1 dex. When we compare our results with Herrera-Camus et al. (2018b, their Fig. 6) (grey dots), it is clear that* EAGLE *simulations do not recover starburst galaxies with* $\log \Delta MS > 0.5$*, but there is a clear indication that deviations from the MS affects $L_{[C\,II]}$/SFR.*

the atomic phase is small, the deficit is more prominent. The [C II] atomic phase contribution decreases at the same time that the total [C II] luminosity decreases. Thus the neutral phase is responsible for the deficit, as we have shown in this work.

However, when Sutter et al. (2019) consider only the ionised component, the [C II] deficit is more prominent compared to the neutral phases. This result is different from what we found in this work, where the [C II] luminosity from the DIG component does not decrease at the higher luminosities ($\log \mathrm{SFR} \approx 0.8 \sim \log \mathrm{L_{TIR}} \approx 10.75$) tested in this work. The methods used to calculate the ionised component may explain the differences between this work and Sutter et al. (2019). The Sutter et al. (2019) calculations come from directly scaling the [N II] 205 μm line deficit, while we calibrate the ionised component taking the two nitrogen emission lines and the SFR into account. The role that metallicity plays in the variation of the $L_{[C\,II]}$–SFR relation may also explain the differences. The sample of galaxies used by Sutter et al. (2019) has, in general, metallicities below solar. Our results (Fig. 2.8) show that in this range the atomic neutral phase dominates and the DIG contributes less at lower metallicities. In other words, this difference could be due to selection bias.

Smith et al. (2017) found a correlation between increased metallicity and deeper $L_{[C\,II]}$

deficits in high-luminosity infrared galaxies. In our case, at higher metallicities, the atomic and molecular gas contributions to the $L_{[\mathrm{C\,II}]}$ seem to be more affected than the DIG contribution (Fig. 2.8), showing more prominent $L_{[\mathrm{C\,II}]}$ deficits. At values below solar metallicity, the structure of the ISM changes, especially the transition from WNM-to-CNM (Bialy & Sternberg 2019). In this work, at $Z_{\odot} < 1$ the DIG seems to be deficient in producing [C II] compared to the atomic gas, but the DIG total contribution is always stable. As discussed in the previous paragraph, galaxies in the work by Sutter et al. (2019) and Díaz-Santos et al. (2017) are biased towards metallicities below solar, where atomic phases dominate $L_{[\mathrm{C\,II}]}$.

The contribution of a given phase to the variations in $L_{[\mathrm{C\,II}]}$–SFR relation depends on the star-formation regulation, due to AGN or star-formation feedback, and metallicity, and these are the keys to understanding the variations observed in line observations. However, we keep in mind that the observed variations in the $L_{[\mathrm{C\,II}]}$–SFR relation can come from selection effects and other biases, such as galaxy brightness, where starburst systems are more easily observed due to their luminosity (Katz et al. 2019). Nonetheless, trends in theoretical models are also dependent on the assumed parameters to estimate line emission (e.g. Olsen et al. 2015, 2017; Vallini et al. 2013, 2015, 2017; Lagache et al. 2018; Cormier et al. 2019; Katz et al. 2019; Popping et al. 2019). We discuss the caveats of our choices in the following section.

2.4.3 Caveats on our predictions

Model assumptions

Our findings may be limited by the assumptions we made in Sect. 2.2.2. We use the model presented by Olsen et al. (2015, 2017) as a basis. However, we have important differences: We calculate the fraction of neutral hydrogen following Rahmati et al. (2013), we change the density distribution in the neutral clouds to a Plummer profile following Popping et al. (2019) and we calibrate the DIG using [N II] emission lines. The main weakness of our model is its geometric simplicity. The model is based on patches of gas (SPH particles) with assumed constant physical parameters, such as the metallicity, which limits comparisons with real galaxies, where the different ISM phases are entangled with each other (Olsen et al. 2018b). Recent observational results (Dessauges-Zavadsky et al. 2019) have shown that the neutral cloud mass function (Eq. 2.15) is only valid in the local Universe but not at different redshifts, where neutral clouds could be more massive (higher surface densities). Fortunately, even though this simple approach does not take all galactic physics into account, the models reproduce observed relations in galaxies in the local Universe.

Olsen et al. (2018a) apply a geometrical correction to their predicted intensities because of the assumed spherical clouds in the model compared with the computed plane-parallel geometry coming from Cloudy. We do not implement this correction because the effect of this correction is much smaller than the general scatter presented in Fig. 2.6; therefore, the computational effort of applying the correction is not justified.

DIG calibration

We calibrate the DIG emission (Sect. 2.2.2) to account for the luminosity coming from this ISM phase in the [C II] line. This approach is biased towards the physical properties of the observations used in the calibration, for example luminous galaxies. At the same time, the assumed distribution function for $R_{\rm DIG}$ (Eq. 2.32) is not entirely physically motivated, as we do not know the actual distribution of this ISM phase in different types of galaxies. As we show in Fig. 2.5, this calibration can affect the DIG contribution of the [C II] line. The spatial distribution presented in Fig. 2.9 could also be affected by the uncertainty of $R_{\rm DIG}$. With smaller (bigger) $R_{\rm DIG}$ the density of the ionised clouds will increase (decrease), the gas becomes less (more) diffuse, thus the total luminosity coming from the DIG will increase (decrease) as well. However, the assumptions used in this work can give us insights into the general behaviour of the DIG (see the discussion in Sect. 2.4.4). Different physical assumptions for the DIG could be applied to the simulations, but most of them only assume the distribution of the ionised gas in disky galaxies (e.g. Haffner et al. 2009; Vandenbroucke & Wood 2019, and references therein). Models with reliably calibrated DIG properties would be very important to understand diagnostic line emissions in galaxy evolution (Kewley et al. 2019), and more observations focusing on this ISM phase are required, such as the GBT Diffuse Ionised Gas Survey*.

In this work, we present luminosity predictions coming from [C II] emission line with calibrated emission from [N II] emission lines, which is ideal in terms of modelling (Olsen et al. 2018b), as other lines can help to constrain the physical properties in the models to improve emission line predictions (Popping et al. 2019). We plan to study the local Universe and at high redshift with more FIR emission lines in future works.

2.4.4 Choice of simulation

The findings of this work may be somewhat limited by the selection of a given EAGLE simulation to estimate the line emission. For example, large cosmological volumes ($L > 50\,$cMpc) provide a more representative range of environments in galaxies compared to small boxes (Schaye et al. 2015; Furlong et al. 2015). In addition, the efficiency of feedback must be calibrated (e.g. with the GSMF) in large hydrodynamical simulations where the ISM scales are unresolved, and therefore calibration is required to make valid predictions for observables (Schaye et al. 2015; Naab & Ostriker 2017).

We have presented results from three simulations (Ref-L0025N0376, Recal-L0025N0752, and Ref-L0100N1504) to test the effects of the box size and resolution on the predictions. We presented some of these comparisons in Figs. 2.5–2.8. In terms of the simulated box size (Ref-L0025N0376 vs Ref-L0100N1504), the predictions are very similar, and only the total number, typical masses and SFR of galaxies change. As remarked above, EAGLE does not contain many galaxies with SFR $> 1\,{\rm M_\odot\,yr^{-1}}$ at $z = 0$ due to the specific implementations (Katsianis et al. 2017; Wang et al. 2019b); therefore, with EAGLE simulations we can predict line emissions for normal star-forming galaxies at $z = 0$, but not starburst-like systems.

The trends in Figs. 2.6 and 2.8 show that $L_{\rm [C\,II]}$ is always lower for Recal-L0025N0752

* https://greenbankobservatory.org/science/gbt-surveys/gdigs/

galaxies. These lower values are due to the GSMF re-calibration. De Rossi et al. (2017) noted that the increase in the resolution for these simulations (Ref-L0100N1504 to Recal-L0025N0752) can affect some features but the fundamental metallicity scaling relations are not altered. Therefore, boxes re-calibrated to the GSMF are a good starting point to predicting emission lines as metallicity scaling relations holds. From our comparison with observations (Sect. 2.3.1), we expect that the observed galaxies in the local Universe behave as in Recal-L0025N0752, where the atomic gas dominates at SFR values higher than $0.01\,\mathrm{M}_{\odot}\,\mathrm{yr}^{-1}$.

Ref-L0100N1504 provides (statistical) clues on how the fractional contribution of the ISM phases changes. Predicted luminosities from Ref-L0100N1504 and Ref-L0025N0376 are overestimated compared with observed galaxies, which are well reproduced by Recal-L0025N0752. This conclusion can also be seen from the right panel of Fig. 2.8, where a reduction of the contribution from atomic and molecular phases in $L_{[\mathrm{C\,II}]}$ increases the dominance of the DIG phase. To summarise, Recal-L0025N0752 gives us the best predictions in terms of reproducing observed local galaxies and Ref-L0100N1504 gives us larger sample sizes for statistical studies.

As Recal-L0025N0752 does not reproduce the large number of galaxies needed for some statistical studies, we suggest that the best way to compare our simulations to observations is a mix between Ref-L0100N1504 and Recal-L0025N0752 to observe the global behaviour of the emission lines, as we do in Figs. 2.10 and 2.11. This gives us a better idea of the contribution of the different ISM phases to the line emission.

2.5 Summary and conclusions

In this work, we have post-processed SPH EAGLE simulations using a multi-phase ISM model and then predicted the luminosity of [C II] emission at 158 μm with Cloudy. We set out to determine the fractional contributions of the different ISM phases to the [C II] line and the effect of these phases on the $L_{[\mathrm{C\,II}]}$–SFR relation. We use three sets of simulations (Ref-L0025N0376, Recal-L0025N0752, and Ref-L0100N1504) with two Cloudy cooling tables for shielded and optically thin regimes (Ploeckinger & Schaye 2020) to characterise ISM composition in terms of dense molecular gas, neutral atomic gas, and diffuse ionised gas (DIG). We validate our model by comparing with observations of local galaxies. Our main conclusions are the following:

1. We find a dependence of the fractional contribution of the different ISM phases to $L_{[\mathrm{C\,II}]}$ on the total SFR and metallicity. Our model agrees with previous works where the [C II] emission comes mainly from neutral ISM regions in observed galaxies. However, this could be a selection bias in the observations towards galaxies with metallicities below solar.

2. In systems where the SFR is low, the DIG plays a dominant role in producing $L_{[\mathrm{C\,II}]}$. Additional resolved observations of local galaxies with low SFRs and their outskirts will be essential to improve the calibration of the DIG fraction in other emission lines (van der Tak et al. 2018).

3. The model for the ISM we have implemented in this work reproduces the observed $L_{[\mathrm{C\,II}]}$–SFR relation in local galaxies, although this result depends on calibrating the DIG with [N II] emission lines.

4. Star-formation regulation and metallicity dependence of the different ISM phases could be responsible for the variations observed in $L_{[\mathrm{C\,II}]}$ at high infrared luminosities, based on the dependence on ΔMS and average R_{cloud}. Further investigations are needed to verify if this result holds at higher redshifts.

5. The use of large boxes (box-size $\gtrsim$ 100 cMpc) and high-resolution simulations (mass resolution $\lesssim 10^5\,\mathrm{M}_\odot$) is key to correctly predict emission lines in different types of galaxies.

In the future, we will take black hole particles (especially the SMBH) into account, which will alter the radiation field in the centres of the galaxies. We will use a random sample of galaxies from the simulations to estimate other fine structure lines. We plan to compare predictions in high-redshift systems with observational results, such as those presented by Le Fèvre et al. (2020) and Neeleman et al. (2019) with a large number of observed galaxies, using for example ΔMS.

Acknowledgements: The authors thank Rodrigo Herrera-Camus for providing the data from the SHINING survey. We acknowledge the anonymous referee for a careful reading of the manuscript and very helpful questions and comments. We acknowledge the Virgo Consortium for making their simulation data available. The EAGLE simulations were performed using the DiRAC-2 facility at Durham, managed by the ICC, and the PRACE facility Curie based in France at TGCC, CEA, Bruyères-le-Châtel. This research made use of Astropy,* a community-developed core Python package for Astronomy (Astropy Collaboration et al. 2013, 2018). We would like to thank the Center for Information Technology of the University of Groningen for their support and for providing access to the Peregrine high performance computing cluster.

* http://www.astropy.org

Appendix

Comments on archival Data

To complement and verify our model in the local Universe, we use archival data from different samples of observed galaxies where the [C II] luminosity is available. In cases where it is possible, we recalculate the luminosities and SFRs with the same cosmology used in this work (Planck Collaboration et al. 2014a). We use the literature samples from (a) the ISO compendium (Brauher et al. 2008) of main-sequence galaxies; (b) xCOLD GASS (Accurso et al. 2017), composed of intermediate-stellar-mass galaxies; (c) GOALS (Díaz-Santos et al. 2013, 2017), composed of LIRGS observed with *Spitzer* and *Herschel*; (d) SHINING (Herrera-Camus et al. 2018b), composed of nearby star-forming galaxies, AGNs, and LIRGs observed with *Herschel*; (e) VALES (Hughes et al. 2017), composed of dusty main-sequence star-forming galaxies observed with *ALMA*; and (f) the Herschel Dwarf Galaxy Survey (HDGS, Cormier et al. 2015, 2019). For the ISO compendium: We ignore close galaxies (below 1 Mpc), we use the median for the line flux in galaxies with more than one measure in the [C II] line, and the infrared luminosity was calculated with the 60 and 100 μm from the IRAS flux as described by Brauher et al. (2008). In most of the literature samples we have calculated the SFR from the IR luminosity as described by Kennicutt & Evans (2012). We use the SFRs from Villanueva et al. (2017) and Saintonge et al. (2016) for VALES and xCOLD GASS samples, respectively.

We compare these literature samples with our SFR and $L_{[\mathrm{C\,II}]}$ estimates in Fig. 2.12. We notice that galaxies in RECAL-L0025N0752 follow similar trend as the dwarf galaxies from Cormier et al. (2019) while REF-L0100N1504 is similar to most of the intermediate stellar mass galaxies from Accurso et al. (2017). Both simulations reproduce main sequence galaxies (Brauher et al. 2008) as well as some galaxies from other samples (e.g. Hughes et al. 2017; Herrera-Camus et al. 2018b). LIRGS galaxies (Díaz-Santos et al. 2017; Herrera-Camus et al. 2018b) are representative of the [C II] deficit, but unfortunately, are not recovered by EAGLE (see Sect. 2.3.2 and Fig. 2.6 to compare with the deficit).

Additionally, we use [N II] observational data from Fernández-Ontiveros et al. (2016) to calibrate the DIG. We describe the procedure in Sect. 2.2.2 and in Fig. 2.1 we show the simulated and observed $L_{[\mathrm{N\,II}]}$–SFR relation for the lines at 122 and 205 μm.

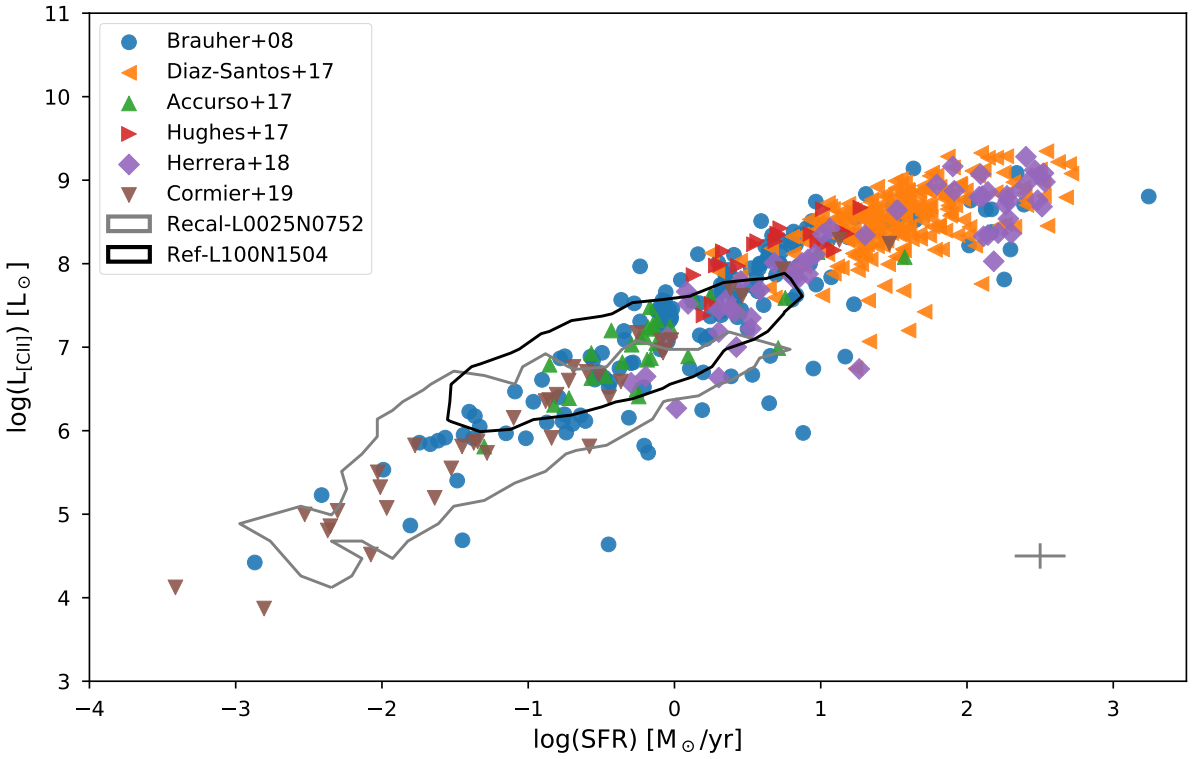

Figure 2.12 – $L_{[C\,II]}$–*SFR relation for the observational sample of local galaxies and two simulations used in this work (*Recal-L0025N0752, Ref-L0100N1504*) presented as contour maps (grey and black). The contours shows where 95% of the galaxies of the respective simulations fall in the relation. We present the mean error from the observational samples in the bottom-right corner of the plot.

Never give up, never surrender
Jason Nesmith (Tim Allen) - Galaxy Quest

Tea please (Don't say you don't like coffee)
Andrés when someone offers him coffee

3

Diagnosing the interstellar medium of galaxies with far-infrared emission lines: [C II], [O I], [O III], [N II] and [N III] up to $z = 6$

A. F. Ramos Padilla, L. Wang, F. F. S. van der Tak and S. C. Trager

Submitted to Astronomy & Astrophysics

Highlights

- Far-infrared (FIR) fine-structure emission lines can be used as a tool to understand the gas conditions and trace the different phases of the ISM.
- We model the most important far-infrared (FIR) emission lines throughout cosmic time back to $z = 6$.
- In this version of our model, we assume four phases of the ISM: dense molecular gas, neutral atomic gas, diffuse ionised gas (DIG) and HII regions.
- Predictions from our model replicate observed galaxies in the SFR–FIR line luminosity relationship over the range $z = 0$–6 for all FIR lines.
- Line ratios like [C II]/[O III] and [N II]/[O I] are useful to trace parameters such as ISRF, metallicity and specific star-formation rate in diagnostic diagrams.

Abstract

Context: Gas cooling processes in the interstellar medium (ISM) are key to understanding how star-formation processes occur in galaxies. Far-infrared (FIR) fine-structure emission lines can be used as a tool to understand the gas conditions and trace the different phases of the ISM.
Aims: We model the most important far-infrared (FIR) emission lines throughout cosmic time back to $z = 6$ with cosmological hydrodynamical simulations. We study how different physical parameters, such as the interstellar radiation field (ISRF) and metallicity, impact the ISM phases traced by FIR line luminosities and connect those with the star-formation rate (SFR).
Methods: We implement a physically motivated multi-phase model of the ISM by post-processing EAGLE cosmological simulation with Cloudy look-up tables. In this model, we assume four phases of the ISM: dense molecular gas, neutral atomic gas, diffuse ionised gas (DIG) and HII regions.
Results: Our model shows good agreement with the observed luminosity–SFR relation up to $z = 6$ in the FIR emission lines analysed and we also provide linear fits. Our predictions also agree with observations in terms of diagnostic diagrams involving various line ratios.
Conclusions: We find that [C II] is the best SFR tracer of the FIR lines even though it traces multiple ISM phases, while [O III] and [N II] can be used to understand the DIG-HII balance in the ionised phase. In addition, line ratios like [C II]/[O III] and [N II]/[O I] are useful to trace parameters such as ISRF, metallicity and specific star-formation rate. These results help to interpret observations of FIR line emission from the local Universe to high-z galaxies.

Keywords: Galaxies: ISM, star formation, high-redshift – ISM: lines and bands, structure – Infrared: ISM – methods: numerical

3.1 Introduction

After the Universe was largely ionised (the period known as "cosmic dawn" at $z \lesssim 6$–8), the composition of the gas and its cooling gradually changed, affecting the star-formation processes in galaxies (Dayal & Ferrara 2018). Since then, star formation and black hole accretion processes co-evolved with cosmic time and shaped the evolution of galaxies (Madau & Dickinson 2014). The study of the interstellar medium (ISM) in the local Universe allows us to comprehend the current star-formation processes, but the ISM gas cooling budget may not be the same at earlier cosmic epochs (Carilli & Walter 2013). Recent observational data have opened new paths for describing and understanding the ISM gas processes of local galaxies (e.g. Malhotra et al. 2001; De Looze et al. 2014; Cormier et al. 2015). However, the complete picture of how the ISM evolves over cosmic time and how its conditions are connected with star formation in galaxies is not well understood.

Far-infrared (FIR) fine-structure emission lines (Table 3.1) are dominant in the gas cooling of the ISM and can help us to understand the star-formation processes, from a theoretical (e.g. Tielens & Hollenbach 1985; Goldsmith et al. 2012; Wolfire et al. 2022) and observational perspective (e.g. Ferkinhoff et al. 2010; Carilli & Walter 2013; Cormier et al. 2015; Díaz-Santos et al. 2017; Herrera-Camus et al. 2018b). These lines are less affected by dust extinction than optical lines and, at high redshift (hereafter

Table 3.1 – *Important FIR fine-structure emission lines. Data from Kramida et al. (2020) and Spinoglio et al. (2017).*

Line Species	λ [μm]	Transition	IP [eV]	$n_{\rm crit}$ [cm^{-3}]
[O III]	51.81	$^3P_2$ – $^3P_1$	35.12	3.6×10^3
[N III]	57.34	$^2P_{3/2}$ – $^2P_{1/2}$	29.60	3.0×10^3
[O I]	63.18	$^3P_2$ – $^3P_1$	···	4.7×10^5
[O III]	88.36	$^3P_1$ - $^3P_0$	35.12	5.1×10^2
[N II]	121.80	$^3P_1$ – $^3P_0$	14.53	3.1×10^2
[O I]	145.53	$^3P_1$ – $^3P_0$	···	9.5×10^4
[C II]	157.68	$^2P_{3/2}$ – $^2P_{1/2}$	11.26	2.8×10^3
[N II]	205.30	$^3P_2$ – $^3P_1$	14.53	4.8×10^1

high-z), are shifted to the (sub-)mm wavelength range accessible to ground-based telescopes (Hodge & da Cunha 2020; Förster Schreiber & Wuyts 2020). The most important ISM emission cooling line is [C II] at 158 μm. This line traces different phases of the ISM: photo-dissociation regions (PDRs), HII regions, diffuse ionised gas (DIG; also known as the warm ionised medium (WIM), e.g. Haffner et al. 2009; Kewley et al. 2019), molecular clouds, and the cold and warm neutral media (CNM and WNM, respectively) (e.g. Cormier et al. 2012; Croxall et al. 2017; Abdullah et al. 2017; Abdullah & Tielens 2020). [C II] is thus a very important cooling line in the range of 20–8000 K due to its low ionisation potential (11.3 eV compared to 13.6 eV for hydrogen: Gong et al. 2012; Goldsmith et al. 2012). Furthermore, it is easily observable, as its luminosity is around 1% of the FIR luminosity of galaxies (e.g. Stacey et al. 1991; Brauher et al. 2008). Other important FIR lines include: i) atomic oxygen ([O I]) at 63 and 145 μm, which traces the denser and warmer neutral ISM environments important for star formation (Malhotra et al. 2001; Goldsmith 2019); ii) ionised nitrogen ([N II]) at 122 and 205 μm, which traces the ionised medium from DIG and HII regions (in the local Universe, Cormier et al. 2012; Goldsmith et al. 2015; Zhao et al. 2016a; Croxall et al. 2017; Langer et al. 2021, and at high-z, e.g. Pavesi et al. 2016); iii) [O III] at 52 and 88 μm, which traces HII regions around young stars (at high-z Ferkinhoff et al. 2010; Inoue et al. 2014); and iv) [N III] at 57 μm, which also traces HII regions (Nagao et al. 2011). Using these lines to understand the ISM evolution of galaxies requires a self-consistent model for all of these FIR lines over the $z = 0$–6 range.

For many decades, the lack of suitable instruments has hampered observations of these FIR lines in high-z galaxies (Inoue et al. 2014). Fortunately, observations taken with telescopes such as IRAM, CSO, *Herschel* and ALMA have provided the first high-z detections of lines like [C II], [N II], [O I] and [O III] (e.g. Maiolino et al. 2005; Ferkinhoff et al. 2010, 2011; Inoue et al. 2016; Uzgil et al. 2016). Moreover, recent emission line surveys like ALPINE (Le Fèvre et al. 2020) and REBELS (Bouwens et al. 2022) are gathering data for larger samples of high-z galaxies, which are ideal to diagnose the ISM of galaxies over cosmic time.

With these new observations, different tools can be used to describe the different physical conditions in the ISM of high-z galaxies. The most common and accessible

tool is the use of emission line ratios that reflect the physical conditions of the ISM. For example, the [C II]/[N II] ratio has been used to describe and constrain whether HII regions and/or PDRs contribute to the ISM phases of galaxies (Decarli et al. 2014), and can be used to estimate the amount of ionised gas in [C II] (Croxall et al. 2017). Another useful line ratio is [O III]/[C II], used to understand ionised and neutral gas in high-z galaxies (e.g. Harikane et al. 2020; Carniani et al. 2020). This ratio has the advantage that [O III] can be brighter than [C II] at redshifts around the reionisation epoch ($z \gtrsim 7$, Inoue et al. 2014, 2016) and can be observed efficiently with ALMA (Bouwens et al. 2022). Finally, the [O III]/[N III] ratio is used to estimate the gas metallicities (Nagao et al. 2011; Herrera-Camus et al. 2018a), although other line ratios can also be used for this purpose (Fernández-Ontiveros et al. 2021).

A more sophisticated tool to describe the physical processes of the ISM is the use of line luminosity predictions from simulations or analytical models. Simple models involve ratios between FIR emission lines and/or the total luminosity in the IR to obtain physical conditions such as hydrogen density, FUV radiation flux or stellar temperatures (e.g. Malhotra et al. 2001; Ferkinhoff et al. 2011), while more complex models use magneto-radiation hydrodynamics simulations of the Universe with different radiative transfer codes to predict emission line luminosities (e.g. Olsen et al. 2021; Katz et al. 2022; Pallottini et al. 2022, and references therein). Some of these studies focus on specific emission lines such as [C II], [O I] or [O III] in analytic models (e.g. Goldsmith 2019; Ferrara et al. 2019; Yang & Lidz 2020) and simulations (e.g. Moriwaki et al. 2018; Leung et al. 2020; Ramos Padilla et al. 2021), while others study interesting line ratios like [O III]/[C II] (Arata et al. 2020; Vallini et al. 2021). A large effort has also been made to model various FIR emission lines in models at different cosmic times (e.g. Vallini et al. 2013, 2015; Olsen et al. 2015, 2017, 2021; Popping et al. 2019; Sun et al. 2019; Katz et al. 2019, 2022; Pallottini et al. 2019; Yang et al. 2021c, 2022b). However, these studies do not examine the emission of these FIR lines in a consistent way in terms of their redshift evolution, from the local Universe to the epoch of reionisation, due to computational constraints or focus on certain cosmic epochs (with the exception of Popping et al. (2019) which modelled the [C II] line). Therefore, we need better models to understand current observations consistently.

With this in mind, we aim to predict luminosities of the main FIR lines in a cosmological context through the use of cosmological hydrodynamical simulations to infer the physical conditions of galaxies across a wide range of redshifts. The goal of this paper is to test the impact of physical parameters on the FIR emission lines tracing different ISM phases in galaxies. We will use these predictions as diagnostic tools, which will be useful for both current and future observations. To do this, we model the emission of FIR lines by post-processing the hydrodynamical simulations of the Evolution and Assembly of GaLaxies and their Environments (EAGLE) project (Schaye et al. 2015; Crain et al. 2015) with a physically motivated model of the ISM presented in Ramos Padilla et al. (2021, hereafter Paper I). We use CLOUDY (Ferland et al. 2017) cooling tables (Ploeckinger & Schaye 2020) to predict the emission from different ISM phases in galactic environments. Throughout this paper, we assume the ΛCDM cosmology from the Planck Collaboration et al. (2014b) results ($\Omega = 0.307$, Ω_Λ=0.693, $H_0 = 67.7$ km s^{-1} Mpc^{-1} and $\sigma_8 = 0.8288$).

In this paper, we first briefly describe the simulation data and the ISM model that we

Table 3.2 – EAGLE *simulations used in this work. The box-size, number of particles and gas particle mass define the resolution of the simulation. The right columns show the number of galaxies used in this work for a given simulation at each redshift.*

Name in Schaye et al. (2015)	Box-size (cMpc)	Number of particles	SPH Gas mass ($M_\odot$)	Number of galaxies						
				$z=0$	$z=1$	$z=2$	$z=3$	$z=4$	$z=5$	$z=6$
RECAL-L0025N0752	25	752^3	2.26×10^5	415	426	339	252	154	75	37
REF-L0100N1504	100	$1\,504^3$	1.81×10^6	1 000	1 000	1 000	1 000	1 000	1 000	579

use to predict FIR emission line luminosities (Sect. 3.2). Next, we present the results of the FIR emission line predictions from the simulations and how they compare with the observations from the local Universe all the way out to $z\sim6$ (Sect. 3.3). In Sect. 3.4 we evaluate some FIR diagnostic diagrams used in various high-z studies. After that, we discuss the potential systematic uncertainties that can affect the predictions and the comparisons with observations (Sect. 3.5). Finally, we present our conclusions in Sect. 3.6.

3.2 Methodology

In this section, we first describe the sets of simulations that we use in this work (Sect. 3.2.1), then we briefly explain the initial model used to characterise the structure of the ISM (Sect. 3.2.2). Finally, we present in detail the addition of HII regions as a new ISM phase (Sect. 3.2.3) in our model.

3.2.1 The EAGLE simulations

EAGLE (Schaye et al. 2015; Crain et al. 2015) is a suite of cosmological hydrodynamical simulations which were run using a modified version of GADGET-3 (last described by Springel 2005), a smoothed-particle hydrodynamics (SPH) code. Briefly, EAGLE adopts an SPH pressure-entropy parameterisation following Hopkins (2013). The simulations include radiative cooling and photo-electric heating (Wiersma et al. 2009a), star formation (Schaye & Dalla Vecchia 2008), stellar evolution and mass loss (Wiersma et al. 2009b), black hole growth (Springel et al. 2005; Rosas-Guevara et al. 2015), and feedback from star formation and active galactic nuclei (AGN) (Dalla Vecchia & Schaye 2012). The simulations provide the properties for gas, dark matter, stellar and supermassive black hole SPH particles.

For this work, based on the results from Paper I, we used two simulations from the EAGLE suite: REF-L0100N1504 and RECAL-L0025N0752, as described in Table 3.2. The main differences between the two simulations are the box-size of the simulation (100 and 25 cMpc (comoving Mpc), respectively), the mass resolution ($\sim10^6\,M_\odot$ and $\sim10^5\,M_\odot$, respectively), and the calibration of the physical parameters of the subgrid routines to reproduce the galaxy stellar mass function (GSMF; Schaye et al. 2015). Both simulations are similar in terms of "weak convergence", which means numerical results converge in different simulations after re-calibrating the sub-grid parameters (Furlong et al. 2015; Schaye et al. 2015).

In this study, we retrieve simulated galaxies from the SPH data (The EAGLE team 2017) by using textscFoF (Friends-of-Friends) and SUBFIND algorithms (Springel

et al. 2001; Dolag et al. 2009) in the dark matter halos. With these algorithms, the sub-halos containing the particle with the lowest value of the gravitational potential are called "central" galaxies. We focus on these "central" galaxies to estimate line luminosities. We use galaxies with at least 300 star particles (i.e. stellar masses higher than $\sim 10^8\,M_\odot$ and $\sim 10^{8.5}\,M_\odot$ for RECAL-L0025N0752 and REF-L0100N1504, respectively) within 30 pkpc (proper kpc) from the centre of the potential. We selected our sample of galaxies from the EAGLE database (McAlpine et al. 2016) in redshifts (the closest snapshot) between $z = 0$ and $z = 6$ in steps of $\Delta z = 1$, using a single snapshot at each redshift (the snapshot closest in time; e.g. at $z = 6$, we use the $z = 5.97$ snapshot from EAGLE), where the total number of galaxies depends on the simulation used. The $z = 6$ cutoff was selected due to the availability of observational data and the number of galaxies recovered in EAGLE required to compare them statistically. For RECAL-L0025N0752, we select all retrieved galaxies, while for REF-L0100N1504, we randomly select up to 1 000 galaxies that fulfil the previous conditions per redshift. In the last columns of Table 3.2, we present the total number of galaxies used per redshift slice in each of the simulations.

We selected a total sample of 8 277 galaxies simulated with EAGLE at redshifts between $z = 0$ and $z = 6$. In each of those galaxies, we model the emission coming from the eight FIR emission lines (Table 3.1) that trace different phases of the ISM.

3.2.2 The multi-phase ISM model

To predict these emission lines, we used the ISM model presented in Paper I, where the luminosity estimations from the [C II] emission line at 158 μm in the local Universe ($z = 0$) showed good agreement with observations. However, this model must be improved if we want to properly account for other FIR lines such as [O III] and [N III], which probe denser ionised regimes. Therefore, we add HII regions as a new phase in our ISM model. In Fig. 3.1, we illustrate the path from the EAGLE simulations (Sect. 3.2.1) to the total luminosity of the lines in the current model. The main difference between Paper I and this work is the addition of the contribution to the line luminosities from HII regions (Sect. 3.2.3). In this subsection, we briefly explain the model presented in Paper I.

After selecting the sample of galaxies for which we want to calculate the line luminosity, we retrieve and post-process the gas and star particle data of the galaxies in the simulation. Physical properties such as total hydrogen number density were estimated for all gas particles using the information available in the SPH particle data. For example, we calculated the total hydrogen number density as $n(\mathrm{H}) = \frac{\rho X_H}{m_H}$, with m_H the hydrogen mass, ρ the density and X_H the SPH weighted hydrogen abundance. In addition, the fraction of neutral hydrogen was estimated for all gas particles according to Rahmati et al. (2013), following ionisation equilibrium as described in Sect. 2.2.1 in Paper I. We calculated the background interstellar radiation field (ISRF) from the star formation rate (SFR) surface density of the gas particles and the local ISRF from the star particles as described by Olsen et al. (2017). The local ISRF is estimated with `starburst99` models (Leitherer et al. 2014) for star particles with a distance below the smoothing length of a given gas particle (as described in Sect. 2.2.1 in Paper I). The sum of the background ISRF and the local ISRF defines the total ISRF

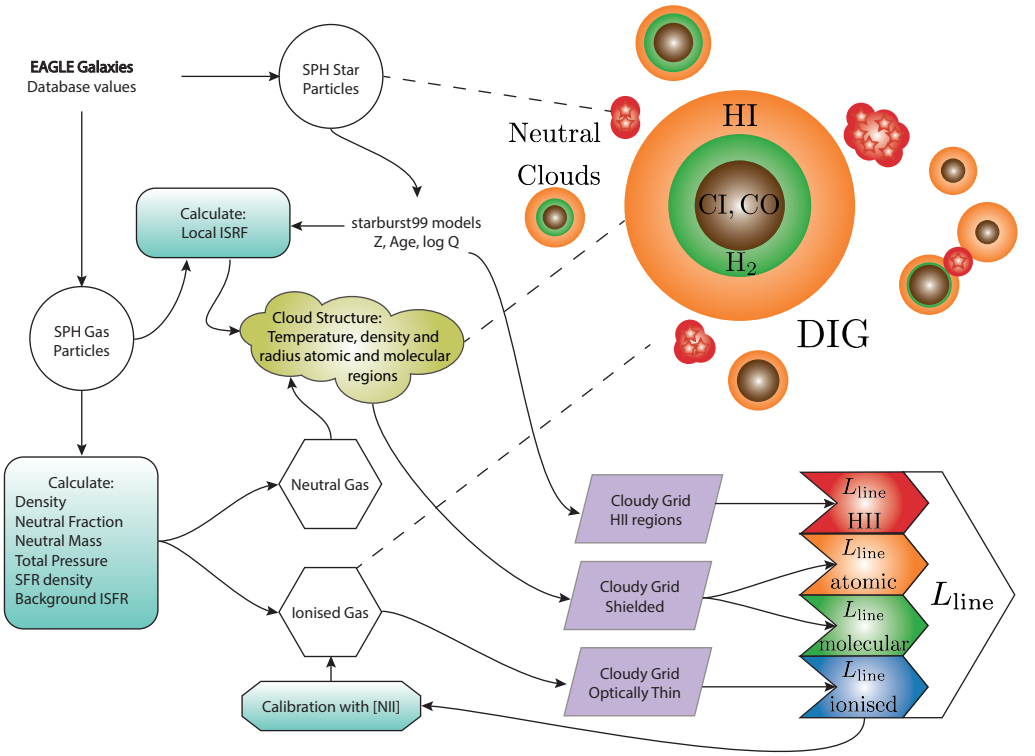

Figure 3.1 – *Flowchart of the sub-grid procedures applied to the SPH to simulate FIR line emission in post-processing. This flowchart is similar to the one presented in Paper I with the main difference being the added* HII *regions as a new ISM phase. The dashed lines connect the gas and star environments to the ingredients of the model.*

impinging on the neutral cloud. With the fraction of neutral hydrogen, we split the gas into neutral and ionised components. The neutral gas mass and the ISRF are used to estimate the sizes of the neutral clouds, while the ionised mass is used to describe the diffuse ionised gas (DIG).

The neutral clouds are defined as concentric spheres with densities following a Plummer profile. Each ISM phase in the neutral cloud is defined by a different radius as described in Sect. 2. in Paper I. The transition between atomic and molecular hydrogen defines the limit between the neutral atomic gas and the dense molecular gas, while the transition between ionised to neutral carbon defines the limit for the inner core region of the cloud that is completely shielded from FUV radiation. This definition assumes that the inner core can only be traced by neutral species such as CO, and the inner core is therefore ignored in the estimation of the total luminosity of the FIR emission lines*.

On the other hand, the volumetric structure of the DIG is assumed to be spherical with a radius drawn from a smoothed broken power-law distribution. The parameters of the power-law distribution are derived based on calibrating the [N II] line (at 122 and 205 μm) predictions of 492 galaxies from the Ref-L0100N1504 and 200 galaxies from Recal-L0025N0752 simulations to the observational dataset from Spinoglio et al. (2015), Fernández-Ontiveros et al. (2016) and Cormier et al. (2019) (containing 8, 53 and 2 galaxies, respectively, of a sample of 63 galaxies). This led us to DIG clouds with average sizes of ∼900 pc for Recal-L0025N0752 galaxies and ∼2 kpc for Ref-L0100N1504, which is around 3 times the maximum softening length of the original simulations.

These structures define the phases of the ISM that contribute to the total luminosity of the emission line: DIG, neutral atomic and dense molecular gas. The estimation of the luminosities for each phase is obtained by using Cloudy (Ferland et al. 2017, v17.01,) cooling tables of shielded and optically thin gas (Ploeckinger & Schaye 2020). The sum of those phases then gives us the total luminosity for a FIR emission line. For a complete description of the model, we refer the reader to Sect. 2.2 and 2.3 in Paper I.

3.2.3 HII regions as a new ISM phase

The luminosity estimations for FIR lines such as [O III] and [N III], which require high ionisation potentials (35.1 and 29.6 eV, respectively), can only be found in HII regions and other dense, ionised regimes. The addition of this phase as an ISM component to the model allows us to compare the predicted luminosities of these lines with observations. Therefore, we update the model described in Paper I to simulate the HII regions production of the most prominent FIR lines, including [C II], [O III], [O I], as shown in Table 3.1, in order to infer the ISM conditions in galaxies from the local Universe out to $z \approx 6$.

In Paper I, we use `starburst99` to calculate the ISRF from the stars close to gas particles (the distance between the particles is less than one smoothing length).

* [O I] emission can come also from the inner core. However, as our results show, the contribution of this inner core to the [O I] luminosity is usually very small.

Table 3.3 – *Sampling of the properties of* HII *regions in the* CLOUDY *grid used in this work. The resulting number of grid points is 3600 per redshift.*

Parameter	Unit	Min.	Max.	Interval
Metallicity	$[Z_\odot]$	0.075	2.991	0.324
$\log(Q)$	$[s^{-1}]$	44	52	1
log(Age)	[Gyr]	−3.0	0.9	0.1

Now, we also use the spectrum from `starburst99` to calculate the emission coming from HII regions. To generate this spectrum, we adopted the Geneva stellar models (Schaller et al. 1992) with standard mass loss for five metallicities ($Z = 0.001, 0.004, 0.008, 0.020,$ and 0.040). We split `starburst99` grid values in young ($\leq$ 100Myr) and old ages ($>$ 100Myr), to improve our estimations at younger ages. For young ages, we estimate the parameters every 1 Myr, while for old ages we calculate 100 steps on a logarithmic scale up to 10 Gyr. We assume a total stellar mass of $10^4\,M_\odot$, with a Kroupa initial mass function (IMF). The SPH star particles are divided to match the stellar mass from the `starburst99` models, assuming a random exponential distribution of the original ages of the SPH particles. By doing this we try to avoid the poor sampling of star formation that can affect luminosity estimates, especially in HII regions.

We use a a photoionisation model from CLOUDY to simulate the line emissitivites of HII regions based on `starburst99` spectra of their underlying stellar populations. We calculate the stellar atmospheres, or spectral energy distributions (SEDs), in the photoionisation models for a given age, metallicity and the ionising photon flux (Q). In this way, the stellar age and metallicity from the star SPH particles are used to obtain Q coming from a stellar cluster from the `starburst99` grids. Thus, we use these three physical parameters (age, metallicity and Q) to construct spherical clouds where the emissivities depend only on them. The range of values for these parameters is presented in Table 3.3, totalling 25 200 grid points for all redshifts in this work.

In terms of the structure, we assume a fixed density of $\sim 30\,\mathrm{cm}^{-3}$ to resemble classical HII regions similar to the Strömgren sphere, where densities are in the range of 10–$100\,\mathrm{cm}^{-3}$ (Draine 2011). Choosing a higher or lower density affects the HII luminosity of some lines as we describe in Sect. 3.5. In ionisation balance the Strömgren radius is

$$R_{\mathrm{Str}} = \left(\frac{3Q}{4\pi\, n(\mathrm{H})^2\, \alpha_B} \right)^{1/3}, \tag{3.1}$$

with Q the ionising photon flux in units of s^{-1}, $n(\mathrm{H})$ the total hydrogen number density in units of cm^{-3} and α_B is the Case B recombination rate coefficient.

These HII regions are radiation bound in terms of ionised hydrogen. Once the ratio of ionised hydrogen to the total hydrogen density drops below 5%, the calculation stops, which defines the outer radius. The inner radius is set to 1% of the expected Strömgren sphere radius (Eq. 3.1), as in Yang & Lidz (2020), allowing us to obtain "ultracompact" HII regions, which can have sizes of $\sim 0.03\,\mathrm{pc}$ (Kurtz 2005). We calculate the emissivities without further iteration on the assumed density, which is adequate for clouds with densities typical of HII regions (Ferland 1996). For the rest

of the parameters in CLOUDY we choose a similar configuration as the one used by Ploeckinger & Schaye (2020).

The `starburst99` outputs are also used to calculate the $L_{\rm FUV}$ and Q for each star SPH particle. Thus, the look-up luminosity tables from CLOUDY are used to interpolate the star SPH particles in terms of age, metallicity and Q. As a result, we obtain the respective HII luminosity for each star particle for the FIR emission lines. This luminosity is then used as part of the total contribution of the ISM phases (DIG + neutral atomic gas + dense molecular gas + HII regions) for a given simulated galaxy.

3.3 Individual line luminosities

In this section, we present the predictions of individual lines from our model and compare them with observations. For each line, we first examine the relation between the line luminosity and the SFR of the galaxy at different cosmic epochs. Our predictions of the line luminosity–SFR relation are also compared with observational measurements collected and homogenised from published work (see Appendix 3.6). Then we check the contribution to the line luminosities from each of the ISM phases, i.e., the DIG, HII regions, neutral atomic and dense molecular gas. We mainly discuss the results of the five FIR emission lines: [C II] at 158 μm, [O I] at 63 μm, [N II] at 205 μm, [O III] at 88 μm, and [N III] at 57 μm. The other three emission lines listed in Table 3.1 behave similarly to their aforementioned pairs (e.g. [O III] at 52 μm is similar to [O III] at 88 μm). We release all the data in a Zenodo repository as described in Appendix 3.6.

3.3.1 [C II] 158 μm

The SFR–$L_{\rm [C\,II]}$ relationship

The most important and brightest of these FIR lines is [C II] at 158 μm. This line follows a clear trend with SFR (e.g. Stacey et al. 1991; Brauher et al. 2008; Stacey et al. 2010), and it is used as a SFR tracer at different redshifts (e.g. De Looze et al. 2014; Herrera-Camus et al. 2015; Schaerer et al. 2020). In Fig. 3.2, we show the relationships between SFR and [C II] luminosity ($L_{\rm [C\,II]}$) for the seven redshift slices analysed in this work. We compare the predictions from the EAGLE simulations REF-L100N1504 and RECAL-L0025N0752 with predictions from other simulations (Olsen et al. 2015, 2018a; Katz et al. 2022), semi-analytic models (Lagache et al. 2018; Popping et al. 2019) and linear relationships derived from observations (De Looze et al. 2014; Schaerer et al. 2020). We also plot the linear relationships that we infer from our model (see Appendix 3.6) and extrapolate them outside the dynamic range covered by the simulations but within $10^{-3.5}{\rm M}_\odot\,{\rm yr}^{-1} < {\rm SFR} < 10^{3.5}\,{\rm M}_\odot\,{\rm yr}^{-1}$. With this extrapolation, we can compare the observations with high SFR values that the simulations do not cover. This is necessary as high-z observations generally only include galaxies with very high SFR due to sensitivity limits. In general, the agreement between our models, observations and other models is good within the typical scatter $\sim$0.4 dex (De Looze et al. 2014), especially at $z = 0$ and $z = 6$, where there are more observational constraints.

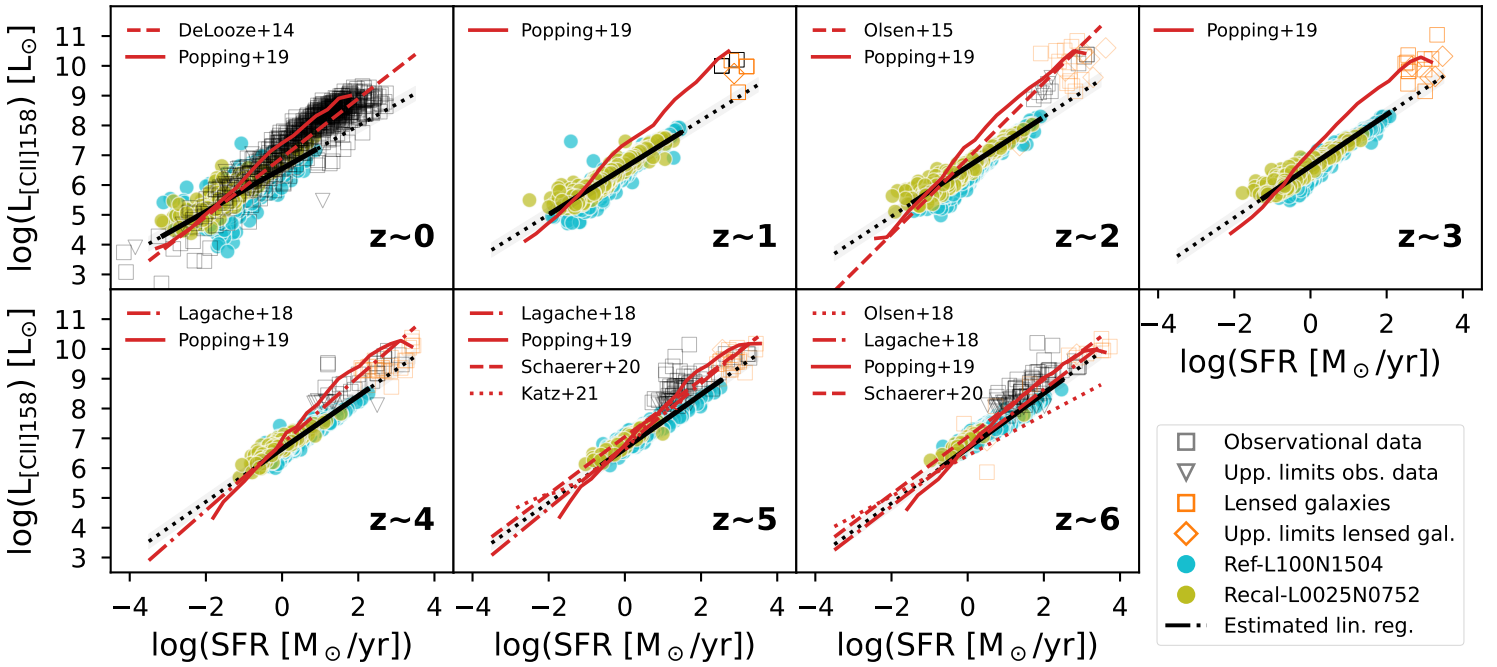

Figure 3.2 – *SFR–$L_{[C\,II]}$ relation for all redshift slices used in this work. We compare the obtained relations from the EAGLE simulations* Ref-L100N1504 *and* Recal-L0025N0752 *with predictions from other simulations (Olsen et al. 2015, 2018a; Katz et al. 2022), semi-analytic models (Lagache et al. 2018; Popping et al. 2019) and linear relations derived from observations (De Looze et al. 2014; Schaerer et al. 2020). Linear relations inferred from our models are shown as black solid lines over the dynamic range covered by the simulations and extrapolated to lower and higher SFRs as black dotted lines, with the grey shaded area representing the 1 σ error. Collections of observational data (Appendix 3.6) are plotted as grey squares for detections and as grey triangles for upper limits. For lensed galaxies, the markers are plotted in and the upper limits that affect both the SFR and luminosity are plotted as diamonds.*

In Ramos Padilla et al. (2021) we showed that the SFR–$L_{[C\,II]}$ relationship at $z = 0$ could be reproduced with a model similar to that implemented in this work. Therefore, it is not surprising that the current model still reproduces this SFR–$L_{[C\,II]}$ relationship. Compared to Popping et al. (2019), the only other model that covers the same redshift range as this work for [C II], we find a flatter slope in the SFR–$L_{[C\,II]}$ relationship. Especially at $z = 1$–3, the differences in the slopes can lead to up to ∼1.8 dex change in $L_{[C\,II]}$ at SFRs around $1000\,M_\odot\,yr^{-1}$, but the relationships are more similar at other redshifts. The reason for these discrepancies may reside in the different galaxy formation physics in EAGLE and the Santa Cruz SAM used in Popping et al. (2019). Unfortunately, the comparison between the two galaxy formation models is out of the scope of this work.

At $z = 2$, the linear relationship of Olsen et al. (2015) predicted from seven simulated galaxies shows a behaviour similar to that of Popping et al. (2019) and agrees with the estimated scatter of our linear regression (0.2 dex) at SFRs around $1\,M_\odot\,yr^{-1}$. Over the redshift range $z = 1$–3, the extrapolation of our relation show a potential small offset compared with observations. However, this discrepancy is not significant, taking into account the small sample size, large scatter (around 0.5 dex), potential bias towards galaxies with high line luminosities, and systematics in deriving luminosities and SFR. Furthermore, altering the assumptions in our modelling process could affect our predictions, as we show in Sect. 3.5.

At $z = 4$–6, most of the models and observations match well with the linear relationships derived from our predictions. Similar predictions from Vallini et al. (2015), Leung et al. (2020) and Carniani et al. (2020), which are not shown in the plot, also agree at $z = 6$, indicating that the SFR–$L_{[C\,II]}$ relationship can be tight at higher redshift, even if there is some scatter in the data related to observational errors (Schaerer et al. 2020). This demonstrates that our physically motivated model of the ISM is valid not only for estimating the luminosity of [C II] in the local Universe but also up to $z = 6$.

[C II] deficit

Although a linear SFR–$L_{[C\,II]}$ relation is commonly used to assume that $L_{[C\,II]}$ is a good SFR tracer, observational data of local galaxies show a decrease of $L_{[C\,II]}$ at IR luminosities above $10^{12}\,L_\odot$, known as the "[C II] deficit" (e.g. Díaz-Santos et al. 2017; Herrera-Camus et al. 2018b). However, recent observations of $z > 4$ galaxies show no evidence of such a deficit (e.g. Matthee et al. 2019; Carniani et al. 2020; Schaerer et al. 2020). In Ramos Padilla et al. (2021), we examined this deficit by comparing the SFR–$L_{[C\,II]}$ relationship with deviations from the star-forming main-sequence (MS), following the suggestion of Herrera-Camus et al. (2018b). In this work, we compare the $L_{[C\,II]}$/SFR ratio with the offset from the MS (ΔMS) in Fig. 3.3. We estimate the specific SFR (sSFR = SFR/$M_\star$) of our galaxies and normalise this by the derived sSFR for the MS from Speagle et al. (2014), which gives us the ΔMS.

We find that the $L_{[C\,II]}$/SFR ratio almost always decreases with ΔMS. If the decrease in the $L_{[C\,II]}$/SFR ratio extends to a higher ΔMS, it supports the idea that starburst galaxies may not follow the SFR–$L_{[C\,II]}$ relationship shown in Fig. 3.2 (Herrera-Camus et al. 2018b; Ferrara et al. 2019). Therefore, $L_{[C\,II]}$ may not be a good tracer of SFR

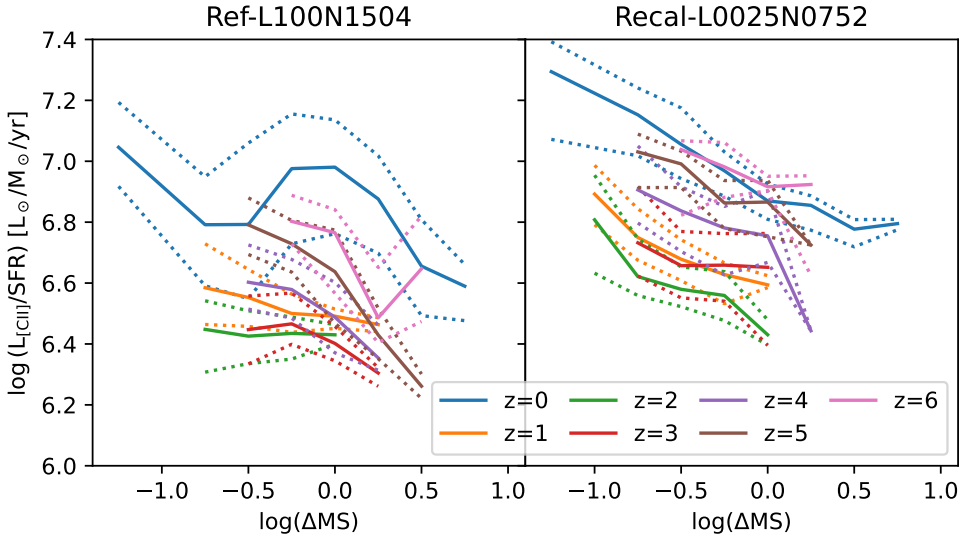

Figure 3.3 – *The ratio [C II]/SFR as a function of the offset from the star-forming main-sequence ΔMS for the simulations* Ref-L100N1504 *(left) and* Recal-L0025N752 *(right). We show the median values of the different redshifts (solid lines) and their 25th and 75th percentiles (dotted lines). We only show the bins with more than 3% of the total sample for the respective simulation.*

for starburst galaxies, as we discuss in Sect. 3.3.6.

Contribution to $L_{[\mathrm{C\,II}]}$ from each ISM phase

We present the average contribution of the ISM phases to $L_{[\mathrm{C\,II}]}$ at each redshift as a function of SFR for the galaxies in the Recal-L0025N0752 simulation in Fig. 3.4. We examine the ISM phase contributions of the Recal-L0025N0752 simulation because we expect this simulation to behave more like observed local galaxies, as shown in Paper I. In general, we find that most of the [C II] emission comes mainly from the atomic phase, especially at $z > 2$, in agreement with the general assumption that the neutral gas is the dominant ISM phase contributing to $L_{[\mathrm{C\,II}]}$ as estimated from observations (e.g. Croxall et al. 2017; Cormier et al. 2019) and suggested by models (e.g. Olsen et al. 2015, 2018a; Lagache et al. 2018). However, the contribution by neutral atomic gas to $L_{[\mathrm{C\,II}]}$ changes with redshift, from ∼20–40% at $z \leq 1$, to ∼70–90% at $z \geq 3$. The most important reason for these differences is the contribution from the DIG. At $z = 0$, the DIG dominates (the contribution is greater than 50%) in most of the galaxies with a SFR$< 0.1\,\mathrm{M_\odot\,yr^{-1}}$, then the DIG contribution reduces to 30% at $z = 2$, and finally it is negligible at $z = 6$. This negligible contribution of the DIG at $z = 6$ does not agree with the estimated contribution of 44% by Olsen et al. (2018a). This result is expected as we estimate the size of the DIG using a physical assumption instead of using the smoothing length as in Olsen et al. (2018a), which leads to a more compact DIG. However, as our modelled SFR–$L_{[\mathrm{C\,II}]}$ relation shows a better agreement with observations, this may imply that a higher contribution from the atomic gas is needed to match the observed galaxies at $z = 6$. Therefore, our

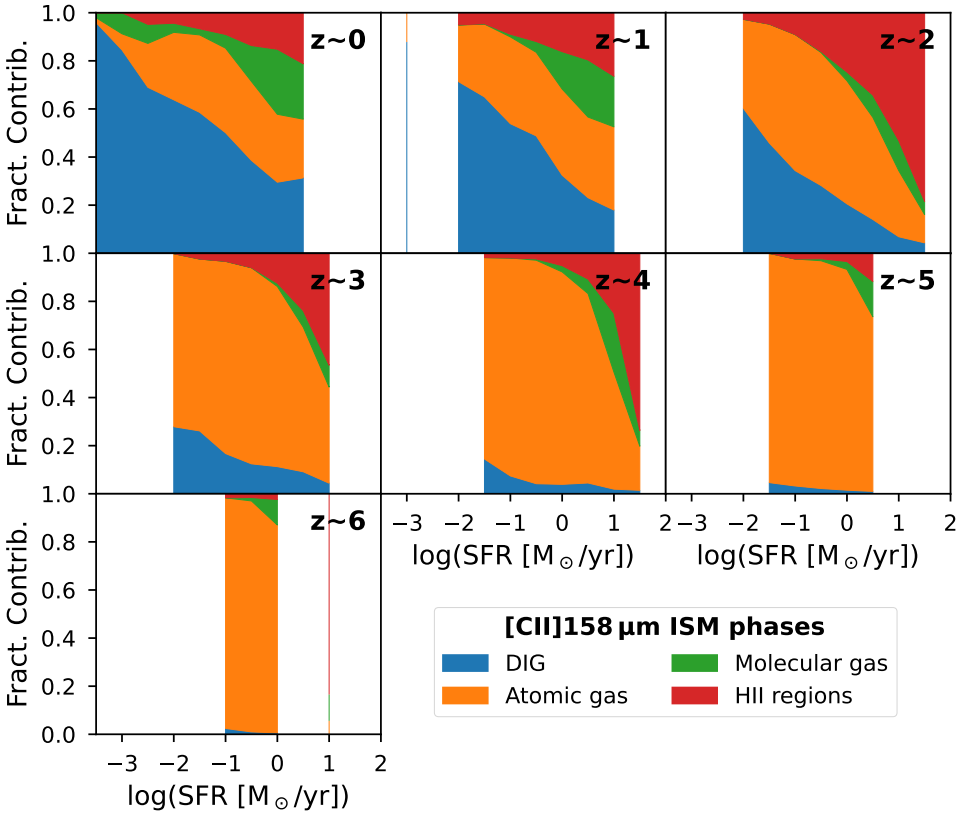

Figure 3.4 – *Contribution from the different ISM phases for the [C II] emission line in* RECAL-L0025N0752. *The regions define the DIG (blue), neutral atomic gas (orange), dense molecular gas (green) and* HII *regions (red).*

estimations seem to favour the atomic gas as the main responsible for the $L_{[\mathrm{C\,II}]}$ at $z = 6$.

On the other hand, the fractional contribution of HII regions to $L_{[\mathrm{C\,II}]}$ reaches its maximum at $z = 2$, where they contribute up to 80% of the luminosity. This trend is expected as HII regions trace young stars, and it is well known that the co-moving star formation rate density reaches its peak value at $z \approx 2$ (Madau & Dickinson 2014). In general, the average contribution by HII regions is ∼20% of $L_{[\mathrm{C\,II}]}$ over all $z < 4$. For molecular gas, the contribution is on average $\lesssim$10% with a maximum at $z = 0$. In both the molecular and HII regions phases we obtain higher fractional contributions to $L_{[\mathrm{C\,II}]}$ at higher SFR. This confirms the results found by Olsen et al. (2015) and in Paper I that the contribution of a given phase to $L_{[\mathrm{C\,II}]}$ depends on the global SFR of the galaxy.

These predictions are not in agreement with recent results from Tarantino et al. (2021) in resolved regions of two local galaxies (M 101 and NGC 6946) where the ionised gas contribution to $L_{[\mathrm{C\,II}]}$ is negligible (average upper limit of 12%). This disagreement

in the ISM phases may be related to the spatial constraints of the observations, as they focus on the arms regions of galaxies where denser neutral gas is expected. For example, Pineda et al. (2018) estimated the $L_{[\mathrm{C\,II}]}$ contribution coming from different regions inside M 51. They found that the region between the arms in M 51, where the diffuse ionised gas is expected be located, is approximately 20% of the total $L_{[\mathrm{C\,II}]}$, compared with ∼80% that comes from the arms and the nucleus. In our model, we calculate the global properties of galaxies (in a 30 pkpc aperture), and thus we expect a higher contribution from the diffuse gas, as this ISM phase can cover more extended regions throughout a galaxy. The information from the diffuse phases may therefore be missing when regions close to the arms of disk galaxies are observed. In general, contributions from different phases will depend on the scales over which we observe the galaxies (Tarantino et al. 2021).

It is important to note that our model seems to overpredict the contribution of the ionised phases (DIG + HII regions) for the RECAL-L0025N0752 simulation. In Fig. 3.5, we examine the relation between the ionised phases to [C II] as a function of metallicity ($12 + \log(\mathrm{O/H})$). We compare the median predictions of REF-L100N1504 and RECAL-L0025N0752 with the relationship obtained by Cormier et al. (2019), based on the data from their work and Croxall et al. (2017), where at a higher metallicity there is a higher contribution to $L_{[\mathrm{C\,II}]}$ from the ionised phase. For REF-L100N1504, we estimate fractional contributions ∼10% higher than those inferred from the Cormier et al. (2019) fitting function, although at $z = 0$ our predictions can be 40% lower for higher metallicities. In RECAL-L0025N0752 we see that the fractional contributions are always above the empirical fitting function, especially at $z = 1$ and $z = 2$, by around 40%. This means that there may be an overestimation in the ionised component of the RECAL-L0025N0752 luminosities.

The simulations and observations presented in Fig. 3.5 are in any case difficult to compare. The simulated galaxies may not have the same metallicity calibration as the observed metallicities in Croxall et al. (2017) and Cormier et al. (2019), and the method by which fractional contributions of the ionised gas are calculated may be different. Nonetheless, in general, we find that the contribution of the ionised phase to [C II] increases with increasing metallicity, as observed by Cormier et al. (2019).

3.3.2 [N II] 122 and 205 μm

The SFR–$L_{[\mathrm{N\,II}]}$ relationship

The emission lines at 122 and 205 μm of [N II] are commonly used to trace ionised gas phases around neutral clouds, because its ionisation potential is only slightly above that of hydrogen. These lines are also used to disentangle the ionised gas contribution to the [C II] luminosity (e.g. Goldsmith et al. 2015; Ferkinhoff et al. 2015; Pavesi et al. 2016; Croxall et al. 2017; Cormier et al. 2019; Langer et al. 2021), as we do in this work. The relationship of these lines with SFR has been explored in the local Universe (e.g. Herrera-Camus et al. 2016; Zhao et al. 2016a); however, at higher redshifts we have very little observational data, especially for the 122 μm line. Due to this lack of data at $z > 0$, we focus here on the [N II] 205 μm line. We present [N II] luminosities at 205 μm as a function of SFR in Fig. 3.6. We compare our results with the linear relationships in the local Universe estimated by Zhao et al. (2016a), Herrera-Camus

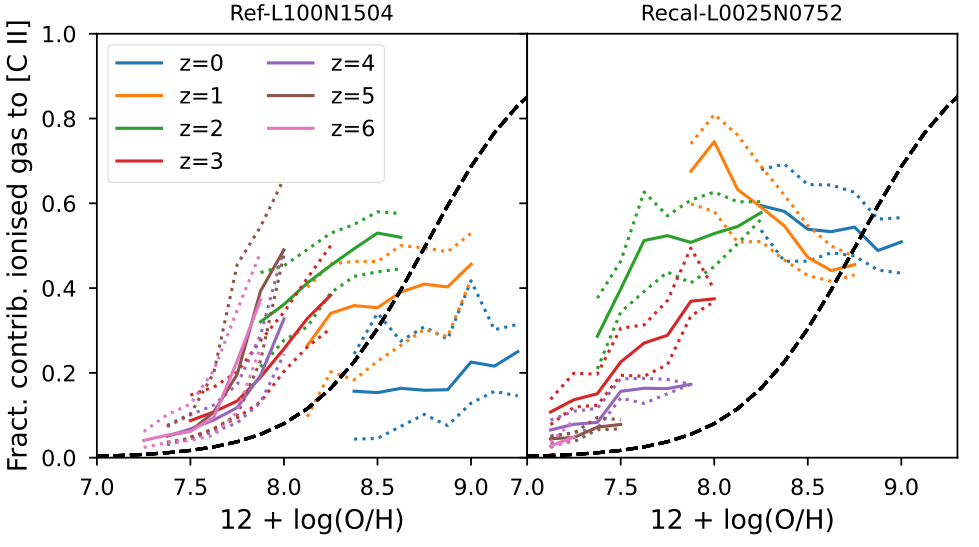

Figure 3.5 – *Contribution of the ionised gas phase to [C II]. We show the median values of the different redshifts (solid lines) and their 25th and 75th percentiles (dotted lines) with respect to the fit to the observational data made by Cormier et al. (2019, black dashed line). In the left panel, we show the predictions of* Ref-L100N1504 *while in the right panel we show the predictions of* Recal-L0025N0752. *We only show the bins with more than 5% of the total sample of galaxies for the respective simulation.*

et al. (2016) and Mordini et al. (2021). For the relation of Herrera-Camus et al. (2016, eq. 10), we assume the values of the collisional excitation coefficients from Tayal (2011), and abundances close to solar. For the relation of Mordini et al. (2021), we use the sample of AGN galaxies assuming the conversion from infrared luminosities (L_{IR}) to SFR of Kennicutt & Evans (2012). We use the AGN sample in Mordini et al. (2021) as those galaxies also follow the relation between SFR and the luminosity of the PAH feature at 6.2 μm (see their Fig. 6).

At $z = 0$, the luminosity predictions of [N II] follow a similar relationship to the observational relations of Zhao et al. (2016a) and Mordini et al. (2021). A potential reason for the difference of one to two orders of magnitude with respect to the Herrera-Camus et al. (2016) relation is that the assumption of solar abundances is incorrect*. If we assume an abundance below solar in the Herrera-Camus et al. (2016) relation, the relation is closer to our model. At $z = 1$ and $z = 2$ our model is consistent with the upper limits of the observational data. At higher redshifts, $3 < z < 6$, our models agree with the observations, although the range of $L_{[N\,II]}$/SFR in the observations is larger than the estimations (almost 0.5 dex at $z = 4$). Most of the data at these redshifts come from the work of Cunningham et al. (2020), who observed 40 gravitationally lensed galaxies with the Morita Atacama Compact Array (ACA) of ALMA. The luminosities of these galaxies depend strongly on the lensing magnification factor,

* Compared to the relations of Zhao et al. (2016a) and Mordini et al. (2021), the N/H abundance that best matches the data is between the values of 0.03 and 0.16 with respect to solar abundance.

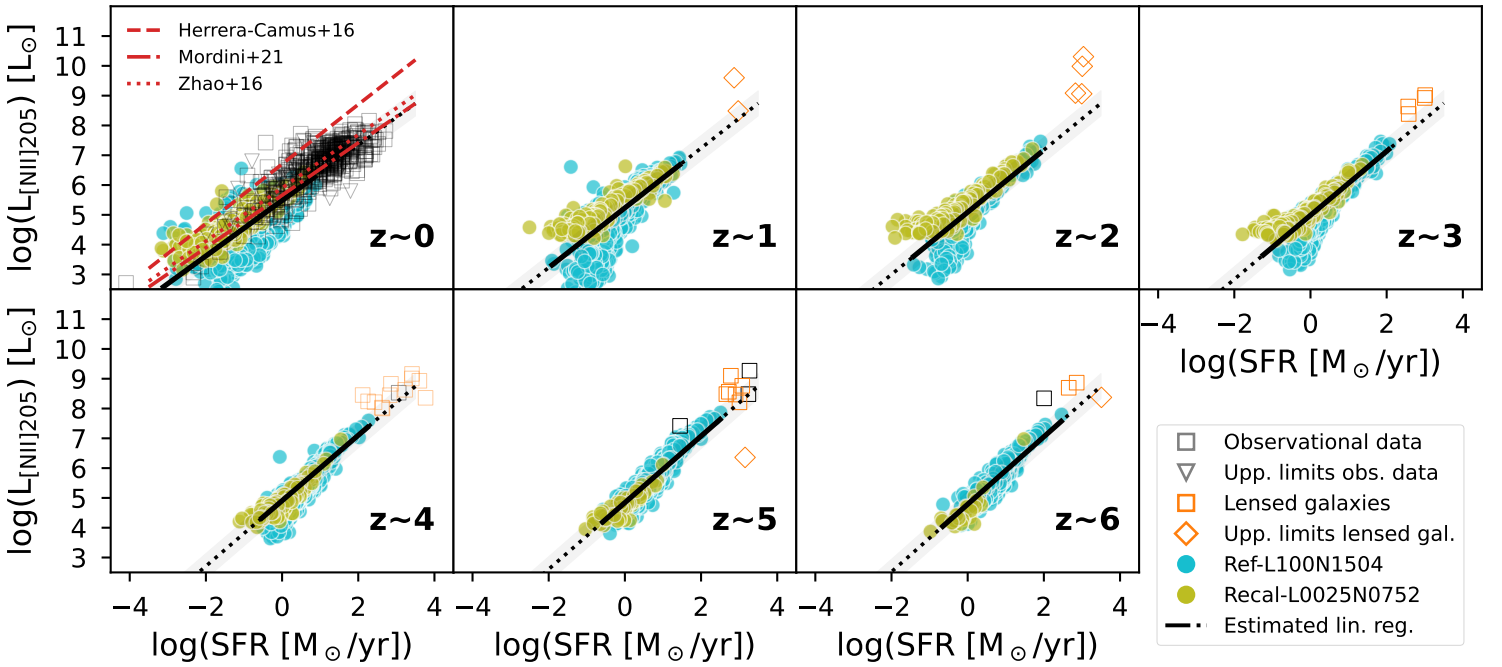

Figure 3.6 – *As Fig. 3.2 for the [N II] 205 μm line. We present the linear relations estimated by Zhao et al. (2016a), Herrera-Camus et al. (2016) and Mordini et al. (2021) at $z = 0$.*

which could create a large scatter in the inferred luminosities (due to uncertainties in the assumed lensing model), as observed in the orange data points. Our model shows a good agreement with the observations, including the lensed galaxies after correction for magnification.

Contribution to $L_{[\mathrm{N\,II}]}$ from each ISM phase

In Fig. 3.7, we show the contribution of the different phases of the ISM to the [N II] line at 205 μm. The luminosity of this line comes mainly from the two ionised phases (DIG and HII regions), as expected from observations (e.g. Langer et al. 2021). Interestingly, the relative dominance of these two phases seems to change with redshift and SFR. At higher SFR most of the luminosity comes from HII regions, while at low SFR, most of the luminosity comes from the DIG phase. The DIG can contribute significantly to the luminosity, which is clearer at lower SFR when there are not so many HII regions. In the local Universe, the contribution from the DIG dominates (∼80%) over the contribution from HII regions. At higher redshifts, $1 < z < 4$, the contribution of the two phases is split relatively evenly. At $z > 5$, HII regions dominate the line emission. In this redshift range, at the highest SFRs, HII regions contribute significantly more than the DIG. The transition point between the phases (the SFR where DIG contributes less than 50% and HII regions more than 50%) is around $1\,\mathrm{M}_\odot\,\mathrm{yr}^{-1}$ at $z = 0$ and decreases to $0.1\,\mathrm{M}_\odot\,\mathrm{yr}^{-1}$ at $z = 6$. There is also a very small contribution from the atomic phase (<4%) at some redshifts, but this is negligible compared to the ionised phases.

3.3.3 [O I] 63 and 145 μm

The SFR–$L_{[\mathrm{O\,I}]}$ relationship

The [O I] emission lines at 63 and 145 μm trace warm gas in neutral clouds and are commonly detected in galaxies in the local Universe (Malhotra et al. 2001). However, the line at 145 μm is fainter than the 63 μm line due to its lower spontaneous decay rate and higher upper level energy (Goldsmith 2019). As we did with [N II], here we focus on the results of the brighter line of [O I], i.e., the emission line at 63 μm.

In Fig. 3.8 we present predictions of [O I] 63 μm luminosity as a function of SFR at different redshifts. Our model at $z = 0$ agrees with the local Universe relationships of De Looze et al. (2014) and Mordini et al. (2021). For the relation of Mordini et al. (2021), we use the sample of star-forming galaxies assuming the conversion from $\mathrm{L_{IR}}$ to SFR of Kennicutt & Evans (2012). At $z > 1$, extrapolation of the linear fits to our model predictions are consistent with the observational upper limits. There are only a few detections at $z = 2$, $z = 3$ and $z = 6$ that can be directly compared with our predictions. There are seven detections in the $z = 2$ bin, of which five are > 0.5 dex from the extrapolated linear fit. However, all of these galaxies are gravitationally lensed (Brisbin et al. 2015), and no corrections were applied to correct for lensing because the magnification factors were unknown (see also Zanella et al. 2018). Therefore, it is reasonable to assume that these five measurements should be treated as upper limits. In the other redshift bins, we observe a very good agreement with the linear relationship because the lensing magnification has been taken into account (Rigopoulou et al. 2018; Zhang et al. 2018; Rybak et al. 2020).

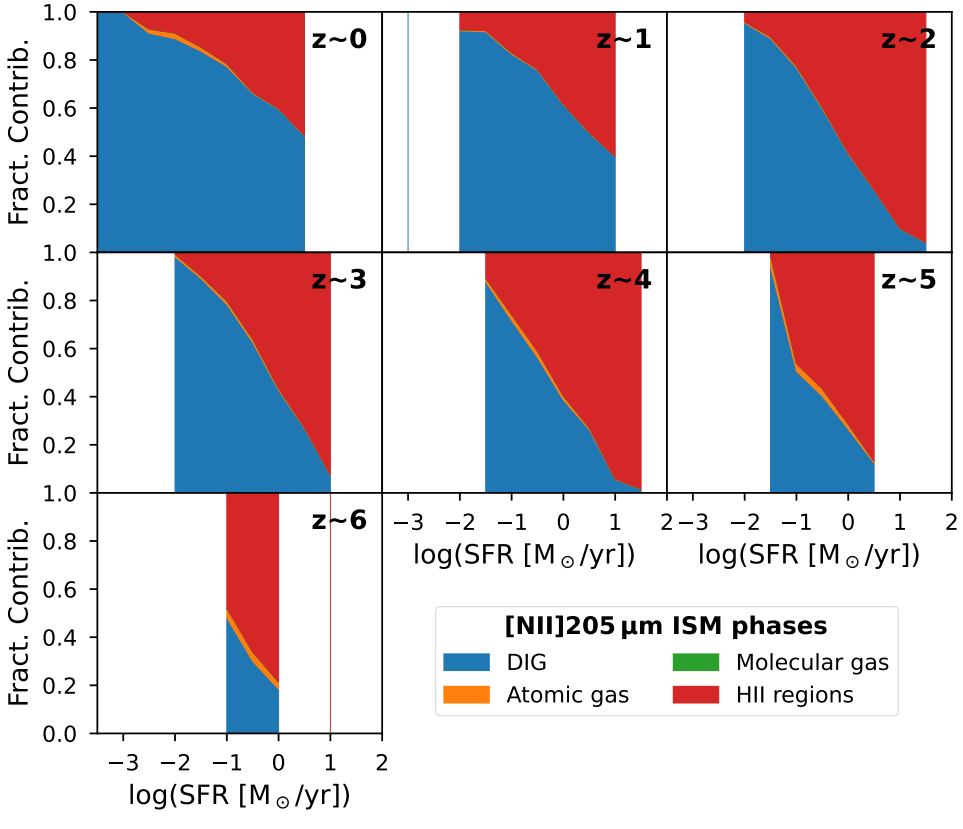

Figure 3.7 – *Contribution from the different ISM phases to the [N II] emission line at 205 µm in* Recal-L0025N0752. *Colour-coding is the same as in Fig. 3.4*

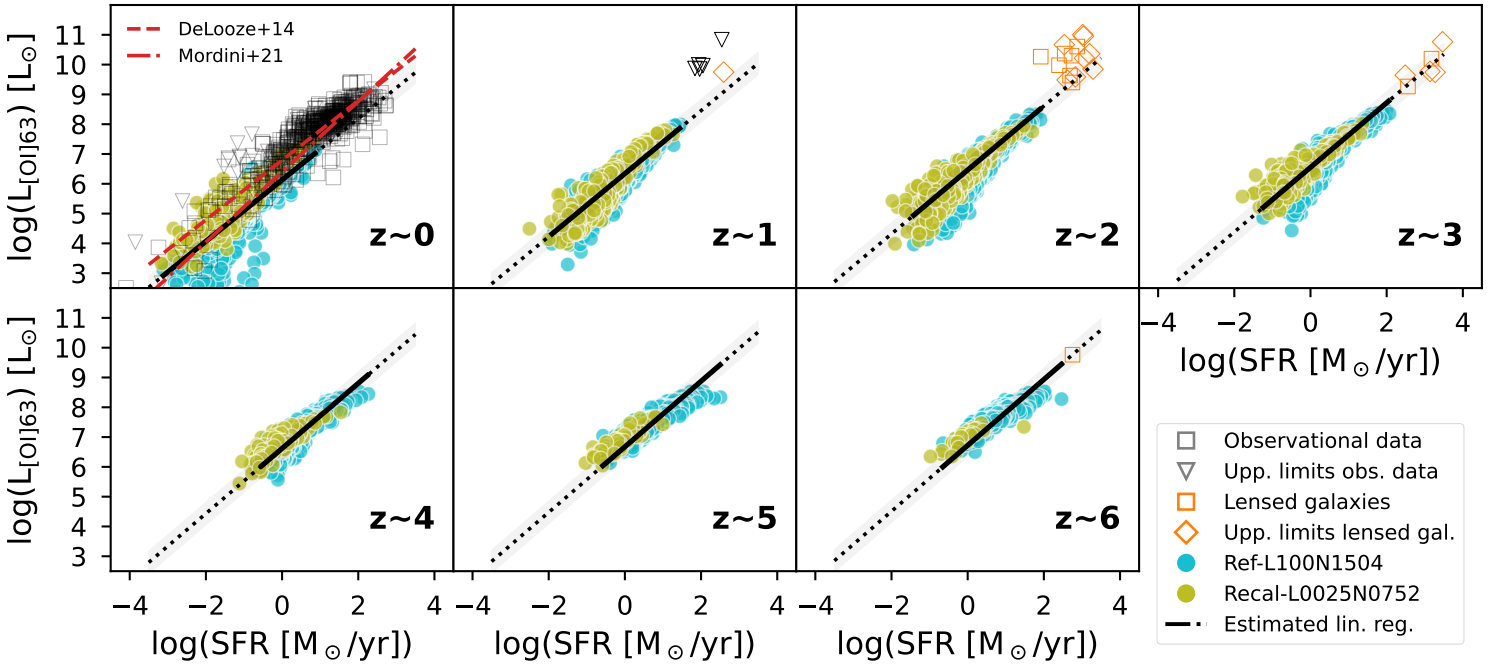

Figure 3.8 – *Similar to Fig. 2 but for the [O I] 63 μm line. We present the linear relations estimated by De Looze et al. (2014) and Mordini et al. (2021) at $z = 0$.*

Other numerical models, not shown in the plots, also predict the [O I] 63 μm line and show similar results. For example, Olsen et al. (2018a) predict the emission of 30 simulated galaxies at $z = 6$ in the MUFASA cosmological simulation. All of their galaxies have SFR$\sim$10 $M_\odot\,yr^{-1}$ and $L_{[O\,I]}\sim10^8\,L_\odot$, very similar to the estimations of the Ref-L100N1504 simulation. However in an updated model, Olsen et al. (2021) use the SIMBA cosmological simulation at $z = 0$, and their predicted [O I] 63 μm luminosities are 1.2 dex above the De Looze et al. (2014) relation shown in Fig. 3.8. This contrasts with our model, which exhibits better agreement with the linear fits of De Looze et al. (2014) and Mordini et al. (2021) at $z = 0$, with a difference of $\sim$0.5 dex at $\log(\mathrm{SFR}[M_\odot\,yr^{-1}]) = 1$. We conclude that the model predictions presented in this work provide a better match with observations across a wide range of redshifts than previous models that only focus on a single redshift slice.

Contribution to $L_{[O\,I]}$ from each ISM phase

Contributions to the [O I] 63 μm line come mainly from neutral clouds, i.e., neutral atomic gas and dense molecular gas, as shown in Fig. 3.9. At $z = 0$, the contribution to $L_{[O\,I]}$ from molecular gas is $\sim$40% while for atomic gas it is $\sim$39%. These percentages change with redshift: the contribution from molecular gas decreases and from atomic gas increases with increasing redshift. At $z = 2$ the percentages are $\sim$7% and $\sim$85%, respectively, while at $z = 4$ they are $\sim$1% and $\sim$98%. The contributions from the other phases are negligible, especially for the HII regions. On average the contribution from the DIG is less than 10%; however, it can be very high ($> 80\%$) in galaxies with very low SFR ($< 10^{-2} M_\odot\,yr^{-1}$) in the local Universe. At $z \geq 3$, the molecular fraction, which is calculated from the line luminosity in the region defined by the radius at which the transition from atomic to molecular H occurs, in the Plummer profile (Equation 31 in Paper I) is very low. At those redshifts, even though the average density of the neutral cloud is higher than in the local Universe, the ISRF is also high, which causes the dominant emission to come from the atomic gas instead of the molecular gas. These results support the understanding that the [O I] 63 μm line originates in warm neutral environments (Malhotra et al. 2001; Goldsmith 2019) even in high-z galaxies ($z\sim6$).

3.3.4 [O III] 52 and 88 μm

The SFR–$L_{[O\,III]}$ relationship

The [O III] emission lines at 52 and 88 μm are the best tracers of ionised gas in the FIR. These lines may be used as SFR tracers as they come mainly from young stars (e.g Ferkinhoff et al. 2010; De Looze et al. 2014; Inoue et al. 2014; Harikane et al. 2020; Yang & Lidz 2020; Yang et al. 2021b). The [O III] 88 μm line has become very important as it may be brighter than the [C II] 158 μm line in galaxies close to the reionisation epoch ($z \gtrsim 7$, Inoue et al. 2014, 2016; Bouwens et al. 2022). As in previous sections, we focus here on the results of only one of the [O III] lines, the [O III] 88 μm line, as we have more observational data for this line. In Fig. 3.10 we present the [O III] 88 μm luminosities as a function of SFR. We compare the predictions from the EAGLE simulations Ref-L100N1504 and Recal-L0025N0752 with observations of individual galaxies and linear relationships between [O III] and SFR derived from

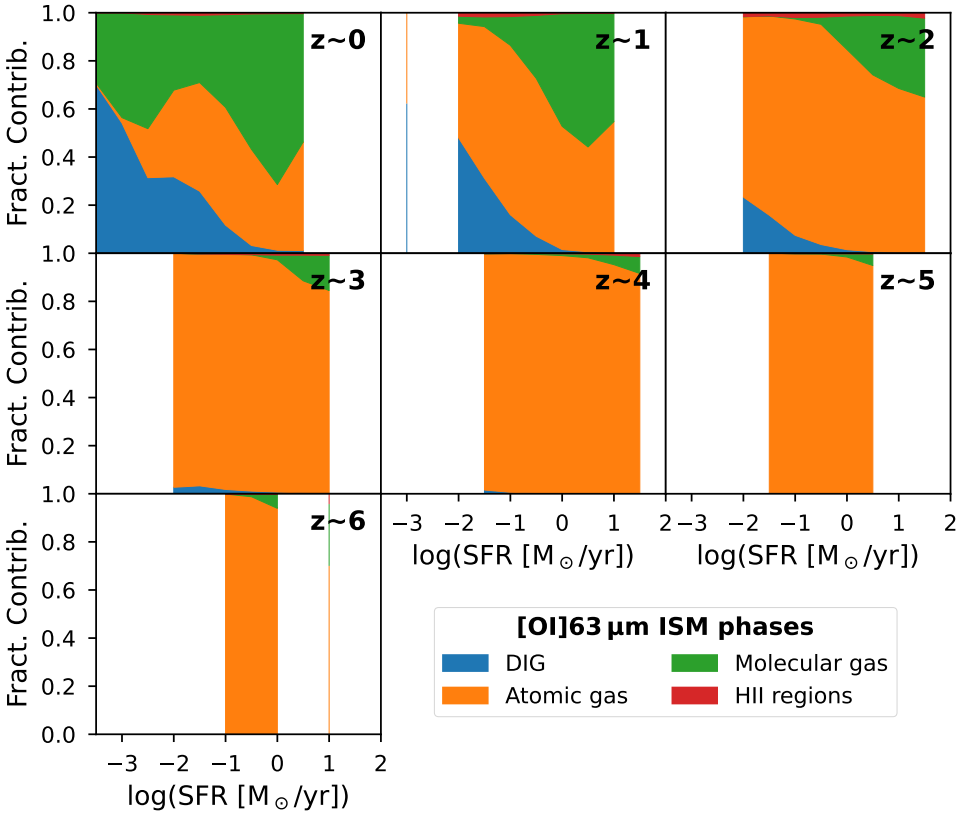

Figure 3.9 – *Contribution from the different ISM phases to the [O I] emission line in* RECAL-L0025N0752. *Colour-coding is the same as in Fig. 3.4.*

observations (De Looze et al. 2014; Harikane et al. 2020; Mordini et al. 2021) and numerical models (Olsen et al. 2018a; Katz et al. 2022; Kannan et al. 2022).

Our predicted SFR–$L_{[\mathrm{O\,III}]}$ relationships largely agree with observations, although it seems that our predicted relationship is steeper than some relations found in the literature. At $z = 0$, we compare our model with the predictions of De Looze et al. (2014) and Mordini et al. (2021). For the relation of Mordini et al. (2021), we use the sample of star-forming galaxies assuming the conversion from $\mathrm{L_{IR}}$ to SFR of Kennicutt & Evans (2012). Both observational relationships appear to have slopes flatter than our model, separated by about an order of magnitude at SFR$= 10\,\mathrm{M_\odot\,yr^{-1}}$. The reason for this difference is that at low SFRs, our luminosity predictions coming from HII regions drop off sharply for both simulations (as we will discuss in the next subsection), leading to a steep slope even though our predicted $L_{[\mathrm{O\,III}]}$ values agree with the observational data at SFR$= 1\,\mathrm{M_\odot\,yr^{-1}}$.

At $z = 2$–4, due to the small sample size in the observational datasets, it is very difficult to properly assess the level of agreement between the observations and our model predictions. In some cases, the observations agree well with the extrapolation

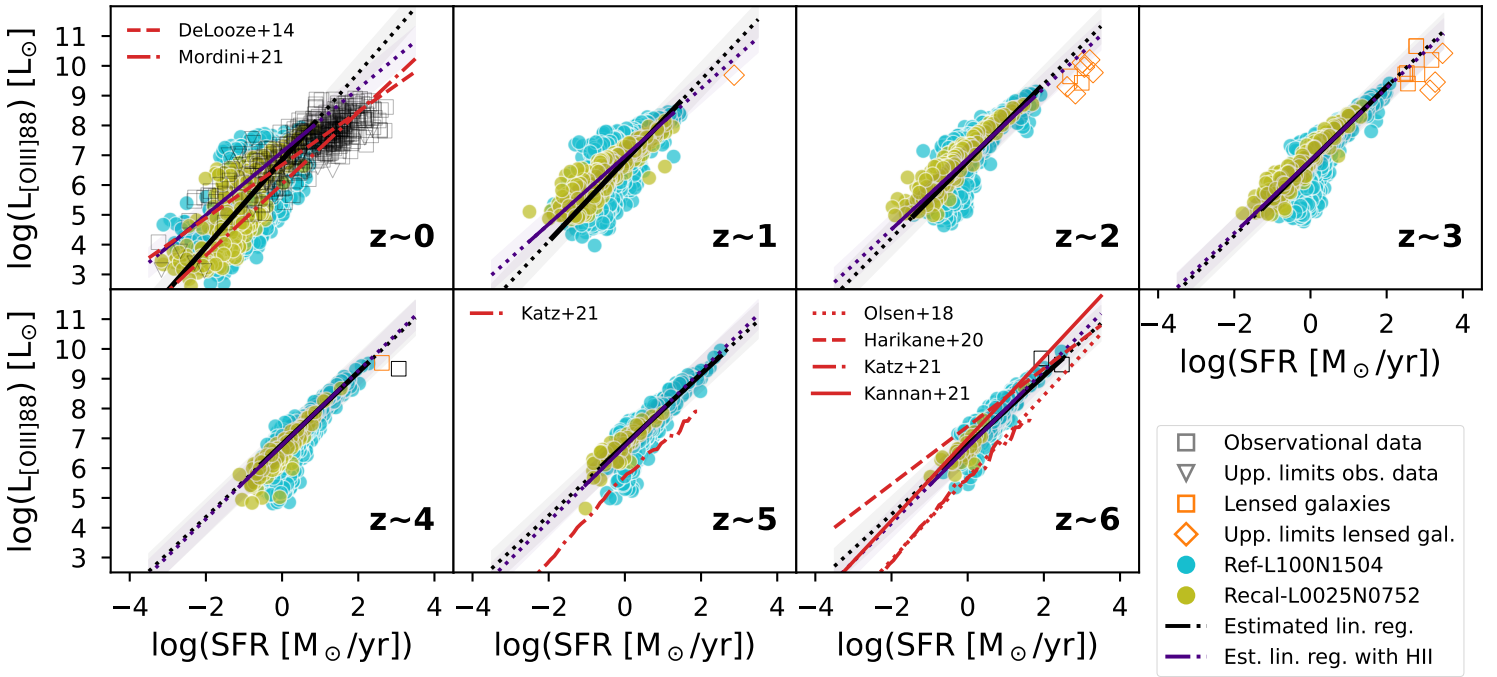

Figure 3.10 – *As Fig. 3.2 for the [O III] 88 μm line. We compare the obtained relations from the EAGLE simulations with relations derived from simulations (Olsen et al. 2018a; Katz et al. 2022; Kannan et al. 2022) and observations (De Looze et al. 2014; Harikane et al. 2020; Mordini et al. 2021). Linear relations inferred from our models using only* H II *regions are shown as indigo solid lines over the dynamic range covered by the simulations and extrapolated to lower and higher SFRs as indigo dotted lines, with the shaded area representing the 1σ error (see Appendix 3.6).*

of our predicted SFR–$L_{[O\,III]}$ relations, with a few exceptions. For example, the galaxy located 1.1 dex below the extrapolation at $z = 4$ is AzTEC 1 (Tadaki et al. 2019). The discrepancy in this galaxy is not related to gravitational lensing, but rather this galaxy has a very low [O III]/[C II] ratio compared with SPT-S J041839-4751.8, a galaxy that closely follows our relation (De Breuck et al. 2019). This means that this galaxy may be an outlier and very different physical conditions can shift its position in the SFR–$L_{[O\,III]}$ relation, as we discuss in Sect. 3.4.

At $z = 5$, we compare our results with those of Katz et al. (2022), which show a similar slope but lower $L_{[O\,III]}$ by ∼1.3 dex. The same model was used to calculate $L_{[O\,III]}$ at $z = 6$, showing that their predictions, together with the models of Olsen et al. (2018a), underestimate $L_{[O\,III]}$ compared with observations by 0.7 dex at SFR∼100 $M_\odot\,yr^{-1}$. This is not the case with our model: it agrees very well with the observations and with the linear relation from Harikane et al. (2020), with a difference of 0.2 dex at SFR∼100 $M_\odot\,yr^{-1}$. It is also interesting to note that the numerical results of Kannan et al. (2022) are very similar to those presented in this work, with the difference of a slightly higher slope in their work, which gives a difference of 0.4 dex at SFR∼10 $M_\odot\,yr^{-1}$.

Contribution to $L_{[O\,III]}$ from each ISM phase

In Fig. 3.11 we show the contribution of the ISM phases to $L_{[O\,III]}$, of which the ionised ISM is the only contributor, as expected. The dominant contributor to $L_{[O\,III]}$ are the HII regions at most redshifts. However at $z = 0$, the emission coming from the HII regions drops sharply when the SFR decreases. The reason for this sharp drop is the low ionising photon flux coming from the star SPH particles in the simulated galaxies at low SFR. This sets the [O III] line luminosities from the HII regions to almost negligible values compared to the DIG, which explains the trend observed in $z = 0$. At $z = 1$, the lack of ionising photon flux affects galaxies less than at $z = 0$, and at these redshifts the HII regions dominate (72%) over the DIG (28%). For redshifts from $z = 2$ to $z = 6$ the contribution from the HII regions changes from 85% to 99%.

3.3.5 [N III] 57 μm

The SFR–$L_{[N\,III]}$ relationship

The [N III] emission line at 57 μm is very similar to the [O III] 88 μm emission line, due to its excitation properties (Table 3.1). Both trace HII regions, with the difference that [N III] is fainter in galaxies with metallicities below solar (e.g. Nagao et al. 2011; Rigopoulou et al. 2018). The relationship between the SFR and $L_{[N\,III]}$ is less well known than the other FIR emission lines presented in this work. In Fig. 3.12, we present the SFR-$L_{[N\,III]}$ relationship. We compare predictions from the EAGLE simulations with the observational sample and the linear relation at $z = 0$ by Mordini et al. (2021) in AGN galaxies assuming the conversion from L_{IR} to SFR of Kennicutt & Evans (2012), as we did in the case of [N II].

At $z = 0$, our model predictions agree with the observations, with a mean offset of 0.2 dex, as they are inside the observational scatter for the overlapping SFR range. However, the $L_{[N\,III]}$ predictions have the same problem as the $L_{[O\,III]}$ predictions: the

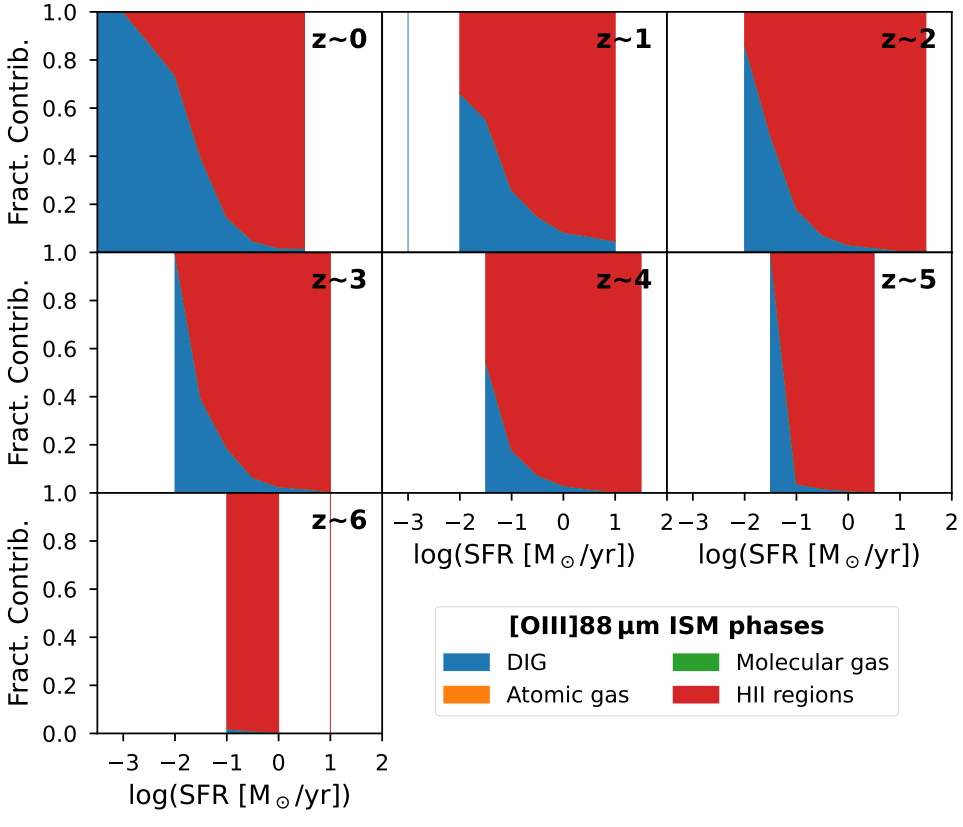

Figure 3.11 – *Contribution from the different ISM phases for the [O III] emission line in* RECAL-L0025N0752. *Colour-coding is the same as in Fig. 3.4.*

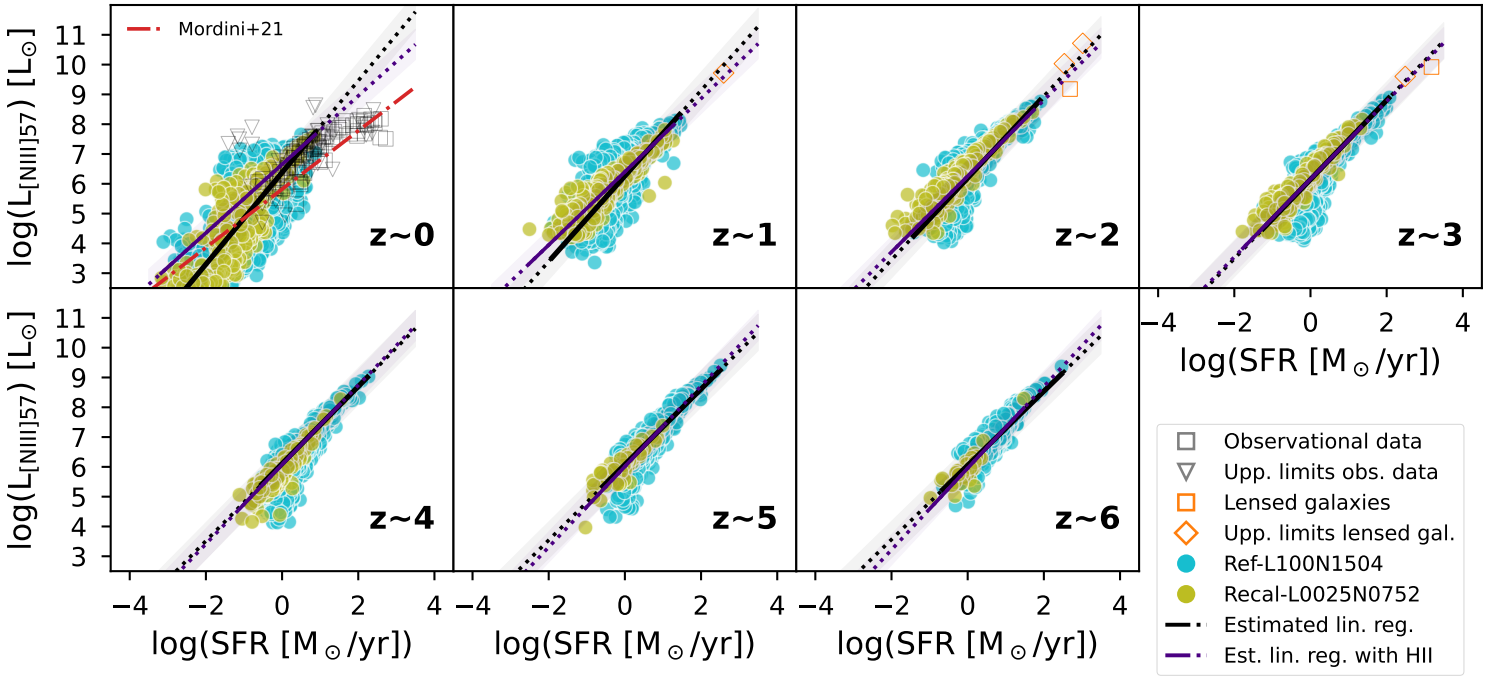

Figure 3.12 – *As Fig. 3.10 for the [N III] 57 μm line. We present the linear relations estimated by Mordini et al. (2021) at $z = 0$.*

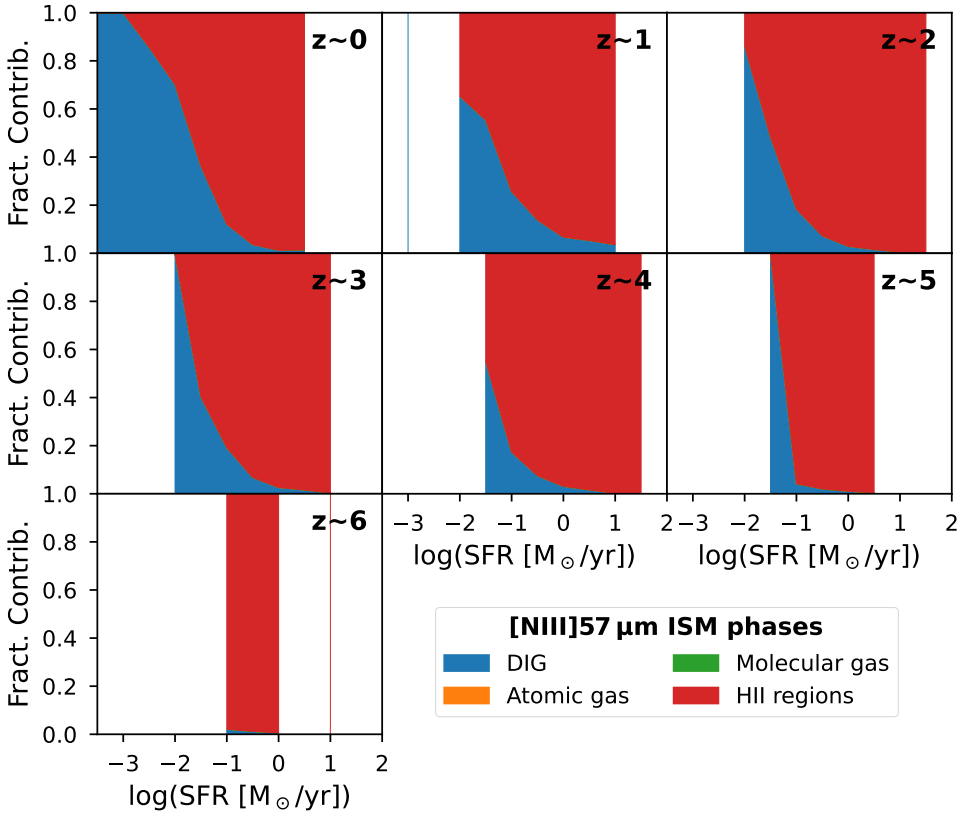

Figure 3.13 – *Contribution from the different ISM phases for the [N III] emission line in* RECAL-L0025N0752. *Colour-coding is the same as in Fig. 3.4.*

low ionising photon flux coming from SPH star particles in simulated galaxies at low SFR. This leads to a steeper relation of $L_{[\mathrm{N\,III}]}$ with SFR than that of Mordini et al. (2021), which follows the other observational results. At $z \geq 1$, even though there is not much information, we find that the extrapolation of our linear fit is consistent with the upper limits and detections in two galaxies at $z \approx 2$ (H-ATLAS J091043.1-000321, Lamarche et al. 2018) and $z \approx 3$ (HERMES J105751.1+573027, Rigopoulou et al. 2018). These two galaxies are magnified by around an order of magnitude (magnification factors of 11.5 and 10.9, respectively), therefore the agreement with our model, after applying the magnification correction, shows that the slope of the linear relationship may be correct for these redshifts.

Contribution to $L_{[\mathrm{N\,III}]}$ from each ISM phase

The contributions of the ISM phases to the [N III] 57 μm emission line are very similar to the contributions of the phases to the [O III] 88 μm line (Fig. 3.11), as shown in Fig. 3.13. The main difference resides in the exact percentages between the two dominant ionised ISM phases (DIG and HII regions). At $z = 0$, the contributions of

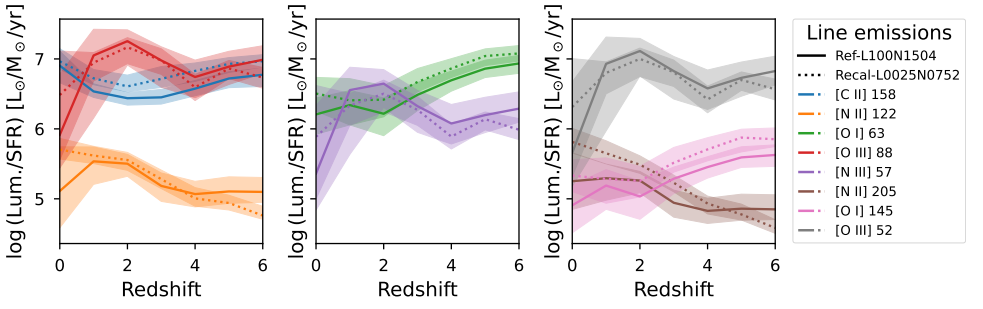

Figure 3.14 – *Evolution of the line luminosity–SFR ratio with redshift for the main FIR lines modelled in this work. We show the median values from* Recal-L0025N0752 *(dotted lines) and* Ref-L100N1504 *(solid lines). The shaded regions correspond to the 25th and 75th percentiles.*

the DIG and HII regions to $L_{[\mathrm{N\,III}]}$ are 38% and 62%, respectively. At higher redshifts these percentages change to 15% and 85% at $z = 2$, 8% and 91% at $z = 4$, respectively. Finally, at $z = 6$ the HII regions are responsible for almost all the $L_{[\mathrm{N\,III}]}$.

3.3.6 Summary of FIR line luminosities

The relationships of the SFR with the different FIR line luminosities presented in Sects. 3.3.1–3.3.5 depend on redshift (Figs. 3.2, 3.6, 3.8, 3.10 and 3.12). We assume a potential redshift dependency in the linear fits of our line predictions (Eq. 3.2), which shows a good agreement with the observations. Nonetheless, we want to quantify how these FIR lines and their dependence on SFR evolve with redshift. To do this, we plot in Fig. 3.14 the ratio between luminosity and SFR (L/SFR) versus redshift for the eight lines modelled in this work.

First, we compare the results of the two sets of simulations Ref-L100N1504 and Recal-L0025N0752. Taking into account the scatter of the predictions, we find a good agreement between them, even though some of the physical properties of galaxies in the simulations are different, as we discuss in Sect. 3.5.

The evolution of the L/SFR ratio is almost flat for most FIR lines at $z \geq 4$. At lower redshifts, $1 < z < 3$, however, there are drastic changes in the L/SFR ratio. For example, the [O III] lines at 52 and 88 μm have higher L/SFR values at $z = 2$ and then decrease sharply towards $z = 0$, by almost an order of magnitude. The opposite occurs for [C II], where at $z = 2$ the L/SFR ratio has lower values than at other redshifts, although the difference is less than 0.5 dex.

These changes are related to the effects present in galaxies during "cosmic noon", where the cosmic star-formation history reaches its peak value at $z \approx 2$ and around half of the stellar mass of the local Universe was formed, affecting the different phases of the ISM traced by the FIR lines in Fig. 3.14 (Madau & Dickinson 2014; Förster Schreiber & Wuyts 2020). This result is expected because the EAGLE galaxies reproduce the observed trend in the SFR density and stellar mass assembly across cosmic time (Furlong et al. 2015) and so the ISM phases evolve accordingly with redshift (Figs. 3.4, 3.7, 3.9, 3.11 and 3.13).

The two [O III] lines have a similar shape in Fig. 3.14, which explains why both lines can be used to constrain gas density and metallicity at high-z (Yang et al. 2021b). The [N II] pair shows a similar behaviour, where the main difference resides in the value of the L/SFR ratio at $z = 0$, although the scatter is especially large at this redshift. This could be important when estimating electron densities from observations from the ratio of these two [N II] lines (Croxall et al. 2017; Langer et al. 2021). Finally, for both [O I] lines, we see a clear difference between the estimated luminosities of the lines of around 1.15 dex over most of the redshift range. This difference of more than an order of magnitude is expected from theoretical models. If the difference in [O I] luminosities is more than 1.15 dex, it may indicate higher kinetic temperatures ($> 400\,\mathrm{K}$) and/or lower gas densities ($\lesssim 10\,\mathrm{cm}^{-3}$) in observations (Goldsmith 2019).

With this information, we can ask which FIR emission line is the best SFR tracer across cosmic time. The FIR line showing the least variation with z is [C II]. However, this tracer may not be ideal in some cases: observations, analytical models and simulations (e.g. Schaerer et al. 2020; Popping et al. 2019; Carniani et al. 2020) suggest that there might be a weak evolution for the SFR–$L_{[\mathrm{C\,II}]}$ relation with respect to redshift. Our model suggests that there is a slight evolution, although less than for the other lines, as shown by the coefficient c_2 in Table 3.7 and Equation 3.2. The luminosity evolution may be related to the active star-formation processes that occur in starburst galaxies, as shown in Fig. 3.3, and in galaxies during "cosmic noon" at $z \approx 2$. The regulation of these star-formation processes is reflected in the changes of the ISM phases (Fig. 3.4), which tend to be more stable in the case of $L_{[\mathrm{C\,II}]}$.

Another possibility is to use $L_{[\mathrm{O\,III}]}$ as an SFR tracer, since the [O III] lines tend to be equal to or brighter than $L_{[\mathrm{C\,II}]}$ in some redshift ranges (e.g. Arata et al. 2020; Carniani et al. 2020; Vallini et al. 2021; Bouwens et al. 2022; Katz et al. 2022). However, our $L_{[\mathrm{O\,III}]}$ predictions at $z = 0$ may be underestimated for lower SFR (see Sect. 3.3.4), which may explain the decrease in the median value of the L/SFR ratio for these lines. It is also possible to use $L_{[\mathrm{O\,III}]}$ together with $L_{[\mathrm{O\,I}]}$ at $63\,\mu$m as an SFR tracer (Mordini et al. 2021) to balance the neutral and ionised components of the ISM in these lines. Nevertheless, this use would require having access to both luminosities ($L_{[\mathrm{O\,III}]}$ and $L_{[\mathrm{O\,I}]}$) at $z < 4$ to confirm the trends presented in Fig. 3.14. The [O I] $63\,\mu$m line alone could also be used to trace the SFR, as this line has similar properties to the [C II] line (Katz et al. 2022). For example, in the predictions of Olsen et al. (2018a), [O I] is brighter than [C II] at $z = 6$, and therefore is more easily observable. The trend of [O I] being brighter than [C II] is also predicted in this work, confirming the results from Olsen et al. (2018a). Therefore, $L_{[\mathrm{C\,II}]}$ may be the best SFR tracer across the entire redshift range of this work ($z = 0$–6), but at high-z other FIR lines such as [O III] and [O I] are also very useful.

3.4 Diagnostic diagrams using FIR lines

We now examine our predictions of FIR emission line strengths from the EAGLE simulations in the context of diagnostic diagrams. Typically, diagnostic diagrams use emission line ratios that reflect the physical conditions of the ISM. In this work, we focus on two diagnostic diagrams: one that normalises the emission line luminosity with SFR for [C II] and [O III] at 88 μm (Harikane et al. 2020), and the other that

uses the ratios between [C II]/[O III] and [N II]/[O I], based on the [O III], [N II] and [O I] lines at 88 μm, 205 μm and 63 μm, respectively. With these two diagnostic diagrams we investigate whether these ratios trace physical quantities related to the ISM, such as radiation field and density.

Other line ratios in the FIR are also of great interest for different types of studies (e.g. Sect. 3.3 of Cormier et al. 2015). We therefore make our model predictions for different emission lines of these simulated EAGLE galaxies publicly available, as described in Appendix 3.6. Similar FIR diagnostic diagrams, such as those presented by De Breuck et al. (2019) and Li et al. (2020a) are not discussed in this work, but plots are provided as supplementary material*.

3.4.1 Comparison with observations

We compare our model predictions with the observational dataset of Appendix 3.6 and the results from the latest version of the SIGAME framework presented by Olsen et al. (2021). The latest version of the SIGAME framework uses the SIMBA simulations (Davé et al. 2019) with the SKIRT (Camps & Baes 2020) radiative transfer code and CLOUDY (Ferland et al. 2017) to predict line luminosities, similar to this work. The SIGAME predictions come from galaxies in two simulated boxes of 25 cMpc and 100 cMpc volumes in the local Universe, similar to the EAGLE box sizes used in this work. Olsen et al. (2021) find that their $L_{[\mathrm{C\,II}]}$ predictions appear to be an extension to higher SFRs of the model predictions presented in Paper I. However, SIGAME tends to have higher line luminosities relative to the SFR–$L_{[\mathrm{C\,II}]}$ relationship (their Fig. 11). In addition, their 25 cMpc box returns even higher line luminosities up to $\sim$0.5–1.5 dex of the 100 cMpc box.

In Fig. 3.15, we compare the $L_{[\mathrm{O\,III}]}$/SFR and $L_{[\mathrm{C\,II}]}$/SFR ratios of our model predictions with the observational data and the SIGAME predictions. For the local Universe, our predictions share a similar range of values with the observational data, which tends to be in the range between $6.5 < \log(L_{[\mathrm{C\,II}]}/\mathrm{SFR}) < 7.6$ and $5.8 < \log(L_{[\mathrm{O\,III}]}/\mathrm{SFR}) < 7.6$. However, most of the simulated galaxies tend to be above ($\log(L_{[\mathrm{O\,III}]}/\mathrm{SFR}) > 6.8$) or below ($\log(L_{[\mathrm{O\,III}]}/\mathrm{SFR}) < 6.0$) the observational data. In contrast, SIGAME predictions tend to have very high [C II] luminosities, which shifts most of the SIGAME simulated galaxies to $\log(L_{[\mathrm{C\,II}]}/\mathrm{SFR}) > 7.5$ values, with the SIGAME values peaking an order of magnitude higher than the observed galaxies and our simulated ratios. This difference is expected from the comparisons presented by Olsen et al. (2021). As noted above, the SIGAME predictions of $L_{[\mathrm{C\,II}]}$/SFR are higher than those described in Paper I, which are similar to those in this work. The SIGAME $L_{[\mathrm{O\,III}]}$/SFR values seem to be similar to our predictions, with values between $6.0 < \log(L_{[\mathrm{O\,III}]}/\mathrm{SFR}) < 8$.

At $z > 0$, there are very few observed galaxies with which we can compare our results. At $z = 2$, we only have two measurements for one galaxy: H-ATLAS J091043.0-000322. These two measurements come from *Herschel* observations following different data reduction methods (i.e. pipeline versions) in Zhang et al. (2018) and Lamarche et al. (2018). ALMA observations presented by Lamarche et al. (2018) show that the [C II] luminosity is around half of the *Herschel* measurement (right data point on

* Zenodo repository at https://doi.org/10.5281/zenodo.6576202

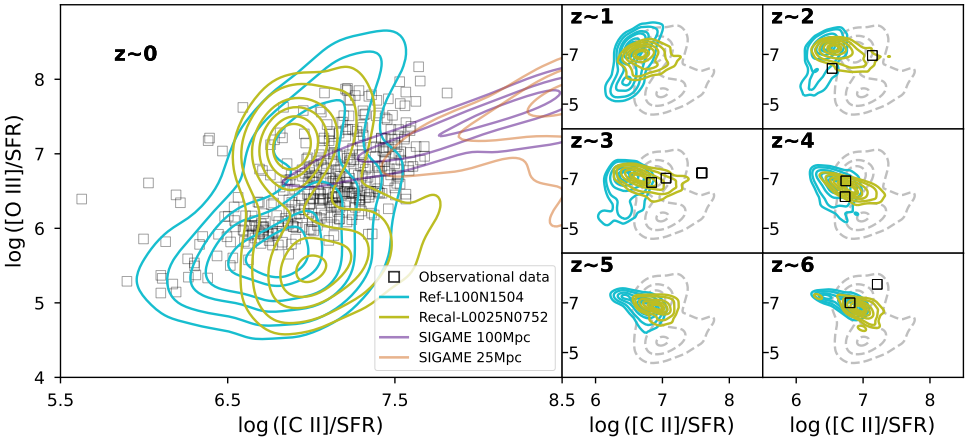

Figure 3.15 – *Diagnostic diagram for the $L_{[\mathrm{O\,III}]}$/SFR and $L_{[\mathrm{C\,II}]}$/SFR ratios, similar to that presented by Harikane et al. (2020). Cyan and olive contours show the model predictions from* Ref-L100N1504 *(cyan) and* Recal-L0025N0752 *(olive). We compare with observational data (black squares) and* SIGAME *predictions (Olsen et al. 2021) for the local Universe in 25 Mpc and 100 Mpc simulation boxes (purple and chocolate contours). All panels with redshifts above zero show the $z = 0$* Recal-L0025N0752 *estimations as grey dashed contours.*

the panel), but this difference cannot be fully explained. Therefore, we assume that H-ATLAS J091043.0-000322 lies somewhere between the two measurements presented in the $z = 2$ panel, agreeing with the predictions from Recal-L0025N0752.

At $z = 3$, we have three measurements for SDP 81 (Valtchanov et al. 2011; De Looze et al. 2014; Zhang et al. 2018) and one for HLock01 (Rigopoulou et al. 2018). For SDP 81, two of the measurements coincide (De Looze et al. 2014 use the values from Valtchanov et al. 2011): the empty squares with $\log(L_{[\mathrm{C\,II}]}/\mathrm{SFR})\sim7.6$. The other measurement for SDP 81 comes from Zhang et al. (2018), which is the leftmost point in the $z = 3$ panel. The main difference between the values of Valtchanov et al. (2011) and Zhang et al. (2018) is the different data reduction methods, and therefore we rely on the more recent results of Zhang et al. (2018). Both SDP 81 and HLock01 are close to the model predictions of Recal-L0025N0752.

At $z = 4$, we have measurements for two galaxies: SPT-S J041839-4751.8 (De Breuck et al. 2019) and AzTEC 1 (Tadaki et al. 2019). Our results agree with the location of both objects, a bit to the left of the results from the $z = 0$ galaxies.

Finally, at $z = 6$ we have measurements for two galaxies: [DWV2017b] CFHQ J2100-1715 companion (Walter et al. 2018) and [MOK2016b] HSC J121137.10-011816.4 (Harikane et al. 2020). [DWV2017b] CFHQ J2100-1715 companion is in the same region where most of our model predictions are at $z = 6$. In contrast, [MOK2016b] HSC J121137.10-011816.4 falls in the same upper region as SPT-S J041839-4751.8 in the $z = 4$ panel. The latter two galaxies have SFRs around $100\,\mathrm{M_\odot\,yr^{-1}}$ and have $L_{[\mathrm{O\,III}]}$ higher than their respective companions in their panels. These high SFRs can explain their positions in the diagnostic diagrams. Harikane et al. (2020) presented two other galaxies in the same redshift range for the ratios presented in Fig. 3.15 be-

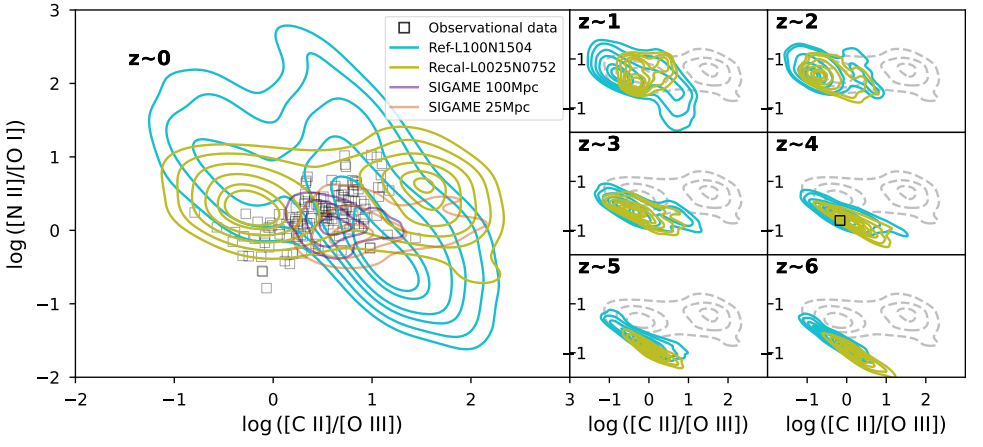

Figure 3.16 – *Diagnostic diagram using four different FIR emission lines comparing the [C II]/[O III] ratio against the [N II]/[O I] ratio. Colour-codes are the sames as in Fig. 3.15.*

sides [MOK2016b] HSC J121137.10-011816.4. Those two galaxies have been removed from the comparison because they have been identified as QSOs.

For the other diagnostic diagram, in Fig. 3.16, we compare the [C II]/[O III] and [N II]/[O I] ratios of our predictions with the observational data and SIGAME predictions. For the local Universe ($z = 0$ panel), we observe that our predictions, observational data and SIGAME predictions tend to be grouped in the region between $-0.5 < \log(L_{[\mathrm{C\,II}]}/L_{[\mathrm{O\,III}]}) < 1.5$ and $-1 < \log(L_{[\mathrm{N\,II}]}/L_{[\mathrm{O\,I}]}) < 1$. The SIGAME predictions match most of the observational data, especially for the 100 cMpc box, although they do not reach the values close to $\log(L_{[\mathrm{C\,II}]}/L_{[\mathrm{O\,III}]}) = 0$ and $\log(L_{[\mathrm{N\,II}]}/L_{[\mathrm{O\,I}]}) = 0$. The difference between the SIGAME 100 cMpc and 25 cMpc boxes comes from the high $L_{[\mathrm{C\,II}]}$ that galaxies in the latter box can have, something that we also commented on at the beginning of this section. Our model behaves similarly to Fig. 3.15, where most of the observations lie between the two concentration regions for Recal-L0025N0752, and the Ref-L100N1504 agree with the observations. Our estimates with a simple model of the ISM are in the same parameter space as observations. Unfortunately, a completely fair comparison cannot be made because of selection bias in both the observational and simulated galaxies.

Unfortunately, as we are using four FIR emission lines, there are few observational data points to which we can compare at higher redshifts. The only galaxy with all four FIR lines is SPT-S J041839-4751.8 at $z = 4$ (De Breuck et al. 2019). Our model matches the position of this galaxy in this diagram, similar to the results presented in Fig. 3.15. The location of this galaxy and most of our predictions at high-z in Fig. 3.16 seem to coincide with some of the observational data at $z = 0$, which may imply that some physical parameters (e.g. sSFR, metallicity and/or density) in these galaxies will be similar at different redshifts.

The comparison presented in Figs. 3.15 and 3.16 show that our model predictions largely match the parameter space of the observational data in the diagnostic dia-

grams. We conclude that our model predictions can be used to interpret the physical parameters of observed galaxies.

3.4.2 Physical parameters in diagnostic diagrams

Now, we use the diagnostic diagrams to infer the ISM physical conditions in galaxies. We use the simulation data of all modelled galaxies and compare the sensitivity of the line luminosity-to-SFR ratio to eight physical parameters as estimated in Paper I: gas mass (M_{gas}), stellar mass (M_*), metallicity ($Z/Z_\odot$), specific star formation rate (sSFR), interstellar radiation field (ISRF), total hydrogen number density in the neutral clouds ($n(H)_{cloud}$), external pressure (P_{ext}) and radius of the neutral clouds (R_{cloud}). We divide these physical parameters into seven ranges to compare them with the observational dataset of Appendix 3.6 and, as a reference, our mock data from RECAL-L0025N0752 at $z = 0$.

We begin by presenting the physical parameters in the $L_{[O\,III]}$/SFR–$L_{[C\,II]}$/SFR diagnostic diagram in Fig. 3.17. We note that the impact of almost all the physical parameters is perpendicular to the observational sample. This effect arises because our mock data also tend to be perpendicular to the observational data, especially at $z = 0$. From these physical parameters, we see that most of the predicted galaxies at the upper-left boundary of the observational contour tend to have higher M_{gas}, M_*, sSFR, ISRF, $n(H)_{cloud}$) and P_{ext}, and lower R_{cloud}. In addition, low values of M_* do not reach the bottom-right limit of the observational contour, while the metallicity spans all over the observational contours and does not have a clear trend.

Harikane et al. (2020) used the $L_{[O\,III]}$/SFR–$L_{[C\,II]}$/SFR diagnostic diagram to explain the physical properties of galaxies at $z = 6$–9 compared with the local Universe. Using simple CLOUDY grids, they found that one of the probable reasons for the location of some of their galaxies in the upper-right region of the diagnostic diagrams was a high ionisation parameter, which is proportional to the ISRF. Their result is similar to what we find from Fig. 3.17; however, it is also important to compare the diagnostic diagram with our metallicity and density results. In terms of metallicity, Harikane et al. (2020) find that $L_{[O\,III]}$/SFR decreases with metallicity while $L_{[C\,II]}$/SFR does not change. We find the same result for $L_{[C\,II]}$/SFR, but the range of change of the $L_{[O\,III]}$/SFR ratio is not directly dependent on metallicity. In terms of density, Harikane et al. (2020) find that both ratios decrease with density, which we also find for our predictions of $L_{[C\,II]}$/SFR – but not for $L_{[O\,III]}$/SFR. A simple reason for this discrepancy is that $L_{[C\,II]}$/SFR depends mainly on density while $L_{[O\,III]}$/SFR is more dependent on other physical parameters: namely metallicity, sSFR and gas mass.

Figs. 3.15 and 3.17 highlight the importance of the [C II] and [O III] emission lines in understanding the physical conditions of gas of galaxies, especially by using their ratio, as other recent studies have done (e.g. Harikane et al. 2020; Arata et al. 2020; Carniani et al. 2020; Vallini et al. 2021; Bouwens et al. 2022) and as we show below. We check its correlation with other FIR lines that trace the neutral and ionised gas components, such as the [N II]/[O I] ratio. We present the physical parameters of the diagnostic diagram of the [C II]/[O III] vs [N II]/[O I] ratios in Fig. 3.18. In this diagnostic diagram, we note how the physical parameters cross the observational data contours in different ways. Interestingly, the physical parameters in the bottom row

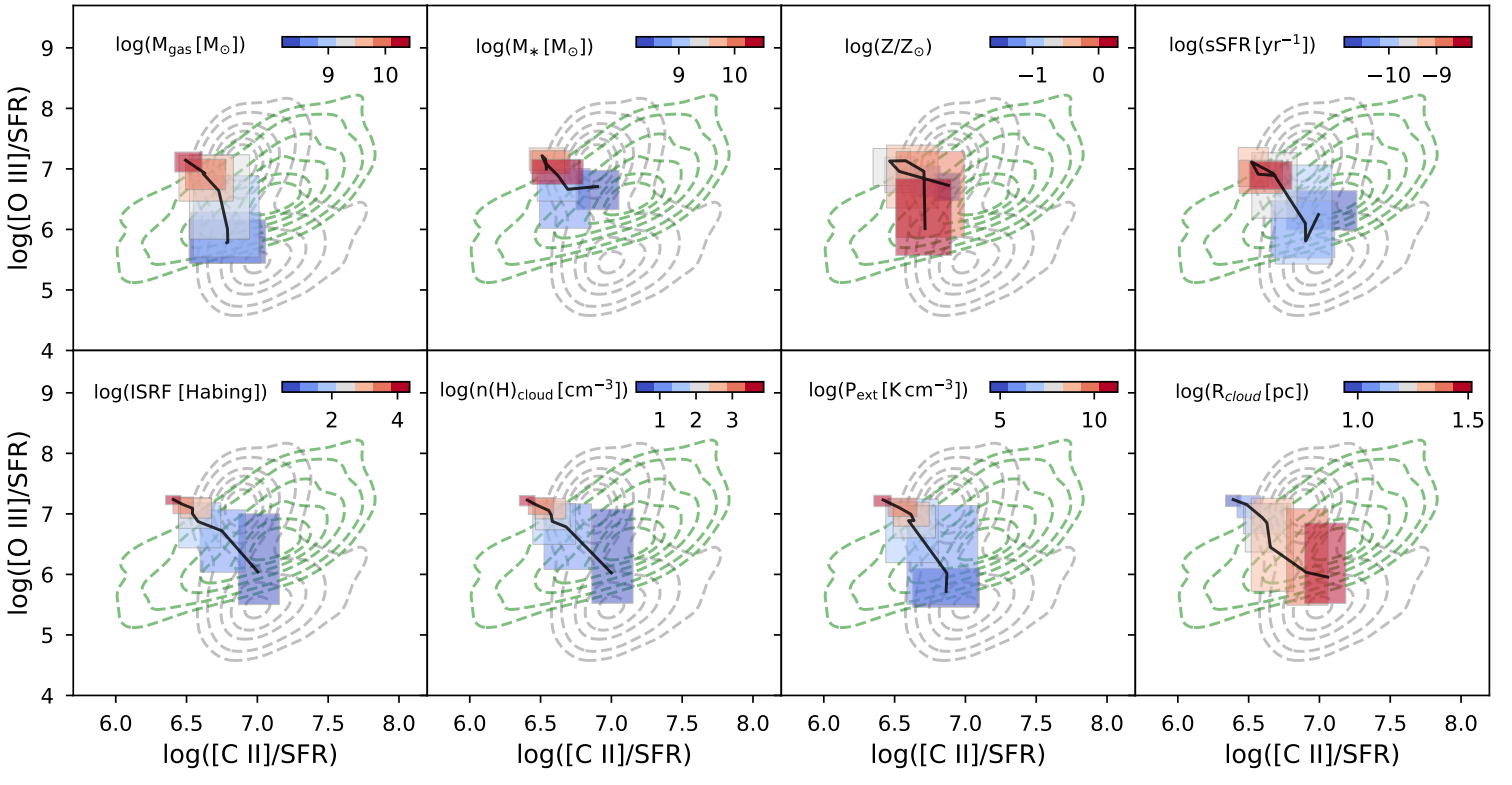

Figure 3.17 – *Physical parameters in the $L_{[\mathrm{O\,III}]88}$/SFR–$L_{[\mathrm{C\,II}]}$/SFR diagnostic plot (see Fig. 3.15). All panels show the $z = 0$ Recal-L0025N0752 model predictions as grey dashed contours and the observational data as green dashed contours. In each panel, colour-coded rectangles represent the 25th and 75th percentiles for a given parameter from low (blue) to high (red) values. The solid black lines connect the median values of each rectangle. The scales for all physical parameters are logarithmic.*

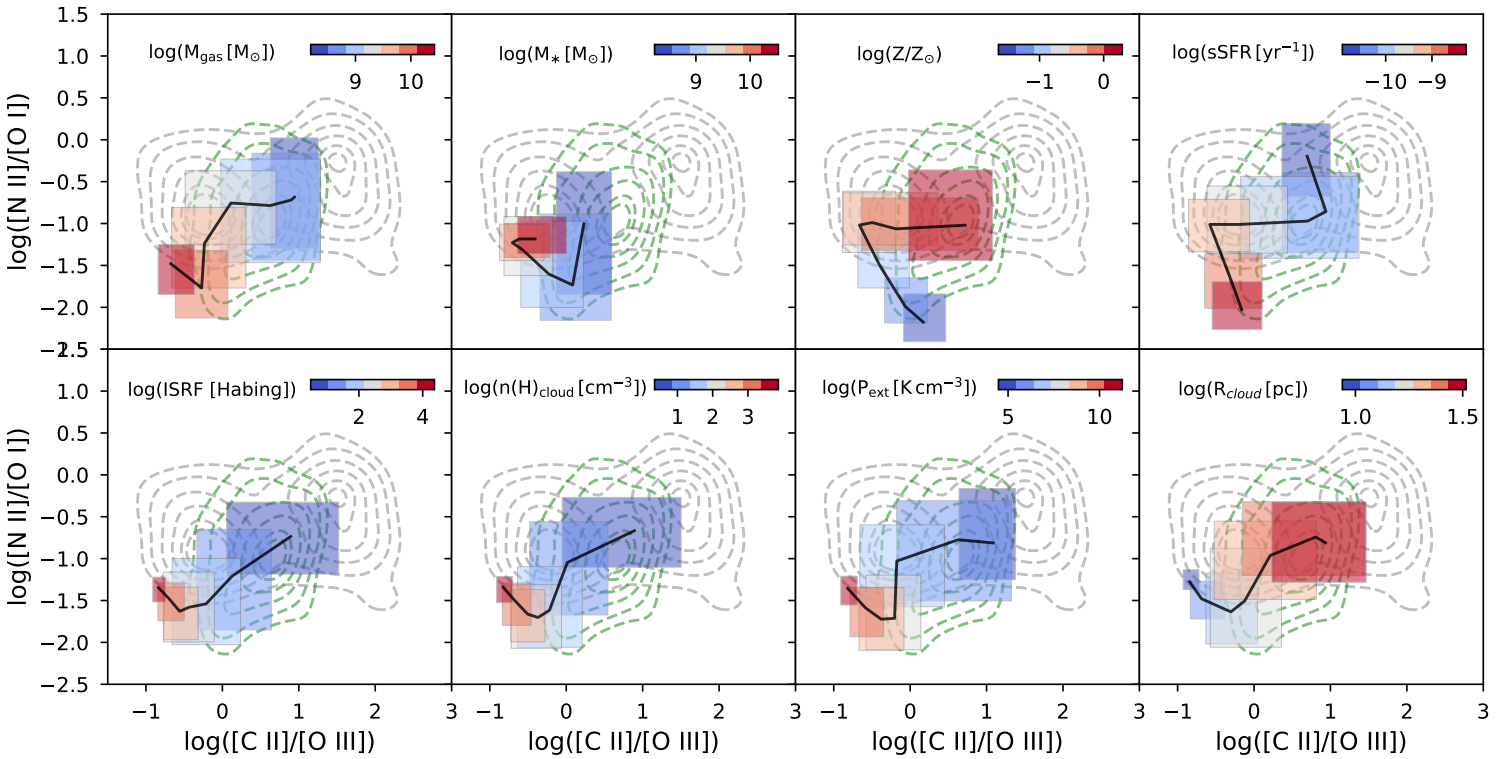

Figure 3.18 – *Physical parameters in the diagnostic plot comparing the [C II]/[O III] ratio against the [N II]/[O I] ratio. Colour-codes are the sames as in Fig. 3.17.*

panels show a "spoon-like" shape showing that most of the observed galaxies have low to moderate values for the ISRF, $n(H)_{cloud}$), and P_{ext}, and moderate to high values for the R_{cloud}. The opposite occurs in the region with low [C II]/[O III], (i.e. high ISRF, $n(H)_{cloud}$ and P_{ext}, and low R_{cloud}), which coincides with our predictions for high-z galaxies ($z > 3$, in Fig. 3.16). In terms of mass, M_{gas} is more extended but follows a similar trend to the parameters mentioned above, while M_* does not show a clear trend. The most intriguing trends in the [C II]/[O III]–[N II]/[O I] diagnostic diagram come from metallicity and sSFR. In terms of metallicity, the [C II]/[O III] ratio seems to be a good tracer for metallicities close to solar, while the [N II]/[O I] ratio is a good tracer for metallicities below $\log(Z/Z_\odot) \lesssim 0.5$, which agrees with some results for high-z galaxies (Arata et al. 2020). In terms of sSFR, both ratios do a good job of separating high and low sSFR of galaxies in a zigzag pattern across the observational sample region. This supports the idea that the [C II]/[O III] ratio can be used for starbursting systems (Vallini et al. 2021).

We have shown that both diagnostic diagrams track the behaviour of the physical parameters presented in the simulated galaxies using the luminosities of the main FIR lines. Our model agrees with the observational data in most of the parameter space, and in some cases with other simulations (e.g. SIGAME). Therefore, it is reasonable to expect that the physical model could be used in inferring the physical parameters in FIR diagnostic diagrams to trace different physical parameters.

Comparing our model predictions with new observational data will give an idea of what kind of physical parameters are expected in those galaxies. At the same time, our model can constrain the expected luminosities of FIR lines when no other measurement is available. However, we emphasise that the modelled FIR lines can also be used to study other types of problems. For example, ratios such as [N II]/[C II] or other diagnostic diagrams can be used to characterise different types of galaxies (e.g. Ultra-luminous infrared galaxies (ULIRGS, where $L_{IR} > 10^{12}\,L_\odot$)) in the local Universe (e.g. Farrah et al. 2013) and at high-z (e.g. Cunningham et al. 2020). Another example is that the ratio between the two [O III] lines can be used to test the efficiency of the black hole feedback in galaxies, as was done with the IllustrisTNG simulations (Inoue et al. 2021). Similarly, ratios like [O III]/[N III] or [O III]/[N II] can be used as a metallicity indicator (e.g. Nagao et al. 2011; Rigopoulou et al. 2018; Fernández-Ontiveros et al. 2021), which will be explored in future studies.

3.5 Discussion

3.5.1 Model assumptions

Although our model emission-line luminosities and ratios are in good agreement with observational results and other simulations, our physically motivated model is based on a number of assumptions. The effect of some of these assumptions is generally not visible in these types of predictions due to uncertainty in the observations and our current understanding of some of the physical processes involved. However, these assumptions may be important for models that try to predict the FIR line emission of galaxies, especially at high-z. In this section we highlight the most important assumptions.

For neutral clouds and HII regions, we assume static spherical geometries to describe the densities and temperatures within those environments. This assumption is made for simplicity, although we know that these structures may not be spherical. Physical processes such as radiation destroy spherical geometries (Deharveng et al. 2010; Peñaloza et al. 2018), which may lead to rough or incorrect line luminosity predictions (Decataldo et al. 2020). However, approximations using mass distributions (Eq. 15 in Paper I) may smooth out these differences, since cloud masses follow scaling relations that seem to be valid for observations at different redshifts (Dessauges-Zavadsky et al. 2019).

A problem with our modelled line luminosities from HII regions is that we assume a fixed density ($\sim$30 cm^{-3}). We can increase the luminosity by increasing this density and vice-versa. The use of different densities could be important when comparing the contributions of DIG and HII regions, especially in lines such as [O III] and [N III] (see Figs. 3.11 and 3.13). This DIG–HII region balance is still unclear in ionised emission lines, and although some estimates exist from optical wavelengths (e.g. Poetrodjojo et al. 2019), a change in this balance can lead to different metallicities, which may affect high-z studies (Sanders et al. 2017). In this work, we calibrate the DIG with observational data at $z = 0$, which may also bias the balance between these two ISM phases. Fortunately, the results presented in this work show that these assumptions seem to be in agreement with the observations, which may represent a likely first step in understanding the DIG–HII region balance.

Finally, our predictions depend on CLOUDY lookup tables, which can give different emissivities depending on the assumed initial abundances or dust configurations (Ploeckinger & Schaye 2020). Different photoionisation models could lead to different interpretations of physical parameters coming from line ratios, such as metallicity or ionisation parameter (e.g. Ji & Yan 2022). Furthermore the intrinsic thermodynamics may not be the same in terms of cooling and heating functions in cosmological simulations and photoionisation models like CLOUDY (Robinson et al. 2022). Therefore, some care must be taken when interpreting the line luminosities predicted by our model.

3.5.2 The use of EAGLE

In this work we use the EAGLE simulation as a proxy of what we would expect to see in the Universe. However, by using a cosmological simulation like EAGLE, our line luminosity predictions of the ISM model are expected to inherit the limitations of the simulation. An example of these limitations is the lack of starburst-like galaxies within EAGLE (Wang et al. 2019b). As we discussed in Paper I, this restricts our comparison at $z = 0$ mainly to SFR below 10 M$_\odot$ yr^{-1} but also limits the comparisons at other redshifts (Katsianis et al. 2017). To compare our predictions with high SFR observations, we extrapolate the linear relations in the range of $-3.5 < \log(\mathrm{SFR}) < 3.5$. However, care must be taken when using this extrapolation.

Another important physical property within EAGLE that affects our predictions is the gas metallicity, which is usually studied through the gas-phase mass–metallicity relation (MZR). Bellstedt et al. (2021) show that the MZR in EAGLE galaxies, as measured by Zenocratti et al. (2020) at $z = 0$, does not behave similarly to other

cosmological simulations or semi-analytical models. As shown in Fig. 4 of Bellstedt et al. (2021), the metallicities in EAGLE have values around $12 + \log(O/H) \sim 9$ for stellar masses between $9 < \log(M_*[M_\odot]) < 11$, which is high compared to observations with metallicities going from $12 + \log(O/H) \sim 8.5$ to $12 + \log(O/H) \sim 9.2$ in the same stellar mass range. This may affect the metallicity that can be recovered from FIR lines if only $z = 0$ is used. However, MZR in EAGLE depends on resolution and box-size used due to the assumed AGN and star formation feedback processes (De Rossi et al. 2017). In Figs. 3.17 and 3.18, we present the predicted luminosities in different boxes and redshifts studied in this work and find that some of the FIR line ratios can be useful to infer the metallicity. Therefore, it will be important to compare the FIR line predictions that trace metallicity with observations in the future more consistently.

Although we base our predictions on EAGLE, we expect that similar physical models and/or cosmological simulations can be used to understand the gas properties with FIR emission lines. For example, Olsen et al. (2021) show that these kinds of predictions can be obtained with a different gas fragmentation scheme using SIGAME, which are similar to observations and our predictions. However, most of the galaxies in SIGAME have higher SFR than those studied in this work. This difference in the sample of simulated galaxies comes from choosing a different simulation (SIMBA instead of EAGLE), which allows for the formation of starburst-like systems. Therefore, although some physical assumptions are different for each model and are limited by the simulation used (e.g. Vallini et al. 2015; Lagache et al. 2018; Popping et al. 2019; Pallottini et al. 2019; Leung et al. 2020), we expect predictions to behave similarly to those presented in this work. In the future, we will compare results from different simulations, such as SIMBA and IllustrisTNG in an efficient way to reduce the bias that the initial assumptions of the simulations may introduce.

3.5.3 Observational data from samples

For our comparison between observations and our model line luminosities, we have collected a heterogeneous sample of observed galaxies with FIR emission line information. We have transformed the luminosities of the lines to the same reference cosmology (Planck Collaboration et al. 2014b) and we have estimated the SFR in most of the cases following the L_{IR}–SFR relation described by Kennicutt & Evans (2012). However, there are some other important issues that may affect this heterogeneous sample.

The most important issue may come from the initial mass function (IMF) assumed to estimate the SFRs. Unfortunately, IMF information is not always present in papers. Assuming a star-formation law that takes into account a standard IMF is a possible solution, as we have tried in this work. However, this may add additional uncertainties. For example, EAGLE uses an instantaneous SFR, which is different from the SFR averaged over the last 10 or 100 Myr typically used in observations. Another problem is that some of the SFR (or L_{IR}) may come from spectral energy distribution (SED) models rather than empirical laws using FIR photometric bands, like IRAS, PACS or SPIRE. In addition, the IR luminosity definition can have different flavours (e.g. L_{IR}, L_{FIR} or L_{TIR}) that use different wavelengths ranges to estimate luminosi-

ties, adding to the spread of SFR estimates (Förster Schreiber & Wuyts 2020) and leading to differences of factors between 1.1 and 1.7 (De Looze et al. 2014; Lagache et al. 2018). Another possibility is to use other SFR tracers, such as Hα-based, UV-based, or radio-based SFRs as shown in many studies (e.g. Wang et al. 2016, 2019a). However, in some cases, such as in high-z galaxies, it can be difficult to obtain such SFRs estimates and other lines like [C II] come into play as a SFR tracer (Le Fèvre et al. 2020), as we have also shown in this work.

In addition to theoretical considerations, some of the FIR line measurements have systematic uncertainties. For example, calibration pipelines or the use of different instruments may affect the comparison, as shown by Lamarche et al. (2018) in the case of H-ATLAS J091043.0-000322, discussed in Sect. 3.4.1. Furthermore, different-sized (or -shaped) apertures in the observational sample may affect the analysis of the fixed-size aperture we selected for the simulated galaxies (30 pkpc). For example, the balance between DIG and HII regions as main contributors to the ionised lines could depend on the selected aperture, as indicated by Mannucci et al. (2021). Finally, gravitational lensing can introduce a large uncertainty in the luminosities, which can be reduced by a factor of 30 to 40 in some galaxies when corrected by the magnification factor (e.g. in Eyelash and SPT-S J041839-4751.8, Zhang et al. 2018; Rizzo et al. 2020, respectively). In addition, these magnification factors may change depending on the observed region of the galaxy (Lamarche et al. 2018). In this work, we highlight those galaxies to warn about these potential complications, but in general, they show good agreement with our predictions.

3.6 Summary and Conclusions

We have modelled FIR emission lines by post-processing EAGLE cosmological simulations using CLOUDY. We have predicted the luminosities of the eight most important lines in the FIR up to $z = 6$, using a physically motivated model that traces four different ISM phases: dense molecular gas, neutral atomic gas, diffuse ionised gas and HII regions. We have also collected a sample of observed galaxies from the literature with which to compare with our predictions. Our main conclusions are as follows.

1. Predictions from our model replicate observed galaxies in the SFR–FIR line luminosity relationship to an average degree of 0.5 dex over the range $z = 0$–6, which is reasonable considering the observational measurements errors. We also compare with different models showing similar level of agreement. We model the SFR–FIR line luminosity relationship for each of the eight lines with a linear relation, each of which shows a slight evolution with redshift.

2. We have presented the expected contributions of each ISM phase to each FIR line. These contributions change as a function of SFR. For the [C II] 158 μm line, the main contributor is the neutral atomic gas, with considerable contributions from HII regions at $z = 1$–4 and the DIG at $z < 2$, which may be related to metallicity. For the [N II] lines at 122 and 205 μm, the DIG contributes more than HII regions in the local Universe, but the opposite is true at high-z, where HII regions seem to dominate over the DIG. For the [O I] lines at 63

and 145 μm, the contribution of dense molecular gas is important in the local Universe. However, the atomic gas is dominant at high-z. Finally, for the [O III] lines at 52 and 88 μm and [N III] at 57 μm, we show that HII regions dominate, with important contribution from the DIG at low SFR in the local Universe.

3. Our predictions indicate that [C II] may not be a good SFR tracer for starburst galaxies, since the [C II]/SFR ratio seems to decrease as a function of the offset from the star-forming main-sequence. However, compared to the other FIR lines, [C II] seems to be the best SFR tracer due to its weak redshift evolution. [O III] and [O I] may also be good SFR tracers. Nonetheless, our predictions of [O III] at $z = 0$ may be underestimated, and more observations of [O I] are necessary at $z < 4$ to confirm our predictions.

4. We compare our predictions in two diagnostic diagrams, and we find reasonable agreement with observations. We compare $L_{[\mathrm{O\,III}]}$/SFR and $L_{[\mathrm{C\,II}]}$/SFR ratios and find that mock galaxies at high-z tend to have higher $L_{[\mathrm{O\,III}]}$/SFR ratios and slightly lower $L_{[\mathrm{C\,II}]}$/SFR ratios than galaxies in the local Universe. We also compare the [C II]/[O III] and [N II]/[O I] ratios and find that mock galaxies at high-z tend to have lower [C II]/[O III] and [N II]/[O I] ratios than galaxies in the local Universe.

5. Finally, we have examined the impact of physical parameters on these diagnostic diagrams. When we compare physical parameters to line luminosities, we find that $L_{[\mathrm{C\,II}]}$/SFR and $L_{[\mathrm{O\,III}]}$/SFR ratios trace hydrogen density and ISRF well in the mock galaxies. However, these ratios are not good metallicity tracers, because $L_{[\mathrm{O\,III}]}$/SFR does not evolve linearly with metallicity and $L_{[\mathrm{C\,II}]}$/SFR does not change with metallicity. Furthermore, we find that [C II]/[O III] and [N II]/[O I] ratios can be good metallicity and sSFR tracers. For example, [C II]/[O III] can trace metallicities close to solar and [N II]/[O I] below solar. On the other hand, we can identify systems with different sSFRs by means of both [C II]/[O III] and [N II]/[O I] ratios, which can be very useful for improving calibrations of [C II] as a SFR tracer.

In the future we expect to use this model and its predictions to understand the effect that AGN can have on the ISM, as well as the physical parameters traced by these lines and their ratios. We make our model predictions and collected observational sample publicly available to allow potential users to compare with their work and/or interpret new observations. We envisage that our predictions will also be useful in planning for future FIR missions.

Acknowledgements: The authors thank Karen Pardos Olsen for providing the data from the SIGAME framework. This research has made use of the 3Mdb database* (Morisset et al. 2015) from which initial experiments to model HII regions were performed. This research made use of Astropy,† a community-developed core Python package for Astronomy (Astropy Collaboration et al. 2013, 2018). This research made use of several Python packages, among them: numpy (Harris et al. 2020), pandas (Wes McKinney 2010) and matplotlib (Hunter 2007). This research has made use of the SIMBAD

* https://sites.google.com/site/mexicanmillionmodels

† http://www.astropy.org

database, operated at CDS, Strasbourg, France. This research has made use of the NASA/IPAC Extragalactic Database (NED), which is funded by the National Aeronautics and Space Administration and operated by the California Institute of Technology. This research has made use of NASA's Astrophysics Data System Bibliographic Services. The EAGLE simulations were performed using the DiRAC-2 facility at Durham, managed by the ICC, and the PRACE facility Curie based in France at TGCC, CEA, Bruyères-le-Châtel. We would like to thank the Center for Information Technology of the University of Groningen for their support and for providing access to the Peregrine high performance computing cluster.

Appendix

Observational sample

We have collected measurements from the literature of the FIR emission lines predicted in this work. The observational sample is a heterogeneous selection of galaxies that covers the redshift range between 0 and 6.5. We present references of the works used in this sample of galaxies together with the number of measurements available per line in Table 3.4. When possible we recalculated the luminosities and SFRs onto the ΛCDM cosmology used in this work (Planck Collaboration et al. 2014b). When the cosmology is not explicitly mentioned, we assume it is the same as used in this work, so no corrections are applied. We flag those measurements where a magnification factor is involved in estimating a luminosity due to gravitational lensing. We estimate SFR in most of the literature samples from the infrared luminosity (L_{IR}) using the relation of Kennicutt & Evans (2012), which assumes a similar IMF (Kroupa) to the one used in EAGLE (Chabrier). Unless stated otherwise, we use the same infrared luminosities as published in the respective works. In cases where L_{IR} is unavailable or is unreliable due to measurement error, we use SFR estimates from other works. Additionally, we remove strong AGN galaxies (i.e. QSO and Blazar). This observational sample is available as a supplementary material in a Zenodo repository, at `https://doi.org/10.5281/zenodo.6576202`.

We now comment on some references for which special attention is needed.

Brauher et al. (2008): We use the median line flux in galaxies with more than one measurement. L_{IR} is calculated using the IRAS 60 and 100 μm fluxes as described in Brauher et al. (2008). For two galaxies (3C 368 and Z 25-7), we adopt the preferred Local Group velocity in NED to estimate the distance to these galaxies (Karachentsev & Makarov 1996). For 19 galaxies (DDO 50, IC 10, IC 1613, IC 342, M 33, M 81, M 82, Maffei 2, NGC 0185, NGC 0247, NGC 0300, NGC 1569, NGC 4236, NGC 4569, NGC 6503, NGC 6822, NGC 6946, VCC 1043 and VCC 92), we use the redshift-independent distances* from NED (Gavazzi et al. 1999; Olivares E. et al. 2010; Tully et al. 2013; Karachentsev et al. 2014). Finally for 3C 368, we use the SFR from Podigachoski et al. (2015), and for NGC 4038 and NGC 4039, we use the SFRs from Herrera-Camus et al. (2018a).

Stacey et al. (2010): We adopt magnification factors of two and eight for SMM J22471-0206 and Hbootes03, respectively. Due to the uncertain magnification in all other galaxies, we flag their luminosities as a precaution. For five galaxies (2XMM J094144.6+385440, 3C 065, 3C 368, IRAS F10026+4949 and IRAS F22231-0512), we adopt the SFRs from other works (De Looze et al. 2014; Podigachoski et al. 2015; Khan-Ali et al. 2015; Vayner et al. 2021).

* These distances come from methods that use standard candles or rules, such as Cepheids and globular clusters.

Table 3.4 – *References and number of measurements per line with a redshift range for the observational sample, after AGN dominated galaxies have been removed.*

Reference	Line measurements								Redshift	
	[O III] 52 μm	[N III] 57 μm	[O I] 63 μm	[O III] 88 μm	[N II] 122 μm	[O I] 145 μm	[C II] 158 μm	[N II] 205 μm	Min.	Max.
Brauher et al. (2008)	45	41	158	87	100	47	223	···	0.0	1.13
Stacey et al. (2010)	···	···	···	···	···	···	9	···	1.12	1.96
Valtchanov et al. (2011)	1	1	1	1	1	1	1	···	3.04	3.04
Decarli et al. (2012)	···	···	···	···	···	···	···	1	3.94	3.94
Coppin et al. (2012)	···	···	6	···	···	···	···	···	1.1	1.62
Farrah et al. (2013)	22	23	23	···	23	23	23	···	0.04	0.26
De Looze et al. (2014)	···	···	10	4	···	···	19	···	1.1	6.6
Magdis et al. (2014)	···	···	···	···	···	8	15	12	0.22	2.96
Brisbin et al. (2015)	···	···	7	···	···	···	8	···	1.41	2.0
Spinoglio et al. (2015)	8	14	26	26	23	23	26	12	0.0	0.3
Rosenberg et al. (2015)	···	···	14	···	···	25	29	···	0.0	0.04
Ferkinhoff et al. (2015)	···	···	···	1	1	···	···	···	2.81	2.81
Gullberg et al. (2015)	···	···	···	···	···	···	20	···	2.12	5.7
Capak et al. (2015)	···	···	···	···	···	···	11	···	5.15	5.69
Béthermin et al. (2016)	···	···	···	···	···	···	1	1	4.77	4.77
Cigan et al. (2016)	···	···	5	4	5	···	5	···	0.0	0.0
Zhao et al. (2016a)	···	···	···	···	···	···	···	196	0.0	0.06
Zhao et al. (2016b)	···	···	7	···	···	···	8	···	0.11	0.2
Oteo et al. (2016)	···	···	···	···	···	···	2	···	4.42	4.43
Kamenetzky et al. (2016)	···	···	···	···	···	···	···	214	0.0	0.25
Pavesi et al. (2016)	···	···	···	···	···	···	3	3	5.29	5.65
Uzgil et al. (2016)	···	···	···	···	···	···	1	···	2.29	2.29
Fernández-Ontiveros et al. (2016)	28	50	141	115	98	81	178	186	0.0	0.44
Knudsen et al. (2016)	···	···	···	···	···	···	1	···	6.03	6.03
Wardlow et al. (2017)	8	4	5	···	···	···	···	···	1.03	3.27
Díaz-Santos et al. (2017)	···	···	226	148	71	···	229	···	0.0	0.09
Olsen et al. (2017)	···	···	···	···	···	···	34	···	5.15	7.6
Decarli et al. (2017)	···	···	···	···	···	···	4	···	6.07	6.59
Lagache et al. (2018)	···	···	···	···	···	···	55	···	4.22	7.91
Vishwas et al. (2018)	···	···	···	1	···	···	···	···	3.13	3.13
Gullberg et al. (2018)	···	···	···	···	···	···	3	···	4.42	4.44
Herrera-Camus et al. (2018a)	···	47	55	46	56	50	56	···	0.0	0.13

Table 3.4 – *continued.*

Reference	Line measurements								Redshift	
	[O III] 52 μm	[N III] 57 μm	[O I] 63 μm	[O III] 88 μm	[N II] 122 μm	[O I] 145 μm	[C II] 158 μm	[N II] 205 μm	Min.	Max.
Lamarche et al. (2018)	1	1	1	1	1	···	2	···	1.78	1.78
Walter et al. (2018)	···	···	···	1	···	···	1	···	6.08	6.08
Rigopoulou et al. (2018)	1	1	1	1	1	1	1	···	2.96	2.96
Zhang et al. (2018)	9	···	25	35	38	35	33	13	1.03	3.63
Zanella et al. (2018)	···	···	···	···	···	···	9	···	1.73	1.94
Cormier et al. (2019)	···	3	30	36	15	12	41	3	0.0	0.04
De Breuck et al. (2019)	···	···	···	1	1	1	1	1	4.22	4.22
Tadaki et al. (2019)	···	···	···	1	···	···	1	1	4.34	4.34
Neeleman et al. (2019)	···	···	···	···	···	···	7	···	6.03	6.59
Lee et al. (2019)	···	···	···	···	1	···	1	1	4.69	4.69
Harrington et al. (2019)	···	···	···	···	···	···	···	1	2.55	2.55
Béthermin et al. (2020)	···	···	···	···	···	···	117	···	4.41	5.87
Rybak et al. (2020)	···	···	1	···	···	···	···	···	6.03	6.03
Harikane et al. (2020)	···	···	···	1	1	···	1	···	6.03	6.03
Cheng et al. (2020)	···	···	···	···	···	···	1	1	4.24	4.24
Cunningham et al. (2020)	···	···	···	···	···	···	27	37	3.07	5.81
Neeleman et al. (2020)	···	···	···	···	···	···	1	···	4.26	4.26
Rizzo et al. (2020)	···	···	···	···	···	···	1	···	4.22	4.22
Fraternali et al. (2021)	···	···	···	···	···	···	2	···	4.54	4.57
Mitsuhashi et al. (2021)	···	···	···	···	···	···	5	···	4.62	4.64
Rybak et al. (2021)	···	···	···	···	···	···	1	···	1.84	1.84
Fujimoto et al. (2021)	···	···	···	···	···	···	2	···	6.07	6.07
Lee et al. (2021)	···	···	···	···	···	1	···	···	4.7	4.7
Rizzo et al. (2021)	···	···	···	···	···	···	5	···	4.23	4.77

Decarli et al. (2012): We use updated values for the magnification factors: for QSO J0831+5245 we adopt the upper limit of eight (Saturni et al. 2016), and for [LWB2009] MM184222+593828 we adopt a value of 12 (Lestrade et al. 2011).

Farrah et al. (2013): We adopt the SFR value from Kamenetzky et al. (2016) for IRAS 00397-1312.

De Looze et al. (2014): We use the magnification factors of other works in cases where the lensing models are better, or for consistency with other references used in this work, for the following galaxies: SDP 81 (Valtchanov et al. 2011), QSO J0831+5245 (Saturni et al. 2016) and [CRR2012] HLS J091828.6+514223 (Lagache et al. 2018). In addition, we take the magnification factor and SFR for G15.v2.779, HFLS3 and SMM J22471-0206 from Cheng et al. (2020), Cooray et al. (2014) and Ferkinhoff et al. (2015), respectively. For the following galaxies, we take SFR values from the literature: 2XMM J094144.6+385440 (Khan-Ali et al. 2015), 3C 065 (Podigachoski et al. 2015), 3C 368 (Podigachoski et al. 2015), 3C 446 (Vayner et al. 2021), IRAS F22231-0512 (Vayner et al. 2021), SMM J02399-0136 (Ferkinhoff et al. 2015), IRAS F10026+4949 (Ruiz et al. 2013) and SWIRE J104704.97+592332.3 (Stacey et al. 2010).

Magdis et al. (2014): For HXMM01, we adopt the magnification factor from Wardlow et al. (2017). For HLock01, we adopt the magnification factor from Gavazzi et al. (2011).

Brisbin et al. (2015): We assume that the cosmology used is the one from Spergel et al. (2003), as they mention that their work is a continuation of Stacey et al. (2010). We assume that all galaxies in this sample are magnified (Zanella et al. 2018).

Spinoglio et al. (2015): L_{IR} is calculated with the IRAS 60 and 100 μm fluxes as described in Brauher et al. (2008).

Rosenberg et al. (2015): We use the redshift from NED and the infrared flux reported in their tables to estimate L_{IR}.

Ferkinhoff et al. (2015): Data for QSO J0831+5245 comes from Ferkinhoff et al. (2010), where we have adopted the magnification factor from Saturni et al. (2016). Although this galaxy is part of the sample, we do not use it for the comparison in this work because the SFR estimated from the L_{IR} is unreliable due to the contribution of the AGN. Reported luminosities for [N II] come from ZEUS for QSO J0831+5245 and SMM J02399-0136, and ALMA for Cloverleaf (QSO J1415+1129).

Gullberg et al. (2015): We take the magnification values from Spilker et al. (2016); we flag those as limits when there is no magnification factor available (e.g. for SPT-S J051258-5935.6). For SPT-S J055138-5057.9 we adopt the magnification factor from Cunningham et al. (2020), while for SPT-S J053816-5030.8 we adopt the magnification factor (18.8) reported by Spilker et al. (2016).

Cigan et al. (2016): L_{IR} is calculated using the IRAS 60 and 100 μm fluxes as described in Brauher et al. (2008). For DDO 155 we use the estimated SFR from Hα by Pokhrel et al. (2020).

Zhao et al. (2016a): We obtain redshifts from NED when available. For five galaxies (M 82, NGC 1569, NGC 2976, NGC 3077 and NGC 4569) we use the redshift-independent distances from NED (Tully et al. 2013; Karachentsev et al. 2014). For NGC 3557 we use the preferred redshift distance. Most of the galaxies in this sample do not have an estimated SFR, so we use the SFRs from Brauher et al. (2008) and Díaz-Santos et al. (2017) for most of the missing estimates. We use SFRs from Kamenetzky et al. (2016) for UGC 2369 and NGC 5010, and Tateuchi et al. (2015) for IC 4518A.

Zhao et al. (2016b): Some of the galaxies are identified to have second velocity components in their spectra; therefore we use the median luminosity value in those cases.

Oteo et al. (2016): The pair of galaxies in this reference is treated as one galaxy (H-ATLAS J000307.2-330250), as the second component does not exist in SIMBAD. We use the IR luminosities derived from SED fitting from Stacey et al. (2021).

Kamenetzky et al. (2016): We assume that the additional regions in NGC 6946 and NGC 4038 (Antennae) are also part of the galaxy. We use SFRs from other references for the following galaxies: 3C 405 (Brauher et al. 2008); Mrk 1298 (Shi et al. 2014); IC 4518A (Tateuchi et al. 2015); 3C 315 and 3C 433 (Westhues et al. 2016); ESO 255-IG007, NGC 2976 and VV 705 (Zhao et al. 2016a); 3C 305 (Fernández-Ontiveros et al. 2016); 3C 31 (Vaddi et al. 2016); IRAS 08355-4944, IRAS F01417+1651, MCG-03-34-064, NGC 877 and VV340a (Díaz-Santos et al. 2017); and, Antennae and NGC 4151 (Herrera-Camus et al. 2018a).

Uzgil et al. (2016): We assume a magnification factor of 30 for IRAS F10214+4724. When correcting the L_{IR} for this magnification factor, the SFR derived from the Kennicutt & Evans (2012) relation is similar to the SFR estimated by Evans et al. (2006). We use the magnification factor of eight as an upper limit for QSO J0831+5245 (Saturni et al. 2016). We assume this work uses the cosmology of Komatsu et al. (2009).

Fernández-Ontiveros et al. (2016): L_{IR} is calculated with the IRAS 60 and 100 μm fluxes as described in Brauher et al. (2008). For LEDA 101275 and Mrk 463E we use the IR luminosity reported by Ashby et al. (2011). For eight galaxies (M 82, NGC 1569, NGC 2976, NGC 3077 and NGC 4569) we use redshift-independent distances from NED (Tully et al. 2013). For NGC 4569 and NGC 6946 we adopt distances reported by Cortés et al. (2008) and Rodríguez et al. (2014), respectively. We use SFRs from other references for the following galaxies: 3C 405 (Brauher et al. 2008); IC 4518A (Tateuchi et al. 2015); 2E 4728, 3C 317 and MCG+05-33-005 (Fraser-McKelvie et al. 2014); 3C 33, 3C 234, 3C 315 and 3C 433 (Westhues et al. 2016); LEDA 3098117, LEDA 4666674, NGC 7592W and Z 468-2 (Díaz-Santos et al. 2017); Centaurus A and Circinus galaxy (Herrera-Camus et al. 2018a); Mrk 266B Smith et al. (2019); and, ESO 141-55, 2MASX J00535615-7038045, IRAS 03450+0055 and UGC 12138 Ramos Padilla et al. (2022).

Knudsen et al. (2016): We use only the information of [ZFM2015] A383 5.1, as observations of the other galaxy (MS0451-H) are only upper limits and are not identified as a galaxy in SIMBAD.

Díaz-Santos et al. (2017): Redshifts are taken from NED to convert from the WMAP cosmology (Komatsu et al. 2009) used for the sample in Armus et al. (2009) to the Planck cosmology used in this work. We obtain IR luminosities from the sample webpage*.

Olsen et al. (2017): We remove two galaxies that are not identified in SIMBAD (MS0451-H and [CFP2010] BDF 3299 sub-region).

Lagache et al. (2018): For some galaxies we use a different magnification factor to those reported in Lagache et al. (2018). For H-ATLAS J142413.9+022304 and SPT-S J034510-4725.6 we use the magnification factor used by Cheng et al. (2020), and for [CLM2003] J022802.97-041618.3 we use the values by Olsen et al. (2017). We use the IR luminosity value for AzTEC 3 from Stacey et al. (2021). For galaxies coming from Aravena et al. (2016) we use the IR luminosities of Faisst et al. (2017). We use the reported SFR instead of converting from the IR luminosities for [CLM2003] J022802.97-041618.3 and [WMH2013] 5. Finally, we use the SFR from Neeleman et al. (2020) for ALMA J081740.86+135138.2.

Vishwas et al. (2018): We use a magnification factor of seven so the dust disk of H-ATLAS J113526.3-014605 can be consistent with other high-z SMG.

Gullberg et al. (2018): The data presented with this reference is an improvement of the results presented in Swinbank et al. (2012).

* GOALS sample at `https://goals.ipac.caltech.edu`

Lamarche et al. (2018): We adopt the IR luminosity from Bussmann et al. (2013). Two velocity components are measured in the [C II] line of SPD 11. We adopt the magnification factor estimated for the dominant (red) component.

Rigopoulou et al. (2018): We adopt a magnification factor of 10.9 according to Gavazzi et al. (2011).

Zhang et al. (2018): We use the magnification factor reported in their Table 4. For H-ATLAS J084933.4+021443 and SDP 81, we adopt 2.8 and 25, respectively, as magnification factors (Wardlow et al. 2017; Valtchanov et al. 2011).

Cormier et al. (2019): We use distances reported by Rémy-Ruyer et al. (2013) in most of the galaxies. For ten galaxies (ESO 495-21, Mrk 209, NGC 1140, NGC 1569, NGC 1705, NGC 4214, NGC 5253, NGC 625, UGC 4483 and UGC 6456) we use redshift-independent distances from NED (Tully et al. 2013). For ESO 350-38 and HS 0822+3542 we use the preferred redshift values from NED. For UGCA 116 we use the distance reported in Bordalo et al. (2009). We use the reported distance used in Cormier et al. (2019) for Mrk 33 and discard the information from the LMC and SMC. For 2MASX J12390403+3920437 we use the SFR reported by Duarte Puertas et al. (2017).

De Breuck et al. (2019): We use the magnification factor of 32.7 by Spilker et al. (2016).

Tadaki et al. (2019): We assume the cosmology used is a ΛCDM cosmology with $\Omega = 0.27$ and $H_0 = 70$ km s^{-1} Mpc^{-1}.

Neeleman et al. (2019): For [NBW2019] J0842+1218C2 we assume that the upper limit for SFR is 100 M$_\odot$ yr^{-1}.

Lee et al. (2019): We use the IR luminosity values and [C II] measurements from Iono et al. (2006).

Hashimoto et al. (2019): The reported values for the SFR are assumed to be upper limits, as those come from QSOs.

Béthermin et al. (2020): These objects have not been ingested in SIMBAD, but we use the coordinates of the closest object. However, this can lead to incorrect identifications. For example, the closest object to the coordinates of VUDS-COSMOS-5100541407 is a star (COSMOS 877137). Caution is therefore required. We use SFRs from SED fitting except for cases where it seems to be overestimated. For COSMOS-DEIMOS-873756 and VUDS-COSMOS-510596653 we assume the SFRs are upper

limits, where for VUDS-COSMOS-510596653 we use the estimated IR luminosity and convert it to SFR.

Venemans et al. (2020): We do not include the companions of the galaxies, as those are not yet identified in SIMBAD.

Cunningham et al. (2020): We use the SFRs from Gullberg et al. (2015) for most of the galaxies. We obtain IR luminosities for SPT-S J234942-5638.2, SPT-S J235339-5010.1 and SPT-S J235718-5153.7 from Greve et al. (2012). For SPT-S J020258-6121.2, SPT-S J045859-5805.1, SPT-S J045912-5942.4 and SPT-S J204823-5520.7 we obtain IR luminosities and magnification factors from Spilker et al. (2020). For SPT-S J231124-5450.5 we use the magnification factor from Spilker et al. (2016). For SPT-S J235149-5722.2 we use the IR luminosity from Faisst et al. (2017) as an upper limit. Finally, we discard the galaxies without coordinates information in SIMBAD.

Rizzo et al. (2020): We adopt the values that take into account the magnification of 32.3.

Mitsuhashi et al. (2021): We adopt the values that take into account the magnification. These galaxies are not yet included in SIMBAD.

Fujimoto et al. (2021): We adopt the values that take into account the magnification (their Tables 3 and 4). These galaxies are not yet included in SIMBAD.

Rizzo et al. (2021): We adopt the values that take into account the magnification. For five galaxies (SPT-S J011308-4617.7, SPT-S J034510-4725.7, SPT-S J044143-4605.5, SPT-S J213244-5803.1 and SPT-S J214654-5507.9), we adopt the magnification factors from Cunningham et al. (2020).

Dataset of estimated emission line luminosities

Predictions of the eight emission lines discussed in this work (Table 3.1) are available at the CDS and in a Zenodo repository at `https://doi.org/10.5281/zenodo.6576202`. The dataset contains the total line luminosities as well as the contributions of the different ISM phases. We show ten rows from the dataset containing these luminosities in Table 3.5 as an example of the format and content. We also present a table of the physical parameters involved in the predicted line luminosities. We show ten rows from the dataset containing the physical parameters in Table 3.6 as an example of the format and content. The first columns of these tables matches the Group Number in the EAGLE database (McAlpine et al. 2016), which is the unique identifier of the FoF (Friends-of-Friends) halo of a given galaxy (unique per snapshot). Although we have applied our physical model on the sample of 8 227 galaxies simulated with EAGLE, the dataset contains 8 224 galaxies since one of the galaxies

does not have enough gas for the estimates, and the other two galaxies fail to predict luminosities in a reasonable amount of computational time.

Linear regressions for lines

Assuming that FIR luminosities change with SFR and redshift (z), we fit the estimated data to obtain linear relations in terms of these parameters for each FIR emission line. At any z there is a linear relation of the form

$$\begin{aligned} \log(\mathrm{L_{line}}) &= c_0 + c_1 \log(\mathrm{SFR}) + c_2 \log(1+z) + \\ &\quad c_3 \log(\mathrm{SFR}) \log(1+z), \end{aligned} \tag{3.2}$$

with SFR in units of $\mathrm{M_\odot\ yr^{-1}}$ and line luminosities in $\mathrm{L_\odot}$. The values for the coefficients presented in Table 3.7 are obtained from the combination of RECAL-L0025N0752 and REF-L100N1504 mock data. We also estimate the coefficients of the [O III] and [N III] line relations using only galaxies where HII regions are dominant (the contribution is higher than 50%).

Table 3.5 – *Example of dataset with the line luminosity estimates derived from this work.*

Group Number	Sim.	z	$L_{NIII_{57}}$ dex($L_\odot$)	$L_{NIII_{57},DIG}$ dex($L_\odot$)	$L_{NIII_{57},HII}$ dex($L_\odot$)	$L_{NIII_{57},ato}$ dex($L_\odot$)	$L_{NIII_{57},mol}$ dex($L_\odot$)	$L_{OIII_{88}}$ dex($L_\odot$)	$L_{OIII_{88},DIG}$ dex($L_\odot$)	$L_{OIII_{88},HII}$ dex($L_\odot$)	$L_{OIII_{88},ato}$ dex($L_\odot$)	$L_{OIII_{88},mol}$ dex($L_\odot$)
688	Recal	0	2.93	2.93	···	···	−4.93	3.51	3.51	···	···	···
11	Recal	1	6.33	5.52	6.26	−0.42	−3.19	6.89	6.17	6.80	0.06	−3.56
1	Recal	5	6.48	4.55	6.48	0.60	···	7.20	5.23	7.20	1.06	···
2442	Ref100	0	6.44	4.21	6.44	−0.67	−3.13	6.90	4.83	6.89	−0.22	−3.35
4634	Ref100	1	5.30	4.97	5.03	−0.34	−2.56	5.90	5.60	5.59	0.18	−2.86
6207	Ref100	2	6.40	4.82	6.38	0.70	−4.07	7.10	5.46	7.09	1.26	−4.07
5662	Ref100	3	6.69	4.49	6.69	0.39	−4.41	7.39	5.17	7.39	0.96	−4.70
4194	Ref100	4	6.41	4.18	6.40	0.33	···	7.08	4.91	7.08	0.85	···
229	Ref100	4	5.93	5.83	5.26	1.18	−2.95	6.55	6.50	5.60	1.75	−3.31
332	Ref100	5	6.61	4.14	6.60	0.85	···	7.28	4.86	7.28	1.37	−4.70

Table 3.6 – *Example of physical parameters dataset derived from this work.*

Group Number	Sim.	z	M_* dex($M_\odot$)	M_{gas} dex($M_\odot$)	$M_{neutral}$ dex($M_\odot$)	Age_* dex(Gyr)	log(Q) dex(1 / s)	ISRF Habing	SFR dex($M_\odot$ / yr)	$Z/Z_\odot$	Pressure dex(K / cm^3)	$n(H)_{cloud}$ dex(1 / cm^3)
688	Recal	0	8.154	7.208	6.930	0.852	34.769	0.530	−2.511	0.356	4.151	0.594
11	Recal	1	10.186	9.665	9.273	0.260	35.761	0.804	−0.409	−0.344	7.816	0.854
1	Recal	5	8.732	10.103	9.947	−0.656	44.669	1.496	0.269	−1.586	8.777	1.473
2442	Ref100	0	9.847	9.735	9.625	0.784	38.941	0.687	−0.319	−0.073	5.975	0.406
4634	Ref100	1	9.458	9.625	9.271	0.296	40.203	2.253	0.123	−0.127	7.756	1.979
6207	Ref100	2	9.151	9.846	9.410	−0.098	41.948	1.828	0.087	−0.562	7.196	1.361
5662	Ref100	3	9.057	9.854	9.566	−0.161	42.137	1.925	0.271	−0.757	7.534	1.560
4194	Ref100	4	8.682	9.814	9.449	−0.430	44.019	1.898	0.064	−1.385	7.391	1.472
229	Ref100	4	10.371	10.088	9.754	−0.346	40.174	1.466	−0.067	−1.629	7.803	1.348
332	Ref100	5	8.735	10.115	9.926	−0.690	45.148	1.757	0.420	−1.413	7.981	1.563

Table 3.7 – *Linear relations derived from this work for each of the FIR emission lines using Eq. 3.2*

	Coefficients				
$\log(L_{line})$	c_0	c_1	c_2	c_3	1σ
[O III] 52 μm	6.72±0.02	1.54±0.02	−0.15±0.03	−0.41±0.03	0.63
[N III] 57 μm	6.38±0.02	1.54±0.02	−0.39±0.03	−0.35±0.03	0.61
[O I] 63 μm	6.13±0.01	1.03±0.01	0.70±0.02	0.09±0.02	0.38
[O III] 88 μm	6.84±0.02	1.46±0.02	−0.07±0.03	−0.35±0.03	0.60
[N II] 122 μm	5.59±0.01	1.13±0.01	−0.68±0.02	0.06±0.02	0.46
[O I] 145 μm	4.96±0.01	1.10±0.01	0.69±0.02	−0.15±0.02	0.41
[C II] 158 μm	6.55±0.01	0.72±0.01	0.14±0.01	0.24±0.01	0.24
[N II] 205 μm	5.48±0.01	0.93±0.01	−0.82±0.02	0.24±0.02	0.42
Coefficients for HII regions					
[O III] 52 μm	6.99±0.02	1.11±0.02	−0.57±0.03	0.25±0.04	0.53
[N III] 57 μm	6.65±0.02	1.15±0.02	−0.82±0.03	0.26±0.04	0.51
[O III] 88 μm	7.11±0.02	1.06±0.02	−0.48±0.03	0.26±0.04	0.50

The future's in the air
Can feel it everywhere
Blowing with the wind of change
Scorpions - Wind of Change

Someday I will make it
Andrés pedalling back home

4

Diagnosing the interstellar medium of galaxies with far-infrared emission lines: Physical parameters of observed galaxies

A. F. Ramos Padilla, L. Wang, F. F. S. van der Tak and S. C. Trager

Available in this thesis

Highlights

- Some available tools used to derive physical parameters are not easy enough to use for simple calculations.
- We present a web app called DiagISM that estimate some physical parameters from FIR line luminosities using a user-friendly environment.
- The models implemented inside DiagISM correctly predict SFR from observed FIR line luminosities.
- We present some number counts from a LETO-like space telescope for future observations.

Abstract

Context: The far-infrared (FIR) emission lines encode valuable information regarding the physical properties of the interstellar medium (ISM). Simple user-friendly tools are needed to allow observers to easily derive such information.
Aims: We present a web app to estimate physical parameters from FIR emission lines observations using a user-friendly environment, called DiagISM. In addition, we estimate the expected number counts of galaxies for planned future FIR space telescope missions.
Methods: We implement an artificial neural network (ANN) to predict parameters related to galaxy physical properties, using data derived from the predicted FIR line luminosities from simulated EAGLE galaxies. We apply the trained ANN to observed galaxies and compare the predicted physical parameters to literature values. Finally, we combine this information with the SIDES lightcone catalogue to predict number counts as a function of various line luminosities at different cosmic epochs.
Results: Models, with eight or at least two FIR emission lines, can successfully retrieve the neutral cloud size (R_{cloud}), star-formation rate (SFR), interstellar radiation field (ISRF), metallicity and gas mass with less than 0.5 dex error compared to input values from the simulations. When applying one of these models to an observational sample of 490 galaxies with $[O\,III]_{88}$ and [C II] FIR line luminosities, the models recover the SFR with a coefficient of determination (R^2) of $R^2 = 0.86$ and a median absolute error of 0.29 dex. We estimated the other physical parameters for that sample of galaxies and made the dataset available at https://doi.org/10.5281/zenodo.6705031. With the DiagISM predictions, we show how the number of galaxies as a function of various FIR line luminosities depend on the observational sensitivity and redshift.
Conclusions: The models in DiagISM show us that observing FIR lines like [C II], [O III] and [O I] at 63 μm will be ideal for obtaining a statistical significant sample of galaxies for future observations from which we can estimate reliable physical parameters.

Keywords: Galaxies: evolution, star formation, ISM, high-redshift – ISM: lines and bands, structure – Infrared: ISM – methods: numerical

4.1 Introduction

The interstellar medium (ISM) plays a crucial role in understanding the physical processes that operate in different environments in galaxies throughout cosmic time. The gas present in the ISM contains information about previous generations of stars (such as metallicities), current conditions of the gas (such as density, temperature and radiation fields) and future generations of stars (such as star-formation rates (SFR) and gas mass) (e.g. Tielens 2010). In general, depending on the conditions of the gas, the ISM can be separated into different phases from molecular clouds (with high densities $> 200\,cm^{-3}$ and very low temperatures $\sim$10 K) to the hot ionised medium (HIM, with low densities $\sim 3 \times 10^{-3}\,cm^{-3}$ and high temperatures $\sim 10^6$ K). In between these ISM phases, we can find other phases such as the cold neutral medium (CNM), warm neutral medium (WNM), warm ionised medium (WIM), and HII regions (e.g. Ryden & Pogge 2021). Each of these phases can be traced by emission or absorption lines in different ranges of the spectrum. These lines contain the physical information necessary to constrain the conditions present in galaxies throughout the Universe.

In the last decades, many spectral line observations have been collected to diagnose the ISM. Some tools have been developed focused mainly on the interface between neutral clouds and ionisation fields, also known as photo-dissociation regions (PDR, e.g. Hollenbach & Tielens 1999; Wolfire et al. 2022), to obtain this information in terms of physical parameters. Some examples of such codes are: PDRtoolbox (Kaufman et al. 2006; Pound & Wolfire 2008, 2011), KOSMA-τ (Röllig & Ossenkopf-Okada 2022, and references therein), Meudon (e.g. Le Petit et al. 2006; Goicoechea & Le Bourlot 2007; Gonzalez Garcia et al. 2008) and Cloudy (Ferland et al. 2013, 2017), which have been compared with other models by Röllig et al. (2007). Some of these PDR codes allow us to obtain physical parameters such as cloud density and interstellar radiation field (ISRF) directly from the spectral line intensities. However, some of these models are not very user-friendly and/or requires a lot of computational time.

One way to significantly decrease the computational time to derive physical parameters from the line emissions is to use machine learning techniques. These techniques are standard tools for data analysis in astrophysics and other sciences (Carleo et al. 2019). One such technique is artificial neural networks (ANN Bishop et al. 1995), which has been applied in the context of astrophysical regression and classification problems. For example, photometric redshifts (e.g. Collister & Lahav 2004; Vanzella et al. 2004), galaxy morphologies (e.g. Domínguez Sánchez et al. 2018), interacting galaxies (e.g. Pearson et al. 2019a) and even physical parameters such as the IR luminosity (Ellison et al. 2016) can be retrieved with trained ANN.

To correctly constrain some of the parameters at different redshifts, specific FIR lines are required. The most commonly used FIR line used is the [C II] emission line at 158 μm. This line can act as a thermostat as it comes from different phases of the ISM (Gong et al. 2012; Goldsmith et al. 2012; Olsen et al. 2015; Ramos Padilla et al. 2021) and its luminosity is around 1% of the FIR luminosity of galaxies (e.g. Stacey et al. 1991; Brauher et al. 2008). Thanks to previous FIR missions such as *ISO* (Kessler et al. 1996) and *Herschel* (Pilbratt et al. 2010) it was possible not only to observe this line but also other important FIR lines, such as [O III] at 88 μm and [N II] at 122 and 205 μm. Nowadays, ALMA can be used to observe some of these FIR lines at high redshifts ($z > 3$, e.g. Mordini et al. 2022) and some tools have been created specifically to understand the physical properties of galaxies at these redshifts (Vallini et al. 2020, 2021). However, ideally, we would like to cover the entire FIR wavelength range at all redshift ranges from $z = 0$ to $z = 10$. For this, a dedicated FIR space telescope is needed to obtain all the important emission lines at different cosmic epochs to bridge the gap between near-infrared and sub-millimetre telescopes. Some concepts are being considered at the moment by some space agencies (e.g. by NASA and ESA), but still nothing concrete in the near future.

With this in mind, the goal of this paper is to estimate the physical parameters from FIR line luminosity observations using a user-friendly environment and estimate the expected counts of galaxies for future FIR space telescope missions. Throughout this paper, we assume the ΛCDM cosmology from the Planck Collaboration et al. (2014b) results ($\Omega = 0.307$, Ω_Λ=0.693, $H_0 = 67.7$ km s^{-1} Mpc^{-1} and $\sigma_8 = 0.8288$).

In this paper, we first describe the ISM model and the simulated FIR emission line in Ramos Padilla et al. (2022, hereafter Paper II), which we use to train machine learn-

ing methods to learn the mapping between FIR lines and galaxy physical properties (Sect. 4.2). Next, we test models with the simulated data (Sect. 4.4) and present the results of the physical parameters we obtain from the observational data (Sect. 4.5). In Sect. 4.6 we show the predictions for future FIR space telescope missions. After that, we discuss the limitations of the predictions presented in this work (Sect. 3.5). Finally, we present our conclusions in Sect. 4.8.

4.2 Data

In this section, we first describe the simulation data used to train the machine learning (ML) models to predict physical properties based on the FIR line emissions. Then, we present the observational sample which we use to compare our model predictions with.

To estimate physical parameters from FIR observations, we made use of the data presented in Paper II. In that work, we implemented a physically motivated model of the ISM to predict the luminosity of the most important FIR emission lines, namely [C II], [O I], [O III], [N II] and [N III]. In addition, we collected the information from 1 702 measurements in galaxies with some of those FIR lines. In this section, we briefly describe the contents of the predicted luminosities from simulations as well as the collected data from observations.

4.2.1 Simulation data

In Ramos Padilla et al. (2021, hereafter Paper I) we present a model of the ISM for the [C II] line at 158 μm in the local Universe assuming three different phases of the ISM: dense molecular gas, neutral atomic gas and diffuse ionised gas (DIG). We implement the model by post-processing the EAGLE cosmological simulations (Schaye et al. 2015; Crain et al. 2015) with the help of Cloudy (Ferland et al. 2017) look-up tables (Ploeckinger & Schaye 2020). Then in Paper II, we expand the ISM model by adding HII regions as an ISM phase and estimated the luminosities of seven additional FIR lines, which are: [O I] at 63 and 145 μm, [O III] at 52 and 88 μm, [N II] at 122 and 205 μm, and [N III] at 57 μm. Furthermore, we use other EAGLE snapshots up to $z = 6$, which allow us to reach other redshift ranges beyond the local Universe. We use the closest EAGLE snapshots in the redshift range $z = 0$–6 in steps of one (i.e. $z = 0$, $z = 1$, $z = 2$, etc.). For each of the 8 224 simulated galaxies, we estimate the luminosity coming from the ISM phases to obtain the total luminosity for a given line (i.e. $L_{line} = L_{line}^{molecular} + L_{line}^{atomic} + L_{line}^{DIG} + L_{line}^{HII}$). Together with these line luminosities, other physical parameters were estimated (e.g. interstellar radiation field and neutral cloud density) or obtained from the global properties of the galaxies, such as star-formation rate (SFR) and metallicity. Information on these physical parameters and luminosities was compared qualitatively with observations using diagnostic diagrams such as [C II]/[O III] vs [N II]/[O I]. Here, we compare them quantitatively.

Table 4.1 – *Sets from observational samples defined to compare the predictions described in this work, ordered by the number of galaxies. The first column shows the adopted name for the set, followed by the number of galaxies in the set and the luminosities used in the second and third columns. The last column shows the maximum redshift in each of the sets.*

Set	# galaxies	Available lines	Max. z
A	699	[O I]$_{63}$, [C II]	3.127
B	490	[O III]$_{88}$, [C II]	6.081
C	144	[O I]$_{63}$, [C II], [N II]$_{205}$	2.170
D	80	[O III]$_{52}$, [N III]$_{57}$	3.043

4.2.2 Observational sample

We use the observational data compilation presented in Paper II which contains measurements of the FIR emission lines in the literature. Luminosities in this sample have been homogenised in terms of the same ΛCDM cosmology (Planck Collaboration et al. 2014b). For most of the galaxies in the sample, a SFR was calculated based on the corresponding IR luminosity according to Kennicutt & Evans (2012). Additionally, when gravitational lensing is present, the galaxies were flagged as a precaution. References and specific comments on the observational sample can be found in Appendix A of Paper II.

Luminosity measurements in the observational sample are scarce except for the [C II] emission line in the local Universe. To compare with the predictions from the ML model trained on the simulated FIR lines from Paper II (see Sect. 4.3), we implement in this work with observations, we decided to divide the observational sample into different sets of galaxies taking into account different line combinations, as detailed below.

For the first set of galaxies, we use the two most commonly available lines in the sample: [C II] and [O I] at 63 μm, with around 70% and 45% availability in the observational sample. This set, called A due to its high availability, contains 699 galaxies with a maximum redshift of $z = 3.1$. We define the second set as the one with the lines of [C II] and [O III] at 88 μm. We selected these lines as those are also the most common and easy to observe at high-z, therefore called B, again due to its availability. The third set, called set C, is defined using the three most common lines in the sample ([C II], [O I] at 63 μm and [N II] at 205 μm). The addition of the [N II] at 205 μm, which has an availability of around 40% in the observational sample, reduces the total number of galaxies observed with the three lines to 144. Finally, we define the last set with the two less frequently available FIR lines: [O III] at 52 μm and [N III] at 57 μm, with around 10% availability in the observational sample. This sample works as the most limited scenario, called set D, as it contains the lowest number of galaxies (80) and those FIR emission lines are more difficult to observe. Most galaxies in the observational sample sets are local galaxies. However, high-z galaxies are also present in the sample and behave similarly in the prediction models (see Sect. 4.5). A summary of the four sets described above can be found in Table 4.1. These sets are used to compare the models described in the following subsection.

4.3 Methodology - ML powered DiagISM

In order to present predictions on the sets of the observational sample, we need to select a model that allows us to predict the different physical parameters easily and efficiently. As we described in the Introduction, ML techniques have helped to obtain predictions in astrophysics and different physical sciences (e.g. Carleo et al. 2019). Among those techniques, simple ANN are accurate enough in supervised learning environments to predict related physical parameters. Therefore, we decided to use this technique to predict physical parameters in the observational data sets using the FIR line luminosities from Paper II for the training.

We implement a small-scale ANN which allows us to obtain predictions in a small amount of time with a non-linear approximation. We use a multi-layer perceptron (MLP, Gardner & Dorling 1998) which consists of a set of interconnected neurons from the input layer (line luminosities and redshift) to the output layer (physical parameters). These layers are connected and weighted in a set of hidden layers that transform the values with an activation function. This activation function defines the output of the neurons given the inputs of the previous layer. We use the MLP regressor implementation available in `scikit-learn` (Pedregosa et al. 2011), `MLPRegressor`.

4.3.1 Hyper-parameter tuning

In order to optimise the retrieved physical parameters from the MLP, we explore which sets of hyper-parameters (parameters that control the learning process) are the best for our predictions. Since we are dealing with different physical parameters, we do this optimisation for each of them, even if they do not directly correlate with the FIR emission-line luminosities (except SFR). We use all the data from simulated galaxies to train these models (i.e. total estimated luminosities and the selected physical parameters).

We use ATOM*, a Python package that allows us to explore and generate some insights on the hyper-parameters, in addition to identifying and removing outliers. We use the `ATOMRegressor` class to obtain different values for the hyper-parameters for the MLP (e.g. activation function, configuration of the hidden layers, maximum number of iterations). For this, we assume the median absolute error (MAE) and the coefficient of determination (R^2) as our metrics of optimisation. The MAE is a robust measure for variability and R^2 measures the quality of replication of the model. We use a stochastic gradient-based optimiser for 100 calls (exploration, global search) and 50 initial points on the grid (exploitation, local search). Those number of calls and initial points are good enough for the purposes of this work (see Sects. 4.5 and 4.7.1). To assess the robustness of the model, we create five random batches† of the sample at each iteration time (i.e. resample the data with replacement), also known as bootstrapping. With those random samples the metrics will statistically decide which hyper-parameters minimise the MAE and maximise the R^2. With this configuration, we obtain the best hyper-parameters used to predict the physical parameters

* https://github.com/tvdboom/ATOM

† There is not rule to select this value, but a 5-fold selection is very common in the field of ML (James et al. 2013).

depending on the set of luminosities we are working with.

4.3.2 Using the eight main FIR lines ("M8all")

The first model we use is set up to take into account the redshift and the eight most important FIR emission lines, referred to as M8all. This model assumes in the training phase the availability of all these lines in order to predict the global physical parameters of the galaxies. However, we know that in most cases it is not possible to have all the eight line luminosities for a given galaxy. Therefore, we create mock luminosities in case any of the eight luminosities is missing.

The mock luminosities are created by taking into account a fixed redshift and assuming a typical error of $\sigma = 0.01$ for each given luminosity (input, in $\log(L_\odot)$ units). We select simulated galaxies that are within a given σ value of the input luminosities to define the distribution of the other non-given (i.e. missing) luminosities. In case there are no simulated galaxies with line emissions near the input values for the input lines, we assume the missing luminosities are similar to the average of the input luminosities. Although this is not ideal, it helps the model to find a close solution for the prediction. We create a total of 2 000 mock luminosities for all the lines so we can estimate the uncertainty of the predictions.

4.3.3 Using selected FIR lines ("M2plus")

The second model we use is set up to take into account any set of lines for a given physical parameter. This scenario is more realistic, as for most of the observed galaxies we only have two FIR lines observed. Therefore, this model only requires as input the redshift and minimum two line luminosities. We refer to this model as M2plus. These inputs are the ones used for training the model depending on the physical parameter that we want to retrieve. Then, these inputs define the model assuming the other lines which are not given (i.e. missing) do not affect the model. Again, we create a total of 2 000 mock luminosities so that we can obtain and estimate the uncertainty of the predictions, this time assuming a random normal distribution with $\sigma = 0.2$ (see Sect. 4.4.1).

4.3.4 Web app

The two models described in the previous sections are implemented and available in a web app interface*, which we call DiagISM. In DiagISM, it is possible to interact with the information presented in the two models (M8all and M2plus) described above, so observers can retrieve physical parameters from FIR line luminosities. The physical parameters that are possible to predict in DiagISM are similar to those used in Paper II: star-formation rate (SFR), interstellar radiation field (ISRF), metallicity ($Z/Z_\odot$), external pressure (P_{ext}), total hydrogen number density in the neutral clouds ($n(H)_{cloud}$), radius of the neutral clouds (R_{cloud}), gas mass (M_{gas}) and stellar mass ($M_\star$). In comparison to Paper II, we do not use the specific star formation rate ($sSFR = SFR/M_\star$) and we use SFR and $M_\star$ as separate parameters for clarity. These physical parameters can be selected manually inside DiagISM.

* https://aframosp-diagism.streamlitapp.com/

In Fig. 4.1 we show an example of the DiagISM web app for the M8all model. For a hypothetical galaxy, we give inputs values of $z = 2$, $\log(\mathrm{C\,II}[\mathrm{L}_\odot]) = 8.00$ and $\log(\mathrm{O\,III}_{88}[\mathrm{L}_\odot]) = 8.00$ to estimate the SFR. The training prediction score is given with this configuration ($\mathrm{R}^2 = 0.969$) and a normalised histogram of the mock luminosities is retrieved. DiagISM also returns the mean/median predicted value and its standard deviation/16th-86th percentiles of the hypothetical galaxy.

Inside the web app, it is also possible to include a comma-separated value (CSV) file in order to estimate more than one set of luminosity values, which is useful for a sample of galaxies. After the results are obtained, the total time it takes to run the configuration of FIR luminosities is returned and it is possible to download the results as a CSV file.

4.4 Performance of the MLP models

We create sets of estimated luminosities to test the models in DiagISM and estimate a realistic error bar for the distributions of the mock luminosities. These sets are similar to those presented in Table 4.1 but come from the simulations instead of the observations. Therefore, these simulated sets cover the redshift range of $z = 0$–6 and all of them contain 8 224 galaxies.

4.4.1 Dependence on the assumed error

We use the simulation sets to see how the median absolute error (MAE) and the coefficient of determination (R^2) change for a given global assumed error inside DiagISM (i.e. the error assumed for the luminosities not given). In Fig. 4.2 we show how the MAE and R^2 change with the assumed error when predicting SFR. From the figure, we can see that depending on the simulation set (selected lines) the behaviour is slightly different in the metrics. For the MAE, almost all simulation sets increase with the assumed error, except for the simulation set C. This set has a minimum at 0.1 and then increases with the assumed error. In terms of R^2, simulations sets A, B and C seem to peak at assumed error values of 0.1 to 0.2 dex. Simulation set D only decreases with assumed error. In case two FIR line luminosities are available, it depends on the combination of the lines which one is better. For example, simulations sets A and B are much better than simulation set D. This may be the result of including [C II] in the former sets, compared to the [O III] and [N III] lines that are mainly tracing HII regions in the latter set.

In addition, it is clear that having more than two lines helps the model to reduce the MAE and increase the R^2 (set C). This set will have the higher R^2 values for an assumed error around 0.2 dex compared to the other sets with a peak at 0.15 dex. Therefore, we decided to use simulation set C as the reference for setting a default assumed error in DiagISM, as adding more physical information will improve the models. In the web app it is possible to change these values for the M8all model in case a higher or lower assumed error is desired.

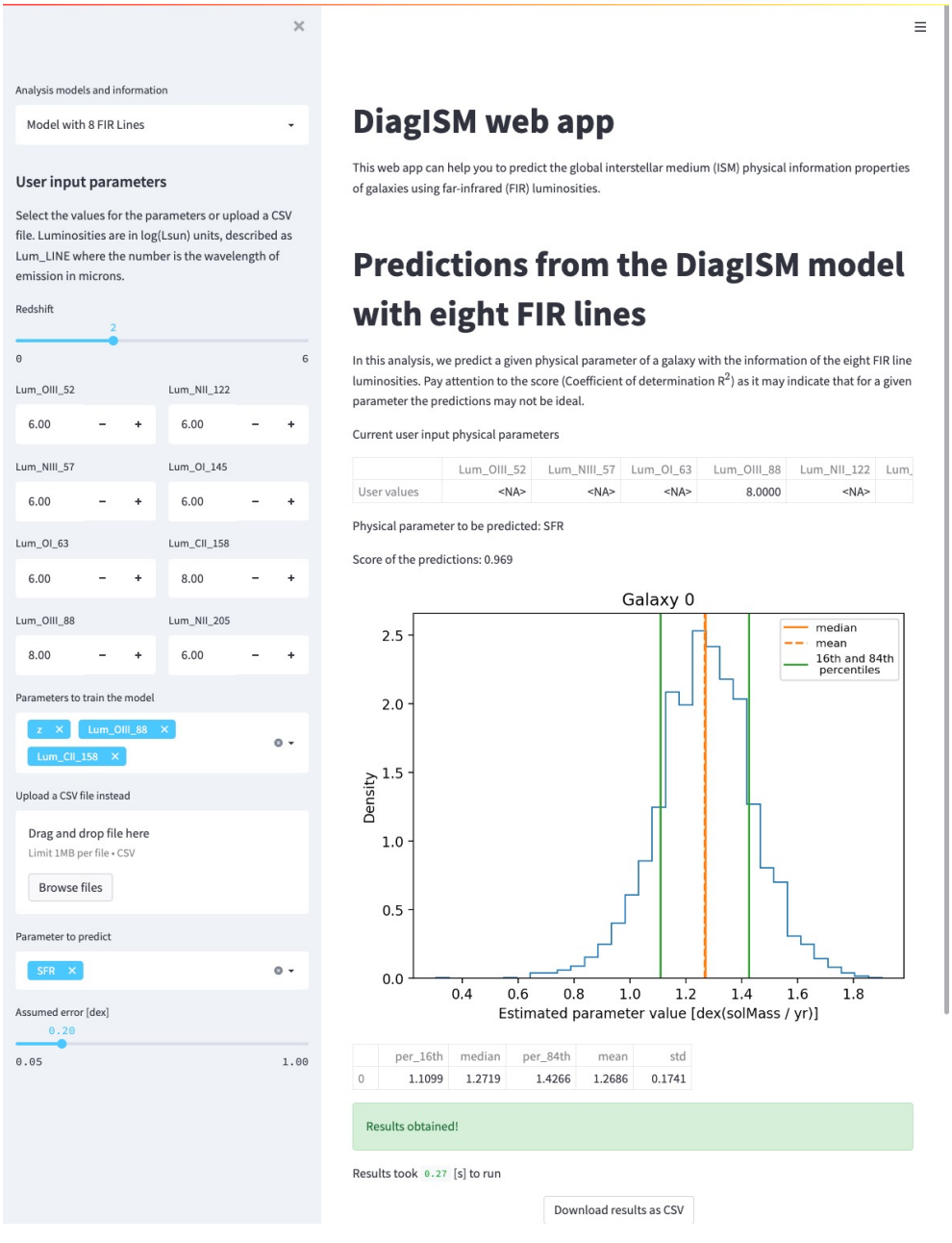

Figure 4.1 – *Screenshot of the* DiagISM *web app showing an example of one of the MLP models presented in this work. The user can select a redshift, different line luminosities and their values which are used to predict a given parameter (SFR in this case).* DiagISM *can also work with CSV files in case the user wants to predict a parameter for more than one galaxy. Final results can also be retrieved as a CSV file.*

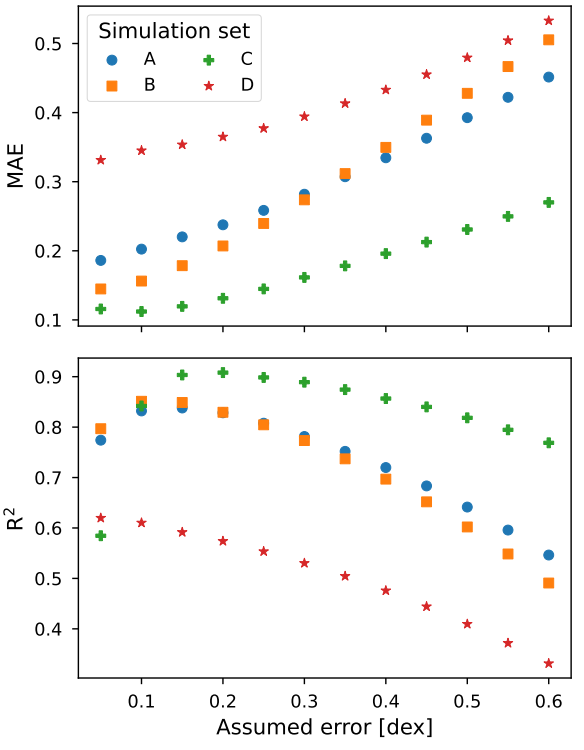

Figure 4.2 – *Change on the median absolute error (MAE, upper panel) and the coefficient of determination (R^2, bottom panel) for a given assumed error when predicting SFR. We use sets of line luminosities estimated from simulations similar to Table 4.1.*

4.4.2 Differences in predictions: M8all vs M2plus

We use the simulation sets to see the difference between the physical parameters estimated in the simulations and the physical parameters predicted from DiagISM. In Figs. 4.3 and 4.4 we present the differences for the eight physical parameters estimated with DiagISM for the two prediction models (M8all and M2plus, respectively). From these figures, we note that both models allow us to recover the physical parameters with differences of less than an order of magnitude, however, there are clear differences between the models. For example, predictions of the M2plus model tend to have a smaller difference from the estimated physical parameters from the simulations compared to the M8all model. However, in most of the parameters, both models are similar.

Regarding the physical parameters, both models allow us to recover SFR, gas mass and the radius of the neutral clouds (R_{cloud}) with a difference of less than 0.3 dex in almost all simulations sets (except simulation set D). In the case of R_{cloud} the differences are even smaller because the range of the parameter is less than an order of magnitude, being the largest difference 0.3 dex. Therefore, these three parameters can be easily recovered with DiagISM.

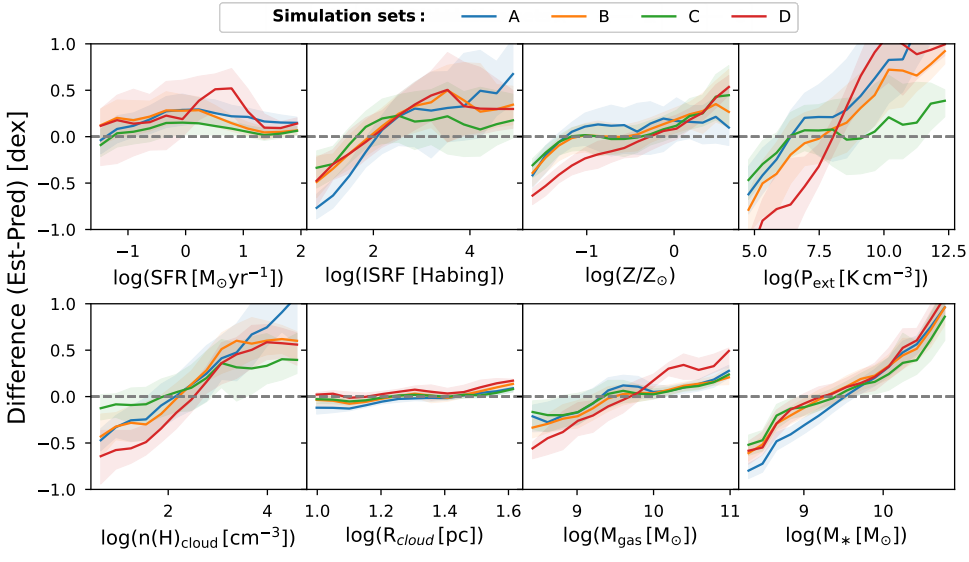

Figure 4.3 – *Difference between the input physical parameter values in the simulations and predicted in* DiagISM *for the M8all model, illustrating how well each physical parameter can be recovered. The different panels show the eight physical parameters that can be retrieved from* DiagISM *with the respective simulation sets in different colours (simulation sets A, B, C and D have blue, orange, green and red colours respectively). We plot a dashed grey line to indicate when there is no difference between the estimated and predicted parameters.*

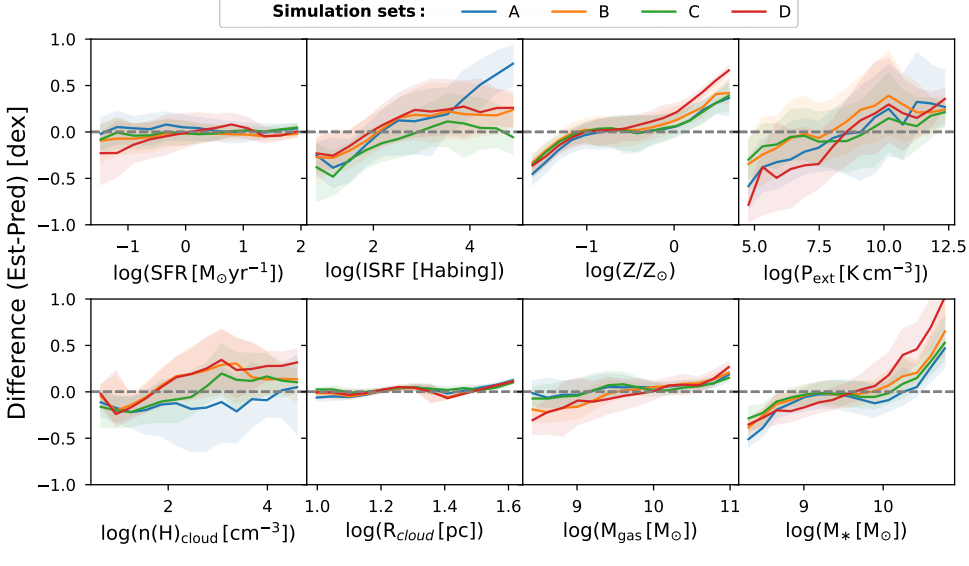

Figure 4.4 – *Difference between the input physical parameter values in the simulations and predicted in* DiagISM *for the M2plus model. The rest is the same as in Fig. 4.3.*

On the other hand, for the other physical parameters, we notice a bias from overestimation (negative difference) to underestimation (positive difference) with an increase in the value for the parameters. The clear example of this bias can be found in the M8all model, where the pressure can be overestimated up to one dex at $\log P_{ext} \sim 5$ and underestimated by more than one dex at $\log P_{ext} > 12$. Similar trends are visible for $n(H)_{cloud}$, $M_\star$ and metallicity but in a lesser extent (i.e. *lesssim*0.5 dex). However the dispersion of the difference in some of the physical parameters is low. For example, for $M_\star$ the dispersion is < 0.3 dex.

There are two parameters where the models differ the most: the external pressure (P_{ext}) and the total hydrogen number density in the neutral clouds ($n(H)_{cloud}$). Regarding these two parameters, in the M2plus model the differences are close to zero and do not evolve much as the value of the parameter increases. However, the mean dispersion of the difference is still very high for this two parameters, ranging from 0.4 to 0.8 dex in both models. In addition, in the M8all model the differences are more bias than in the M2plus model. This suggests that the M2plus model may be better at recovering these two parameters but still with a large dispersion. Finally, metallicity and ISRF are similar in both models, with a mean dispersion of 0.3 dex and 0.5 dex, respectively. In general those predictions will depend more on the selected FIR line luminosities that are used.

It is important to note that some of these differences clearly depend on the simulation set used. For example, the predictions for the simulation set D always tend to differ more than the other sets, while for simulation set C the differences in the predictions are always smaller than those of the other sets. The difference between these two sets is that simulation set D uses two similar lines that trace the DIG and the HII regions, while simulation set C uses three lines that also come from the ISM phases of neutral and molecular gas. This shows the importance of using more than two different lines with a different ISM origin. In conclusion, the results of these figures show us that both models can be used to predict the physical parameters of galaxies with certain considerations in the emission line sets and models used.

4.5 Inferences from observational data

The MLP models presented in this work (Sect. 4.3), available in the DIAGISM web app (Sect. 4.3.4), allow us to predict physical parameters from the observational sample data that we describe in Sect 4.2 and Table 4.1. We present the results on SFR and other physical parameters in this section.

4.5.1 Star formation rate

One of the important physical parameters of galaxies is the SFR. Some FIR emission lines, such as [C II] and [O III] are commonly used to trace the SFR in galaxies (something that we also show in Paper I and Paper II). We use this physical parameter as the baseline to compare with predictions from our MLP models trained on simulations from different sets of galaxies from the observational sample. SFR estimates for individual galaxies may vary depending on the chosen MLP model. Having the two sets of MLP models described in this work gives us an idea of how well we can

constrain the SFR based on FIR emission lines.

In Fig. 4.5, we show an example of two galaxies (Circinus and ESO 173-15) in the set A for the two MLP models presented in DiagISM. For Circinus galaxy, the estimated SFR from M8all is $\log(\mathrm{SFR}[\mathrm{M}_\odot \mathrm{yr}^{-1}]) = 0.66 \pm 0.34$ while for M2plus is SFR$=0.69 \pm 0.14$. This two values are close enough to $\log(\mathrm{SFR}[\mathrm{M}_\odot \mathrm{yr}^{-1}]) = 0.4$ we collected in the observational sample (Paper II and Brauher et al. (2008)). Similarly for ESO 173-15, the SFR from M8all is $\log(\mathrm{SFR}[\mathrm{M}_\odot \mathrm{yr}^{-1}]) = 1.75 \pm 0.16$ while for M2plus is $\log(\mathrm{SFR}[\mathrm{M}_\odot \mathrm{yr}^{-1}]) = 1.69 \pm 0.17$, compared with $\log(\mathrm{SFR}[\mathrm{M}_\odot \mathrm{yr}^{-1}]) = 1.56$ from the observational sample. These SFR values are also consistent if we take into account that the observations also contain systematic errors and we are comparing with models that use data from simulations. From the figure, the difference in the distribution of the estimated parameter is an effect of the construction of the two models. We expect that the M8all model can be well constrained (make better predictions) in general when more than two lines are given for the estimation, whereas for M2plus model we are more likely to have a lower spread distribution since it only depends on the given FIR emission lines input.

From the different sets of observed galaxies, we can get an idea of how the prediction will behave in a real scenario of having different samples of galaxies. In Fig. 4.6, we present the predicted SFR values from the DiagISM compared to the SFR from the observational sample. From the different panels, the first result we can see is that the predictions do a good job to retrieve predicted SFR values similar to the observational SFR of galaxies. Most of the coefficients of determination (R^2) are higher than 0.7. However, there are a few cases where the predictions from DiagISM may not be good enough. For example, predictions from the set D have a large scatter (median absolute error (MAE) values of 0.83 and 0.79 dex for M8all and M2plus, respectively) even if they follow the one-to-one relationship between the predicted and observational SFR. The estimated error bars from the M8all model show that only using $[\mathrm{O\,III}]_{52}$ and $[\mathrm{N\,III}]_{57}$ is not sufficient to constrain the SFR. Similar things happen in set A for the low SFR galaxies, where the information for $[\mathrm{O\,I}]_{63}$ and [C II] is also not well recovered by the model.

Numerical comparisons between the different sets and models are presented in Table 6.4. From them, we notice that in both models (M8all and M2plus) set B has the higher R^2 and a lower MAE. This means that with the lines of $[\mathrm{O\,III}]_{88}$ and [C II] we get better SFR predictions than with $[\mathrm{O\,I}]_{63}$ and [C II] (set A) and with three FIR emission lines ($[\mathrm{O\,I}]_{63}$, [C II] and $[\mathrm{N\,II}]_{205}$, set C). Interestingly, set C with the M2plus model gives the higher R^2 when training the model but their R^2 is worse than with sets A and B.

As a sanity check, we define a baseline by assuming that all the predictions are the average of the sample set. When comparing the metrics with a baseline, we found that any model with set D does even worse than the baseline in terms of the MAE. Therefore we confirm that even having the luminosities of $[\mathrm{O\,III}]_{52}$ and $[\mathrm{N\,III}]_{57}$ is not sufficient to retrieve a correct SFR.

In summary, SFR can be retrieved with DiagISM with an uncertainty of around or below 0.3 dex. The best-case scenario in the models and sets compared in this work show that set B with the model M2plus seems to be the reasonable starting

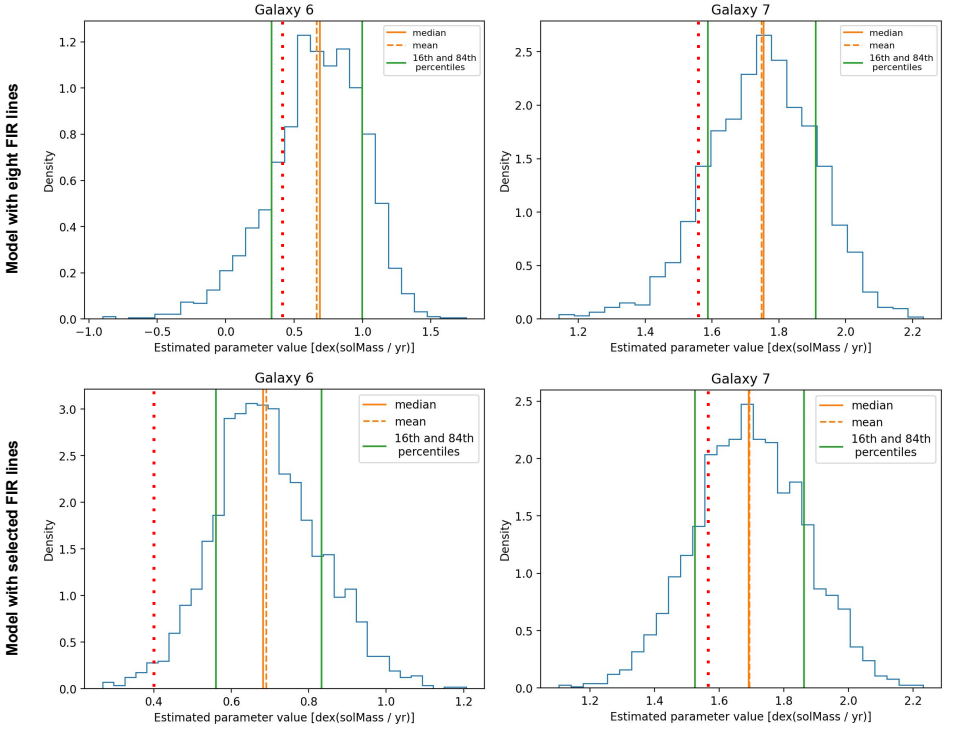

Figure 4.5 – *Example histograms from the* DiagISM *web app with the MLP models presented in this work. We present two galaxies (Circinus and ESO 173-15, galaxy 6 and 7, respectively) that come from the set A. The histograms show the density distribution of the inferred SFRs (blue line) with their respective mean (orange dashed line), median (orange solid line) and 16th and 85th percentiles (green solid lines). SFR values from the observational sample (Paper II and Brauher et al. (2008)) are presented as red dotted lines.*

Table 4.2 – *Prediction metrics for the MLP models and sets compared in* DiagISM*. We show the coefficient of determination (R^2) for the initial training of the model, and the R^2 and median absolute error (MAE) for the combinations of models and sets. In bold, we highlight the highest value for R^2 and the lowest value for MAE for a given combination of model and observational set. The baseline metrics are defined by assuming that the prediction is the average of the sample set.*

Model	Set	Metrics		
		R^2 train	R^2	MAE
Eight FIR lines	A	0.969	0.712	**0.41**
(M8all)	B	0.969	**0.753**	**0.41**
	C	0.969	0.356	0.56
	D	0.969	0.308	0.83
Selected FIR lines	A	0.929	0.796	0.33
(M2plus)	B	0.934	**0.859**	**0.29**
	C	**0.952**	0.729	0.41
	D	0.857	0.378	0.79
Baseline	···	···	0.000	0.690

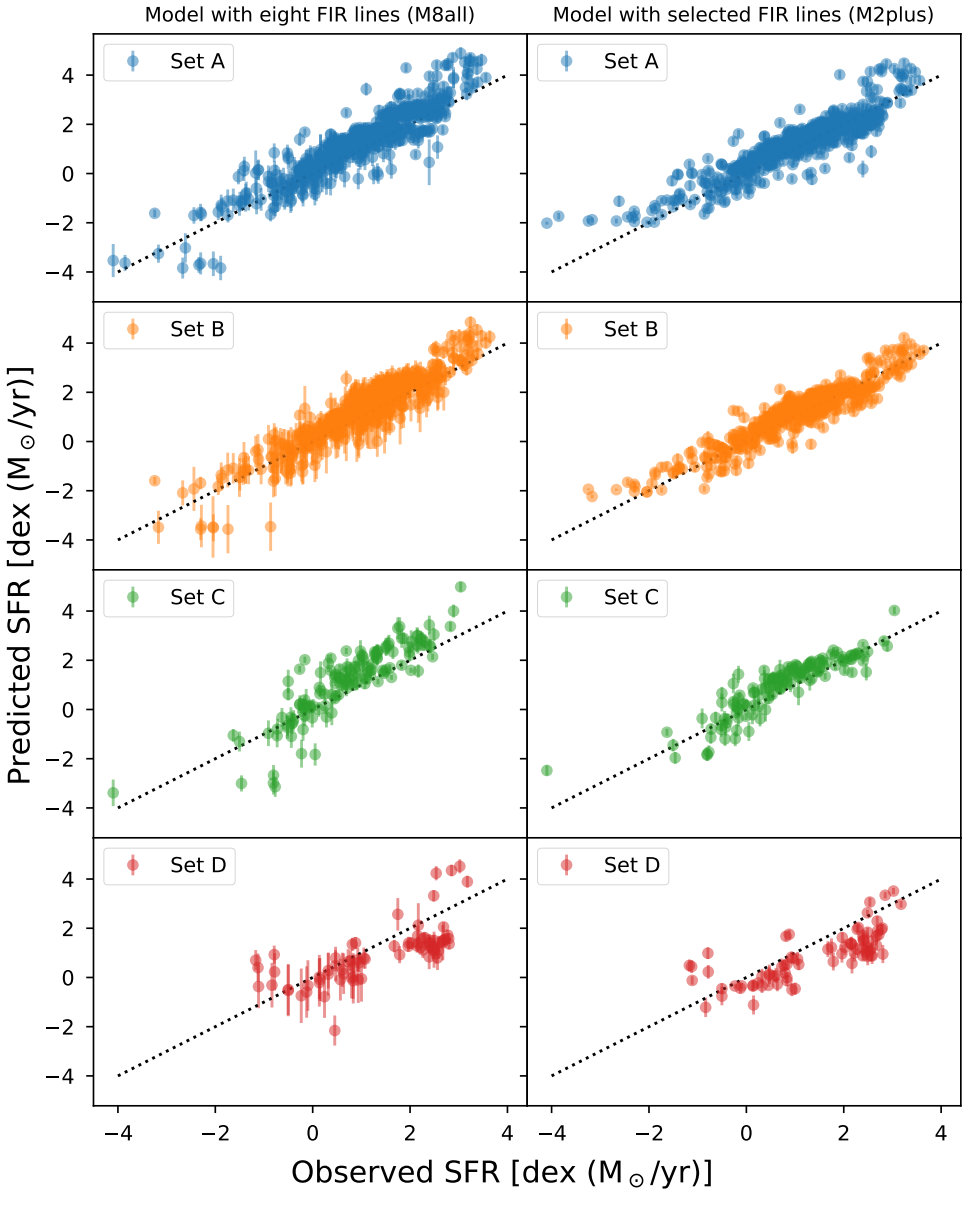

Figure 4.6 – *Comparison between observational estimates of the SFR and predicted SFR values with the* DiagISM. *The first column shows the model with eight FIR lines (M8all), while the second column shows the model with the selected FIR lines (M2plus). At each row on the panel, we show the sets we use to compare the predictions with a different colour (blue, orange, green and red for the respective sets A, B, C, and D). We plot the one-to-one relation as a dotted black line to guide the eye.*

point to predict the SFR of a given galaxy. For this reason, we decided to use that combination to obtain some impressions on how well DiagISM will be able to retrieve other physical parameters.

4.5.2 Other physical parameters

In the observational sample presented in Paper II some information about other physical parameters has been collected for the galaxies that have at least one FIR line measurement. Unfortunately, other physical parameters besides SFR are not easy to get from the literature. Specific samples of galaxies may have stellar masses or metallicities for the galaxies, but the total number of them is not enough to analyse them as we did for SFR. Therefore, we decided to estimate the other seven physical parameters described in Paper II and assume that their difference with the observations will be similar to the difference in the simulation sets (Fig. 4.4). We only use set B from the observational sample (see Table 4.1) with the M2plus model as it appears to be the most similar in terms of SFR (Fig. 4.6).

It is important to notice that some of the galaxies have a very high SFR according to the M2plus model ($\log(\mathrm{SFR}[\mathrm{M}_\odot \mathrm{yr}^{-1}]) > 3.5$). Almost all of these galaxies with overestimated SFR are lensed galaxies originally from Zhang et al. (2018). This may suggest that the luminosities of lensed galaxies may not be correct, suggesting lower luminosities in these galaxies. Something similar happens for the M8all model and galaxies with high SFR in set A. Therefore, caution must be taken when interpreting the physical parameters from DiagISM in lensed galaxies.

Although, ideally we will want to compare the estimated values from DiagISM with other estimates using observational data, compiling this information in a large number of galaxies is outside of the scope of this work. We make available the predicted physical parameters from DiagISM for the 490 galaxies in set B to be used as a comparison in future works. This dataset is available at `https://doi.org/10.5281/zenodo.6705031`. An example of the contents of the dataset is shown in Table 4.3 and Fig 4.9.

4.6 Predictions for future observations

The initial step to observe galaxies with future ground- and space-based observatories is to estimate how many galaxies we expect to find. From a theoretical perspective, it is possible to obtain such estimates based on the specifications (such as line sensitivity and mapping speed) of a future mission. In this section, we estimate line number counts from the linear relationships between line luminosities and SFR presented in Ramos Padilla et al. (2022) and the ML models (Sect. 4.3) for future observations of the FIR emission lines described in this work. To do this, we need to assume a configuration for a future IR telescope and assume the abundance of galaxies to estimate the number counts as close to the Universe as possible. Although there are many ways to obtain the abundance of galaxies (e.g. IR luminosity functions or SFR distribution functions), we decided to use a lightcone catalogue that contain similar information.

Table 4.3 – *Predicted physical parameters for the galaxies inside set B. Example with the top 10 galaxies. The full dataset can be found at* https://doi.org/10.5281/zenodo.6705031.

Galaxy ID	log(SFR) [$M_\odot yr^{-1}$]	log(ISRF) [Habing]	log($Z/Z_\odot$)	log(P_{ext}) [$K\,cm^{-3}$]	log($n(H)_{cloud}$) [cm^{-3}]	log(R_{cloud}) [pc]	log(M_{gas}) [$M_\odot$]	log(M_*) [$M_\odot$]
Mrk 1506	1.85 ± 0.18	3.39 ± 0.24	0.55 ± 0.07	9.05 ± 0.41	3.30 ± 0.29	0.97 ± 0.08	10.42 ± 0.04	10.48 ± 0.01
Centaurus A	0.63 ± 0.16	2.10 ± 0.21	0.32 ± 0.08	7.83 ± 0.29	1.77 ± 0.33	1.12 ± 0.08	10.09 ± 0.07	10.53 ± 0.04
Circinus Galaxy	0.51 ± 0.15	1.97 ± 0.25	0.26 ± 0.08	7.60 ± 0.29	1.52 ± 0.36	1.18 ± 0.10	10.04 ± 0.07	10.51 ± 0.06
ESO 173-15	1.76 ± 0.16	3.10 ± 0.23	0.60 ± 0.08	9.37 ± 0.32	3.19 ± 0.27	0.91 ± 0.07	10.39 ± 0.04	10.50 ± 0.02
ESO 317-23	0.82 ± 0.17	2.25 ± 0.16	0.39 ± 0.08	8.18 ± 0.28	2.11 ± 0.29	1.06 ± 0.06	10.16 ± 0.06	10.55 ± 0.03
ESO 350-38	1.55 ± 0.22	3.08 ± 0.22	0.44 ± 0.08	8.44 ± 0.38	2.81 ± 0.32	1.09 ± 0.07	10.34 ± 0.06	10.47 ± 0.02
Mrk 477	1.51 ± 0.18	2.79 ± 0.22	0.47 ± 0.08	8.91 ± 0.40	2.76 ± 0.31	1.01 ± 0.08	10.33 ± 0.05	10.49 ± 0.01
IC 10	-1.97 ± 0.10	0.38 ± 0.10	-0.30 ± 0.12	5.15 ± 0.15	0.41 ± 0.02	1.36 ± 0.03	8.43 ± 0.05	8.95 ± 0.15
IC 1953	0.23 ± 0.19	1.42 ± 0.33	0.11 ± 0.08	7.10 ± 0.30	0.91 ± 0.26	1.36 ± 0.09	9.88 ± 0.10	10.32 ± 0.17
IC 2554	0.75 ± 0.17	2.13 ± 0.20	0.33 ± 0.08	8.05 ± 0.30	1.91 ± 0.32	1.10 ± 0.07	10.13 ± 0.06	10.53 ± 0.04
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

4.6.1 Lightcone catalogue

Before estimating the expected number counts from the luminosity estimates of Ramos Padilla et al. (2022) and this work, we need to select a catalogue that can be used for future studies considering SFR as one of the main physical properties. Therefore, we chose to use the Simulated Infrared Dusty Extragalactic Sky (SIDES, Béthermin et al. 2017; Bethermin et al. 2022) catalogue*. In brief, the SIDES catalogue was created from dark-matter halos from the Bolshoi-Planck simulation (Rodríguez-Puebla et al. 2016) that have been constrained with observational data for the stellar mass function, the main sequence of star-forming galaxies, and the evolution of the spectral energy distributions (SEDs).

The catalogue contains 5 584 998 galaxies in an area of 2 deg^2. These galaxies span a redshift range of $0 < z < 10$, with a stellar mass range of $7.0 < \log(M_\star[M_\odot]) < 11.6$ and SFR between $-3.65 < \log(\mathrm{SFR}[M_\odot \mathrm{yr}^{-1}]) < 3.00$. Furthermore, for almost all of these galaxies, the catalogue contains estimates of some important submillimeter emission lines, including the [C II] emission line. The $L_{[C\,II]}$ has been estimated in SIDES according to two different methods, showing that different estimates for the $L_{[C\,II]}$ can be obtained using this lightcone catalogue. The first estimate comes from De Looze et al. (2014), where $L_{[C\,II]}$ depends only on SFR. The second estimate comes from Lagache et al. (2018), where $L_{[C\,II]}$ depends on SFR and redshift.

Using the SFR and z available in the catalogue, we estimate the $L_{[C\,II]}$ following the linear relations described in Paper II. In addition, we estimate $L_{[C\,II]}$ from predicting SFR in DiagISM assuming a second-degree polynomial in terms of SFR, as explained in Appendix 4.8. In Fig. 4.7, we compare the two available [C II] luminosities from the works of De Looze et al. (2014) and Lagache et al. (2018) with the estimates we obtain of the linear relation derived from the EAGLE simulations in Paper II and those obtained with DiagISM. We note that most of the galaxies in the SIDES catalogue have luminosities in the range of $5 < \log(L_{[C\,II]}) < 6$ no matter which relation is used to estimate them.

Differences at the high-luminosity end of the distribution of $L_{[C\,II]}$ are important when estimating the number counts. We note that the De Looze et al. (2014) and Lagache et al. (2018) relations tend to estimate a much larger number of galaxies at $\log(L_{[C\,II]}) > 9$ than the approximation used by DiagISM. However, this may be a result of not considering the high standard deviation ($\sim$1 dex, Table 4.6) in the model we are using. Therefore, this distribution will change depending on the observed z and emission line when taking the dispersion into account, as we will see in Sect. 4.6.3.

4.6.2 Future MIR/FIR telescope

To test the predictions of the theoretical models carried out for FIR line emission, we need to obtain observational data that allow us to improve our models to better understand the physics within galaxies. However, the facilities currently exploring the Universe at the required wavelengths are scarce. ALMA, IRAM and SOFIA are the only ground/airborne facilities that we could use for these purposes, but more statistical coverage is needed in the MIR and FIR wavelength regimes. Space

* Publicly available at https://cesamsi.lam.fr/instance/sides/home

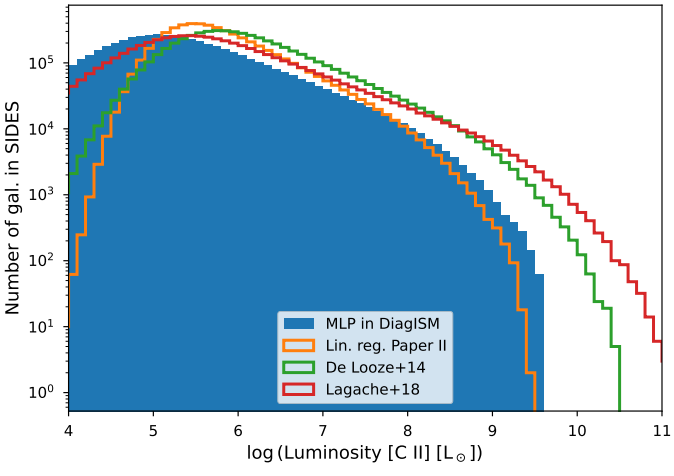

Figure 4.7 – *Histogram of* $L_{[C\,II]}$ *for the total number of galaxies in SIDES. We show estimates of the linear relationships from De Looze et al. (2014) (green line), Lagache et al. (2018) (red line) and Paper II (orange line) along with the estimates obtained from* DiagISM *(blue area).*

Table 4.4 – *Possible configuration of LETO-like receivers. The frequencies and wavelengths are approximations of the central position of the band close to the species that gives it its name.*

Band Number	1	2		3		4
Name (Species)	CI [370]	NII [205]	CII [158]	OI [145]	NII [122]	OI [63]
Frequency (THz)	0.809	1.450	1.900	2.060	2.450	4.700
Wavelength (μm)	370.6	206.8	157.8	145.5	122.3	63.8

telescopes in the IR are the best tools we could use to solve this issue. However, space telescope concepts such as SPICA (Roelfsema et al. 2018) and ORIGINS (Leisawitz et al. 2018) have not reached completion, even though the importance of such missions leaves a gap in the MIR/FIR wavelength range for the next decade.

Recently, the Line Emission Terahertz Observatory (LETO, initially described as FIRSS, Rigopoulou et al. 2021) was one of the candidates for the ESA M7 Mission*. The main goals of LETO were to focus on the assembly of galactic ISM, the onset of star formation and galaxy evolution. LETO was designed to obtain large-area velocity-resolved spectroscopic surveys to fulfil its goals. Its design concept consisted of a 2m telescope passively cooled to ∼50K with shields to minimise the thermal background, covering a wavelength range of 63 to 370 μm (0.81 THz to 4.7 THz) in 4 bands (centred on specific FIR emission lines). With these bands it was possible to obtain information about six lines ([C I] at 370μm, [N II] at 205μm, [C II] at 158μm, [O I] at 145μm, [N II] at 122μm and [O I] at 63μm) with a bandwidth of 16 GHz. In Table 4.4, we present the probable configuration of the receiver channels. The basic configuration of a LETO-like telescope is used in this work to compare the expected number counts in the next section.

* https://www.cosmos.esa.int/web/call-for-missions-2021

Another option is to assume a space telescope such as PRIMA* (previously, also known as Galaxy Evolution Probe (GEP), Glenn et al. 2021), which expects to use imaging and spectroscopy at low and high resolution to measure properties of galaxies with large and unbiased surveys. As a simple comparison, we calculate the number of counts for PRIMA by assuming a 5σ sensitivity per hour of $3 \times 10^{-19}\mathrm{Wm}^{-2}$. Compared to LETO the difference will be in the mapping field of view, where LETO is around 20 arcseconds and for PRIMA is expected to be 10 arcseconds in the best case scenario.

4.6.3 Number counts

After estimating the luminosities of the eight FIR lines described in this work for the SIDES catalogue and assuming the configuration of a space telescope like LETO, we estimate the expected number counts per square degree for four different integration times: 1, 10, 100 and 1 000 hours. We use five of the possible LETO bands, without the band with the shortest wavelength (i.e. band 4, the 63 μm band). In Fig. 4.8, we show the cumulative number counts we expect for the luminosity estimates estimated with the linear relations in Paper II and the second-degree polynomial with DIAGISM in different redshift ranges.

First, when we assume the uncertainty in the relations, the number counts vary significantly. For example, the number counts of the [C II] line can reach luminosities higher than those presented in Fig. 4.7, where we use the coefficients without taking into account the dispersion. This significantly affects the DIAGISM estimates where number counts increases to $\log(\mathrm{L}_{[\mathrm{C\,II}]}) = 9$ and $\log(\mathrm{L}_{[\mathrm{C\,II}]}) = 10$ at $z = 1.3$–1.4 and $z = 0.29$–0.32, respectively. In addition, it also leads to a difference of an order of magnitude or more with the linear regression estimated in Paper II.

Second, for highly ionised gas lines like [O III] and [N III], we see that the number counts for the linear regression with all ISM phases are very high compared to the HII regions regressions and DIAGISM. For example, for the [O III] luminosity at 88μm in the Band 2-CII [158] we estimate ∼11 galaxies at luminosities above $\log(\mathrm{L}[\mathrm{L}_\odot]) = 9$ for the linear relation, while for the HII regions regressions and DIAGISM we have ∼7 ∼2 galaxies, respectively. A more contrasting case is for the [N III] luminosity in the Band 2-CII [158] we estimate ∼23 galaxies at luminosities above $\log(\mathrm{L}[\mathrm{L}_\odot]) = 10$ for the linear relation, while for the HII regions regressions and DIAGISM we have ∼0.5 ∼0.02 galaxies, respectively. This is expected since the linear regression of these lines uses galaxies where the ionising photon flux coming from the star SPH particles in the simulation is very low. Choosing this regression for the number counts can have a big impact on high-luminosity galaxies, especially in the lower redshift ranges. Linear regressions with HII regions are more reliable, but the information about the other ISM phases (e.g. the contribution of the DIG) may be missing. However, the DIAGISM model we use to derive the polynomial regression (M8all) uses the information from the other FIR lines. Therefore, for the [O III] and [N III] lines, we can rely more on the DIAGISM estimates.

Third, the number counts of some of the lines will be very low even with integration times of around 1 000 hours. This is especially the case for [N II] 122 μm and [O I]

* https://workshop.ipac.caltech.edu/farirprobe/page/factsheet

Table 4.5 – *Example table of the number counts for an integration time of 10 hours for the five LETO bands and the eight FIR emission lines. These number counts presented in this table were retrieved assuming the regression from the MLP in* DiagISM *(green line in Fig. 4.8).*

Line \Band	CI [370]	NII [205]	CII [158]	OI [145]	NII [122]
[N II] 205 μm	0.06 ± 0.25	···	···	···	···
[C II] 158 μm	61.74 ± 7.72	0.96 ± 0.95	···	···	···
[O I] 145 μm	0.00 ± 0.06	···	···	···	···
[N II] 122 μm	···	···	···	···	···
[O III] 88 μm	36.42 ± 5.96	3.89 ± 1.93	2.78 ± 1.67	2.72 ± 1.64	2.11 ± 1.43
[O I] 63 μm	3.65 ± 1.87	1.57 ± 1.25	2.99 ± 1.67	2.00 ± 1.42	0.95 ± 0.98
[N III] 57 μm	0.02 ± 0.13	···	0.33 ± 0.56	0.13 ± 0.36	0.03 ± 0.16
[O III] 52 μm	0.38 ± 0.62	1.35 ± 1.17	1.51 ± 1.20	1.86 ± 1.34	1.90 ± 1.35

145 μm. For [O I], the estimates show similar results no matter which regression is used. Only in the redshift range between $z = 1.5$–1.6 it is possible to obtain around 10 galaxies with 1 000 hours of integration time. For [N II], although the estimates are different they show similar results, especially at high redshift ($z = 1.98$–2.10) where it is possible to observe dozens of galaxies with 1 000 hours of integration time. Therefore, these two lines are the most difficult to observe with this type of space telescope design.

Fourth and last, the highest number counts in terms of statistics can be obtained by observing lines like [C II], [O III] and [O I] at 63 μm. For almost all the redshift ranges it is possible to obtain at least 10 galaxies with an integration time of 100 hours. For example, if we use only the CI [370] band we can get a total of ∼450 just from those four FIR emission lines with 100 hours of integration time.

In Table 4.5, we present an example of the number counts presented in Fig. 4.8 with an integration time of 10 hours with DiagISM. From this table it is easy to see that from lines like [N II] at 122 μm we are not going to obtain any detection, while for lines like [C II] and [O III] at 88 μm the numbers are good enough to try to perform blind surveys with some of the bands.

4.7 Discussion

In this section we discuss the repercussions of predicting physical parameters and number counts taking into account the ML techniques. Discussions of the data we use and their accuracy can be found in the discussion section of Paper II.

4.7.1 Using MLP for the estimates

The estimates presented in this work and available in the DiagISM web app use simple MLP regression models that have been fine-tuned for specific physical parameters. These models, based on simulation data, have shown that it is possible to correctly recover SFRs in some galaxy samples with available FIR line luminosities. In the same way, we expect the other physical parameters presented in DiagISM to behave similarly when compared to observed galaxies. Therefore, using a simple MLP to

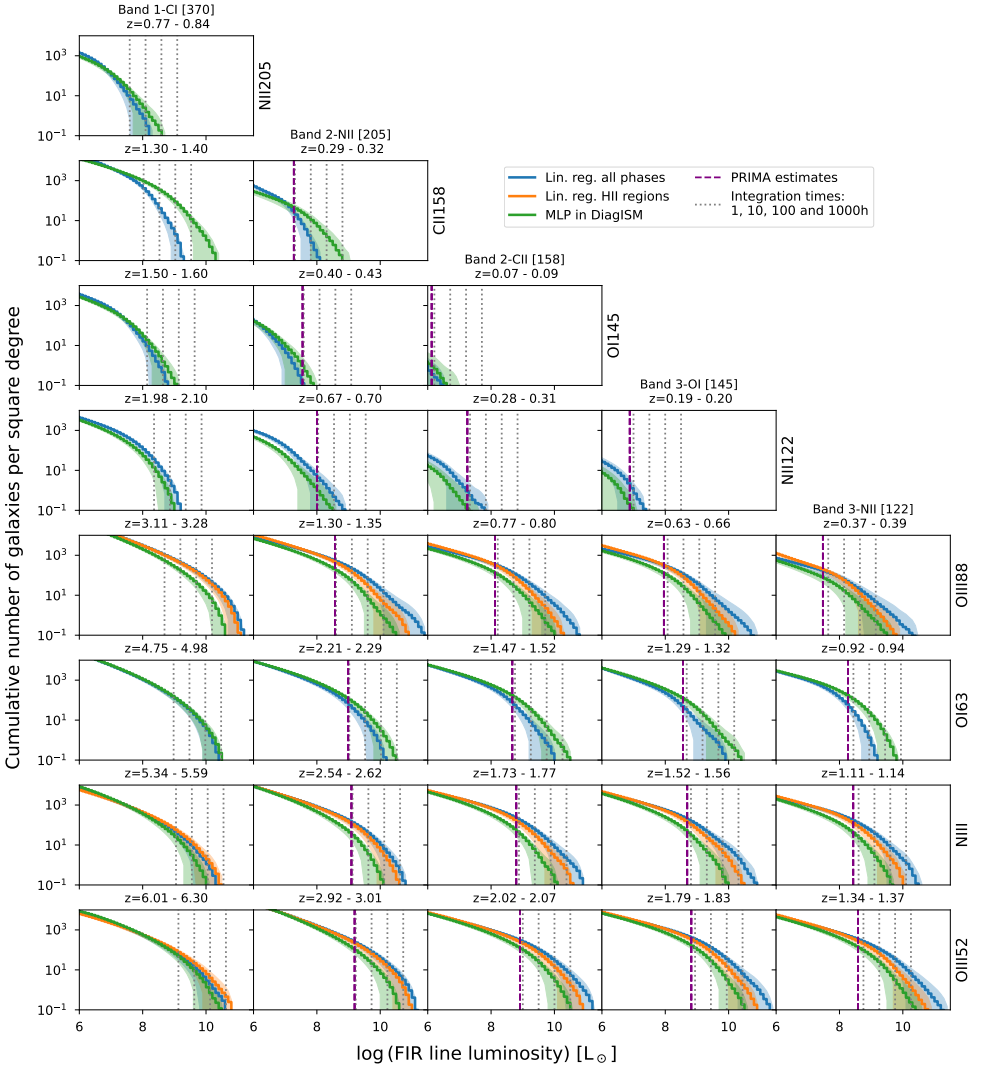

Figure 4.8 – *Cumulative number of galaxies per square degree for the eight FIR emission lines and five bands from LETO estimated according to the linear relations in Paper II (blue and orange lines) and the second-degree polynomial with* DiagISM *(green lines). The shaded areas follow the estimated uncertainty of the estimations. Dotted lines show the integration times of 1, 10, 100 and 1 000 hours from right to left in all the panels. The upper label for each column shows the name of the band in LETO, while the right label in each row shows the FIR emission line. Above each panel, we set the redshift range in which the line is observable for a given band. PRIMA estimates are shown as a purple dashed line for each panel. PRIMA values will be comparable with the those of LETO by one order of magnitude (i.e. the vertical lines of 1000 hours for LETO is similar to 100 hours of PRIMA when taking the field of view into account).*

recover the parameters is just a numerical trick to obtain the physical information hidden in the FIR line luminosities that we still need to understand (e.g. the ISM phases).

More complex ML tools could be used to increase the accuracy of the models presented in this work (M8all and M2plus). However, we argue that the use of such complex tools may be unnecessary as there are still some constraints that we cannot control. For example, observations of some of the bright FIR lines can be helpful (e.g. [C II]), but they only tell us part of the story of the complex ensemble of galaxies in the Universe. Another problem is that simulations reside in physical approximations constrained by computational resources that depend on calibrations of some scaling relationships (e.g. stellar mass functions). Therefore, the use of elaborated regression setups is not (yet) necessary for the current purpose of this work.

The purpose of the DiagISM web app is not to give an exact answer to the values of the physical parameters that we are predicting. Instead, we expect that this tool can be used to lighten and pave the way for future observations. Those observations will help to narrow down the physics we might likely find in distant galaxies, and that will allow us to iterate on to improve current models. That is why we use a combination of simulation and theoretical models to predict the line luminosities and use them as a first step to observe galaxies with future space IR telescopes.

4.7.2 Accuracy of number counts

We estimate the number counts based on the results of the DiagISM web app and the linear relationships of Paper II. From these estimates it was possible to see that for a given integration time of a LETO-like telescope, the cumulative number of galaxies in one deg^2 could differ up to two orders of magnitude depending on the relationship used. Some of these differences can be explained in terms of issues of the model coming from the simulation (e.g. [O III] at 88 μm, see Sect. 4.6.3). However, other relations can also be used for the number counts(e.g. De Looze et al. (2014), Lagache et al. (2018), and others also listed in Paper II for the luminosity–SFR relation) which can give different values. Different luminosity functions, or related relationships, just show the possible scenarios we would expect when observing the FIR line luminosities. Bethermin et al. (2022) discuss scenarios for the [C II] line at $z = 6$, where the different predicted relations show at least the region we expect the real Universe to be in. In addition, other observational considerations in the number counts, such as gravitational lensing, can lead to differences in the estimates. Therefore, the accuracy that we can achieve in the estimated number counts will only be an approximation of the real scenario. Fortunately, these scenarios can be tested with different observing strategies, from a simple shallow and then deep survey (Bonato et al. 2015, e.g.) to a more elaborate and focused strategy like the one propose for a SPICA-like mission (Spinoglio et al. 2021).

4.8 Summary and Conclusions

In this work, we have presented an user-friendly web app called DiagISM which estimates the physical parameters of the ISM present in galaxies up to $z = 6$. This

web app uses the FIR line luminosity model data of Paper II to predict physical parameters such as star-formation rate (SFR), interstellar radiation field (ISRF), metallicity ($Z/Z_{\odot}$), external pressure (P_{ext}), total hydrogen number density in the neutral clouds ($n(H)_{cloud}$), radius of the neutral clouds (R_{cloud}), gas mass (M_{gas}) and stellar mass ($M_{\star}$). We test two multi-layer perceptron (MLP) models within DiagISM (M8all and M2plus) and predict the physical parameters for a sample of observed galaxies with FIR line emissions. In addition, we estimate the number counts that we would expect for future MIR/FIR telescopes. Our main conclusions are as follows:

1. The ISM physical parameter predictions presented in DiagISM depend on the set of FIR emission lines used and the physics contained in those lines. Using three FIR lines ($[O\,I]_{63}$, [C II] and $[N\,II]_{205}$), we obtain more accurate predictions than using any other two lines, suggesting that the more FIR lines we have, the better the model is at constraining the physical parameters. In addition, using FIR lines that cover different ISM phases ($[O\,III]_{88}$ and [C II]) is better than using FIR lines which mainly come from neutral ISM phases ($[O\,I]_{63}$ and [C II]) to estimate SFR.

2. The two MLP models present in DiagISM web app (M8all and M2plus) can reasonably recover some of the physical parameters from simulations. Both models can retrieve R_{cloud} with very small error. In addition, M2plus can retrieve SFR and gas mass with an error below 0.3 dex, while parameters such as ISRF, metallicity and neutral cloud density are retrieved with an error of less than 0.5 dex. Similarly, the M8all model can retrieve parameters such as SFR, ISRF, metallicity and gas mass with an error below 0.5 dex. In addition. there is some predictable bias for some of the parameters, e.g. for the pressure and stellar mass. Therefore, some of the results should be taken as an initial estimate and not as the ground truth.

3. The model that best resembles SFR estimated from observational data in this work is the one that uses two selected $[O\,III]_{88}$ and [C II] FIR line luminosities (M2plus), with a coefficient of determination of $R^2 = 0.859$ and a median absolute error of MAE= 0.29. We used this model to predict the physical parameters from an observational sample and made this dataset available at https://doi.org/10.5281/zenodo.6705031.

4. The expected number counts from a LETO-like space telescope with DiagISM predictions shows that observing FIR lines like [C II], [O III] and [O I] at 63 μm will be ideal for obtaining a statistical sample of galaxies useful to derive physical parameters for future observations. Although some of the number counts can vary significantly depending on the relationship from which they were derived.

Acknowledgements: This research made use of Astropy,* a community-developed core Python package for Astronomy (Astropy Collaboration et al. 2013, 2018). This research made use of several Python packages, among them: streamlit, `scikit-learn` (Pedregosa et al. 2011), `numpy` (Harris et al.

* http://www.astropy.org

2020), `pandas` (Wes McKinney 2010) and `matplotlib` (Hunter 2007). This research has made use of NASA's Astrophysics Data System Bibliographic Services.

Table 4.6 – *Linear relations derived from this work for each of the FIR emission lines using Eq. 4.1*

	Coefficients					
$\log(L_{line})$	c_0	c_1	c_2	c_3	c_4	1σ
[O III] 52 μm	6.227	1.176	0.392	-0.121	-0.024	1.04
[N III] 57 μm	5.740	1.172	0.393	-0.133	-0.026	1.11
[O I] 63 μm	6.245	1.229	0.411	-0.153	-0.032	0.89
[O III] 88 μm	6.241	1.187	0.401	-0.126	-0.028	1.03
[N II] 122 μm	4.741	1.206	0.402	-0.122	-0.029	0.91
[O I] 145 μm	4.662	1.137	0.365	-0.130	-0.007	1.13
[C II] 158 μm	6.228	1.176	0.391	-0.130	-0.024	0.89
[N II] 205 μm	4.650	1.135	0.369	-0.143	-0.005	1.16

Appendix

4

Polynomial regression from MLP in DiagISM

In Paper II we assume that line luminosities changed with SFR and redshift (z), therefore, we fit the estimated data assuming a linear relation for a given z. This time, we assume that the behaviour of the luminosities coming from the SFR predicted with multi-layer perceptron (MLP) in the M8all model of DiagISM follows a second degree polynomial in terms of SFR (i.e. adding a SFR^2 term) in the form:

$$\begin{aligned}\log(L_{line}) \quad = \quad & c_0 + c_1 \log(\mathrm{SFR}) + c_2 \log(1+z) + \\ & c_3 \log(\mathrm{SFR}) \log(1+z) + c_4 \log(\mathrm{SFR})^2, \end{aligned} \tag{4.1}$$

with SFR in units of $M_\odot\ \mathrm{yr}^{-1}$ and line luminosities in $L_\odot$. The values for the coefficients presented in Table 4.6 are obtained by assuming a mock data sample of a million galaxies with luminosities between $2 < \log([L_\odot]) < 11$. We assume a difference between [C II] luminosity and [N II] lines and [O I] at 145 μm of -1.5 dex, and a difference between [C II] luminosity and [N III] of -0.5 dex. We use these values to resemble the distribution of luminosities estimated with EAGLE. In Table 4.6, we present the coefficients down to the third decimal because the estimated errors were similar for all the lines and in general below 0.005. The estimated 1σ shows that the standard error for this approximation is very high as we are covering a large range of luminosities values (9 orders of magnitude) and assuming a random flat distributions in the luminosities.

Histograms of predictions in observable sample

In Fig. 4.9 we show the distributions of the predictions of the observational set B. We can see that the galaxies in set B span different ranges of values for the physical parameters. However, for the stellar mass most of the galaxies tend to be close to $\log(M_\star[M_\odot]) \sim 10.5$. This indicates that the model M2plus in DiagISM is not good at finding good solutions on the stellar mass range that this observation set can have. A deep look at physical parameters other than SFR is needed to clearly understand the possible bias in the other parameters when compared to observed data, especially

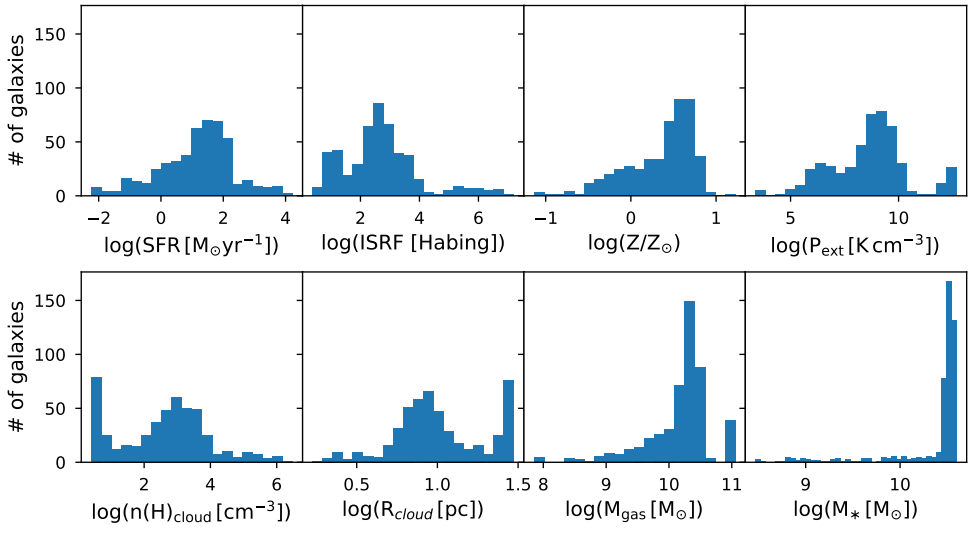

Figure 4.9 – *Histograms of the physical parameters retrieved for the sample of galaxies used in the observational set B. Each parameter is described with 20 bins except for the stellar mass ($M_\star$) where we use 40 to better compare with the number of galaxies in the other panels.*

for $M_\star$. The process of compiling these physical processes with other methods or tools is outside the scope of this paper.

A child blind from birth doesn't even know he's blind until someone tells him.

Stephen King - It

Are you telling me I forgot to put my name?

Andrés filling forms

5

The AGN contribution to the UV-FIR luminosities of interacting galaxies and its role in identifying the main sequence

A. F. Ramos Padilla, M. L. N. Ashby, Howard A. Smith, Juan R. Martínez-Galarza, Aliza G. Beverage, Jeremy Dietrich, Mario-A. Higuera-G., and Aaron S. Weiner

Published in Monthly Notices of the Royal Astronomical Society, Vol. 499, Issue 3, Pages 4325–4369

Highlights

- Interacting systems are crucial to understand the enhancement of star formation in galaxies.
- We use samples of different stages of interacting galaxies along with starburst and AGN galaxies with multi-wavelength coverage.
- We assess the impact of the AGN on interacting systems through SED analysis to infer the physical conditions of AGN and star formation.
- Accounting for the AGN contribution can be important to uncover buried AGN using both line emission and SED estimates.
- We found a modest correlation between AGN fraction and interacting system stages. This can affect their position in the main-sequence of star-forming galaxies.

Abstract

Context: Emission from active galactic nuclei (AGNs) is known to play an important role in the evolution of many galaxies including luminous and ultraluminous systems (U/LIRGs), as well as merging systems. However, the extent, duration, and exact effects of its influence are still imperfectly understood.
Aims: To assess the impact of AGNs on interacting systems, we present a Spectral Energy Distribution (SED) analysis of a sample of 189 nearby galaxies. We gather and systematically re-reduce archival broad-band imaging mosaics from the ultraviolet to the far-infrared using data from *GALEX*, SDSS, 2MASS, *IRAS*, *WISE*, *Spitzer* and *Herschel*. We use spectroscopy from *Spitzer*/IRS to obtain fluxes from fine-structure lines that trace star formation and AGN activity.
Methods: Utilizing the SED modelling and fitting tool CIGALE, we derive the physical conditions of the ISM, both in star-forming regions and in nuclear regions dominated by the AGN in these galaxies. We investigate how the star formation rates (SFRs) and the fractional AGN contributions ($f_{\rm AGN}$) depend on stellar mass, galaxy type, and merger stage.
Results: We find that luminous galaxies more massive than about 10^{10}M$_*$ are likely to deviate significantly from the conventional galaxy main-sequence relation. Interestingly, infrared AGN luminosity and stellar mass in this set of objects are much tighter than SFR and stellar mass.
Conclusions: We find that buried AGNs may occupy a locus between bright starbursts and pure AGNs in the $f_{\rm AGN}$-[Ne V]/[Ne II] plane. We identify a modest correlation between $f_{\rm AGN}$ and mergers in their later stages.

Keywords: Galaxies: active, evolution, interactions, starburst – Techniques: photometric, spectroscopic

5.1 Introduction

Over the past decade, a significant body of evidence has accumulated that supports the existence of a so-called main sequence (MS) of star-forming galaxies (e.g. Elbaz et al. 2011; Speagle et al. 2014), a tight correlation between galaxy stellar mass and the star formation rate (SFR). This scaling relation is claim to be independent of redshift and luminosity (Elbaz et al. 2011), but its normalisation does evolve with redshift (Speagle et al. 2014). Outliers above the MS are often interpreted as merger-driven starbursts with enhanced SFRs (Renzini & Peng 2015; Martínez-Galarza et al. 2016; Pearson et al. 2019a). The relatively tight correlation suggests that the bulk of the stars in star-forming galaxies form via secular processes rather than in violent events, such as mergers (Ciesla et al. 2015, and references therein). However, this correlation depends in part on the assumptions used to calculate SFRs, star formation histories (SFHs), halo properties, and the degree to which galaxy interactions enhance star formation (e.g. Hayward et al. 2014; Matthee & Schaye 2019).

Interacting systems are therefore crucial to our understanding of galaxy assembly over cosmic time, and of the mechanisms that shape the observed scaling relations. In the local Universe, the most luminous infrared galaxies are almost exclusively systems undergoing significant mergers (Stierwalt et al. 2013). In these systems, star formation is significantly enhanced by the funnelling of gas and dust into the nuclear

region, and the thermal emission from obscured star-forming regions outshines the UV and optical radiation from massive young stars. Systems with luminosities greater than 10^{11} $L_{\odot}$ (so-called Luminous InfraRed Galaxies, or LIRGs) are typically found in interacting systems, which results in a strong correlation between enhanced SFR and galaxy interaction (Sanders & Mirabel 1996; Su et al. 2013). However, this simple description does not capture the full range of observed behaviour. For example, Lanz et al. (2013) found no correlation between specific star formation rate (sSFR) and galaxy mergers (see also Silva et al. 2018).

Nuclear starbursts may exist in galaxies that are not undergoing a merger, with about 20% of all spiral galaxies displaying starburst activity in nuclear rings (Brandl et al. 2012). In many of these systems, the active galactic nucleus (AGN) contribution to the luminosity from activity around the supermassive black hole appears to be negligible. These *pure* starbursts are the opposite extreme of systems that are almost entirely dominated by the infrared emission from a dusty torus surrounding an AGN, such as Seyfert galaxies and more distant quasars. To put those two extremes in context, a thorough understanding of the energetics of systems with intermediate AGN contributions is needed. Although star formation dominates the bolometric luminosity of nearby systems during most of the merger, during the later stages an AGN is thought to become active (Sanders & Mirabel 1996; Brassington et al. 2015, and references therein). Presumably, AGNs are fed by the same infalling material that feeds star formation, and the mid-infrared thermal emission from the dusty torus around AGNs can be comparable to that of the dusty star-forming regions (Genzel et al. 1998).

There exists strong theoretical evidence from simulations of mergers (Lanz et al. 2014; Hayward et al. 2014; Dietrich et al. 2018) that AGNs dominate bolometric luminosity during coalescence, and are responsible for quenching star formation in the post-coalescence stages (Dixon & Joseph 2011). This process underlies their transition from the star-forming "blue cloud", through the so-called "green valley", and onto the passively evolving "red sequence" (Ciesla et al. 2015). There is widespread support of this evolutionary path moving from star-forming galaxies to AGN-dominated galaxies (see Sturm et al. 2002; Veilleux et al. 2009; Tommasin et al. 2010; Wu et al. 2011, and references therein), and it is also supported by simulations showing that AGN activity is strongly correlated with the merger stage (Hopkins et al. 2006). Merging galaxies at different interaction stages, ranging from first encounter to post-coalescence, are a natural choice to study AGN evolution and star formation of composite galaxies that combine both starburst and AGN processes.

Uncertainties regarding the energy budget in composite starburst-AGN systems, and about how the two energy generation processes impact one another and evolve, are among the most pressing open questions in astrophysics. For example, buried AGN have been discovered in systems previously catalogued as pure starbursts (e.g. Higuera-G et al. 2009; Dixon & Joseph 2011), and physical models have been proposed to describe the interplay between the two (Ishibashi & Fabian 2016). Discriminating between the two processes based on spectral energy distribution (SED) studies is relatively straightforward and reliable when just one dominates the emission, especially at mid-IR wavelengths. Unfortunately, disentangling them becomes much more difficult when their IR luminosities are comparable (Abel & Satyapal 2008). Optical

and infrared spectroscopy can potentially separate the two if they cover specific fine-structure lines which are prominent in the vicinity of AGNs and weak or non-existing in star forming regions. The best-known example is the BPT diagram (Genzel et al. 1998; Fritz et al. 2006) which separates AGNs from starbursts according to their [O III] λ5007/Hβ and [N II] λ6584/Hα line intensity ratios, among others. The BPT diagrams are not always reliable, however, because high dust opacities toward AGNs can significantly attenuate emission lines at optical wavelengths. For this reason, the absolute strengths of specific mid-infrared emission lines have also been used to estimate AGN contributions (e.g. Genzel et al. 1998). Others have used the silicate attenuation in the SED, or other SED features (Groves et al. 2008; Ciesla et al. 2015). But these techniques, however useful for signaling the presence of AGNs, aren't capable of straightforwardly disentangling the relative importance of star formation and AGNs in composite systems.

To understand the physical mechanisms underlying scaling relations such as the MS, it is of crucial importance to account for the AGN contribution to the total luminosity of merging systems and estimate the SFRs and sSFRs at different interaction stages. The picture at present is somewhat confused. For example, Lanz et al. (2013) and Silva et al. (2018) find no significant change in sSFR with interaction stage, but Lanz et al. (2014) do find that sSFR increases during the relatively short times around nuclear coalescence because the SFR increases but the total mass of stars do not change. Furthermore, combining simulations and multi-wavelength observations, Martínez-Galarza et al. (2016) find that the SEDs of interacting galaxies do change with interaction stage, due to changes in stellar mass and SFR, and that these changes affect the location of galaxies within the MS.

Using SED modeling, Ciesla et al. (2015) showed that the AGN emission could modify the MS slope. Overestimations of the SFR due to the presence of a buried AGN are plausible especially at later stages, and the AGN emission can contribute to the observed MS scatter. Ciesla et al. (2015) verified that these effects can be reduced through broadband SED fitting methods such as CIGALE (Burgarella et al. 2005; Noll et al. 2009; Serra et al. 2011) by taking into account the continuum emission from the AGN to obtain a better interpretation of the star-forming galaxies.

In this work, we apply those SED modeling techniques to four galaxy samples, estimate the fractional contributions of AGNs to their output, and elucidate how that depends on interaction stage. Our approach includes photometry from the UV to the far-infrared to account for multiple emission processes that blend with the AGN emission: UV emission from young stars, optical and near-infrared stellar photospheric emission, mid-infrared emission from warm dust heated by star formation and evolved AGB stars, and cold dust emission. We incorporate photometry from dozens of instruments and surveys.

This paper is organized as follows. In Sec. 5.2, we present the sample selection and in Sec. 5.3 describe the data reduction. Section 5.4 describes how we use CIGALE to analyse our photometric data and the MIR emission lines. We present the derived galaxy parameters in Sec. 5.5 and discuss their implications in Sec. 5.6. We present our conclusions in Sec. 5.7. Our photometry and spectroscopy, as well as the derived parameters for all the galaxies, are presented in the Appendix (available online).

Throughout this paper we adopt $H_0 = 67.7$ km s^{-1} Mpc^{-1} (Planck Collaboration et al. 2016).

5.2 The Four Study Samples

AGN activity ranges from nonexistent to dominant in any particular galaxy. During a galaxy merger, AGN activity can increase over time, so that immediately after coalescence, it is – at least briefly – the dominant contributor to the luminosity (Narayanan et al. 2010; Blecha et al. 2018). Star formation activity is also influenced by mergers, reaching high star-formation intensity in many well-known cases (Veilleux et al. 2009; Stierwalt et al. 2013, among others). But not all AGNs arise in mergers, and not all starburst galaxies host detectable AGNs. Here our approach is to address this ambiguity in a statistical sense by comparing samples of galaxies selected in different ways. Specifically, we attempt to understand how galaxy interactions influence AGN activity by analyzing systems that span wide ranges of 1) interaction stage, from isolated galaxies to coalescing systems, and 2) activity, from AGN-dominated to star-formation-dominated.

We analyze four galaxy samples in the present work. First, we consider a sample of nearby systems selected to span a wide range of interaction stages from isolated systems to strongly interacting systems, the Spitzer Interacting Galaxies Sample (SIGS, Brassington et al. 2015, hereafter B15). Our work in 100 SIGS galaxies (see Sec. 5.2.1) builds on Lanz et al. (2013, 2014) and B15, but is based on a more complete sample, and includes spectroscopic diagnostics. The second sample is selected on the basis of *Spitzer*/IRS emission line ratios to be dominated by star formation (the SB sample, 21 galaxies; Sec. 5.2.2). The third sample is comprised of 29 AGN-dominated galaxies drawn broadly from the literature (Sec. 5.2.3). Finally, the fourth sample is a set of 49 late-stage merging systems chosen to be in or approaching final coalescence (the Late-Stage Merger or LSM sample; Sec. 5.2.4). We include the LSM galaxies specifically to address a gap in SIGS, which lacks late-stage mergers.

Thus our work includes not only systems with *a priori* known dominant activity (AGN or star formation) selected without regard to interaction stage, but also systems with *a priori* known interaction stage selected without regard to activity. We add that none of the galaxies in our four samples are radio-loud based on the identification criteria of Yun et al. (2001) that $L_{1.4GHz} \geq 10^{25}$ W Hz^{-1}. A summary of the four samples is presented in Table 5.1 and are described in detail below.

5.2.1 The Spitzer Interacting Galaxies Sample

Our first sample is drawn from SIGS (B15, see Table 5.4). The SIGS galaxies are relatively bright, nearby systems compiled by Keel et al. (1985) in a manner designed to construct a sample free from morphological bias. Specifically, Keel et al. (1985) identified systems containing a spiral galaxy with a companion seen in close projection, subject to area and magnitude restrictions. These systems comprise the so-called "Complete sample". To augment the Complete sample with more strongly interacting systems, Keel et al. (1985) also compiled a sample of close pairs with pronounced morphological signs of interaction (i.e., tidal tails and asymmetries). This second

Table 5.1 – *Basic data for the four study samples.*

Sample	# Galaxies	References	Description
SIGS	100	1–3	Nearby interacting galaxies presented by Brassington et al. (2015).
SB	21	4,5	Galaxies dominated by star-formation.
AGN	29	5–14	Galaxies dominated by AGN.
LSM	49[a]	15–17	Galaxies close to coalescence with a numerical interaction strength of 4 and 5 (Dopita et al. 2002).
Total	199[b]		All galaxies.

Notes: [a] We present in this work 38 of the galaxies as 11 of the LSM sample galaxies are presented by Dietrich et al. (2018). We re-introduce physical parameters of NGC 2623 as part of the SB sample.
[b] We found reliable SEDs in 189 galaxies (see Sect. 5.5), including the 11 galaxies presented by Dietrich et al. (2018).
References: (1) Keel et al. (1985), (2) Lanz et al. (2013, 2014), (3) Brassington et al. (2015), (4) Brandl et al. (2006), (5) Higuera-G. & Ramos P. (2013), (6) Stierwalt et al. (2013), (7) Keremedjiev et al. (2009), (8) Tommasin et al. (2010), (9) Weaver et al. (2010), (10) Pereira-Santaella et al. (2010), (11) Wu et al. (2011), (12) Dasyra et al. (2011), (13) Wu et al. (2011), (14) Guillard et al. (2012), (15) Wang et al. (2014), (16) Lintott et al. (2008, 2011) and (17) Dietrich et al. (2018).

sample is known as the “Arp sample." The basic properties of all SIGS galaxies are given in Table 5.4, in which the Complete and Arp galaxies are indicated with C and A, respectively. We adopted the distances given in B15 for all SIGS galaxies.

The SIGS galaxies' merger stages were classified by B15, who assigned a numerical interaction strength to each system following Dopita et al. (2002). The classification is based on the degree of morphological disturbance, as follows.

- Stage 1 galaxies are isolated systems without discernible companions and are therefore, by construction, not present in SIGS.
- Stage 2 galaxies are weakly interacting systems, inferred on the basis of their very mild or absent morphological distortions.
- Stage 3 galaxies are moderately interacting, have apparent tidal features, and display moderate morphological distortions.
- Stage 4 galaxies are strongly interacting, with prominent tidal features, but have two separate nuclei that can still be resolved.
- Stage 5 mergers are at the point of coalescence or are merger remnants, and have only one apparent nucleus (the progenitor nuclei cannot be distinguished).

As described in B15, the stage of each system was put to the vote among the authors of that paper, using Digital Sky Survey (DSS) images, and the stage receiving the most votes for each system was assigned. The merger stages classified by B15 are noted in column “Interaction Stage" of Table 5.4. Ultimately, the original SIGS sample was

found to consist of 35 Stage 2 galaxies, 34 Stage 3 galaxies, 33 Stage 4 galaxies, and just 1 Stage 5 galaxy. The SIGS objects treated here are predominantly early-to-intermediate mergers, with just a few late-stage mergers. Thus SIGS is most useful as a means of quantifying AGN activity in mergers *before coalescence*.

SIGS groups 40 and 41 (galaxy pairs NGC 5544/NGC 5545 and NGC 5614/NGC 5615, respectively) overlap too closely to be reliably photometered separately in the *Herschel*/PACS and SPIRE bands. We therefore photometered and subsequently modeled these systems as if they were single objects. NGC 5846 and NGC 5846A from SIGS group 42 were similarly entangled, and we treated them the same way, although we photometered and modeled the other group 42 galaxy, NGC 5850, separately. Thus the apertures given in Table 5.4 for NGC 5544, NGC 5614, and NGC 5846 encompass merging pairs instead of individual galaxies. Taking these considerations into account, the SIGS sample is effectively comprised of 100 galaxies.

5.2.2 The Starburst Sample

Our second study sample consists of galaxies dominated by star formation. This sample, which we refer throughout this work as the SB sample, consists of 21 relatively bright, nearby galaxies known from existing high-quality *Spitzer*/IRS (Houck et al. 2004) spectra taken in Short-High (SH) mode to be dominated by star formation. This requirement for IRS spectra was imposed to facilitate interpretation of the energetics and support the modeling effort, as diagnostic lines of the energetics (e.g. [Ne V] or [Ne II]) fall in the SH bandpass.

The SB sample is a heterogeneous group comprised of two subsamples. First, it includes 16 *Spitzer*-selected "classical" starbursts galaxies from Brandl et al. (2006), selected from its enhanced nuclear star-formation. To these objects we added a selection of bright well-known starburst galaxies also having SH IRS spectra, some of them also form Brandl et al. (2006), including NGC 23, NGC 253, NGC 660, NGC 1797, NGC 3256, NGC 4088, and NGC 4945. A few of the galaxies do have weak AGN signatures as, for example, NGC 253 is known to host a weak AGN (Müller-Sánchez et al. 2010; Higuera-G. & Ramos P. 2013). In addition, the following systems are reported to be undergoing interactions: NGC 660, NGC 1222, NGC 1614, NGC 2623 (see Sect. 5.2.4), NGC 4194, NGC 4676, and NGC 7252. By using this heterogeneous SB sample, we can compare the other samples and check evolutionary connections between them, from the different levels of intensity of star-formation and AGN (Sturm et al. 2002; Veilleux et al. 2009; Tommasin et al. 2010; Wu et al. 2011).

For the SB sample (as well as for the AGN and LSM samples described in detail below), we adopted the redshifts given in NED.

5.2.3 The AGN Sample

Our third sample consisted of 29 strongly AGN-dominated galaxies. We created our AGN sample by selecting galaxies with both strong neon emission lines indicative of high ionising flux (i.e., integrated line intensity ratios [Ne V]/[Ne II] > 0.6; see Sec. 5.4.2), and available archival *Herschel*/PACS and/or SPIRE photometry, as described below.

Our AGN sample includes three galaxies from the Great Observatories All-Sky LIRG Survey (GOALS, Stierwalt et al. 2013, a collection of Ultra-Luminous Infrared Galaxies (ULIRGs) with available *Spitzer*/IRS spectra), that meet our selection criteria: NGC 1068, NGC 7674, and MCG-03-34-63. We also include NGC 4151, a composite AGN/starburst galaxy (Higuera-G. & Ramos P. 2013) in which the AGN is the dominant contributor.

To these we added galaxies from a batch SIMBAD query (Wenger et al. 2000) for suitable targets. Specifically, we retrieved the brightest 20000 galaxies classified by SIMBAD as nearby ($cz <= 29999\,\mathrm{km\,s^{-1}}$) and as AGNs, which also had available photometry from *Herschel*/PACS and/or SPIRE. Of the 20000 galaxies satisfying the proximity, classification, and data availability constraints, we then searched for suitable neon line ratios. We required detections of both [Ne V] and [Ne II], and set a lower limit on the measured ratio [Ne V]/[Ne II] > 0.6.

Estimates of the neon line ratio [Ne V]/[Ne II] for some galaxies appear in different works. In total, we obtained 54 measurements for 26 different galaxies from Keremedjiev et al. (2009), Tommasin et al. (2010), Weaver et al. (2010), Pereira-Santaella et al. (2010), Wu et al. (2011), Dasyra et al. (2011), Wu et al. (2011), and Guillard et al. (2012). The 26 objects satisfying all our selection criteria are classified primarily as Seyferts, including some with hidden broad-line regions. For example, Tommasin et al. (2008) classify MCG-03-34-63 as a non-Seyfert galaxy, but Tommasin et al. (2010) and Weaver et al. (2010) discuss a hidden broad-line region in this galaxy. In some cases the estimated line ratios were discrepant. When multiple measurements were available, we used the most recent, to make use of the best available calibration and pipeline for the data in question.

In summary, our sample of 29 AGN-dominated galaxies consists of one object drawn from Higuera-G. & Ramos P. (2013) (NGC 4151), one object from GOALS (MCG-03-34-064), two objects appearing in both GOALS and Higuera-G. & Ramos P. (2013), and 25 objects drawn from our SIMBAD search. The Fundamental properties of the AGN sample galaxies are given in Table 5.6.

5.2.4 The Late-Stage Merger Sample

The LSM sample is an extension of SIGS emphasizing mergers whose morphology is consistent with the system being close to coalescence. Although SIGS was designed to span the full range of galaxy interaction parameters by selecting strictly on the basis of interaction probability rather than morphology, activity, luminosity, or other derivative indicators, SIGS has relatively few systems at stages 4 and 5. In order to more thoroughly explore the full range of galaxy interactions, we assembled the LSM sample by filtering two catalogs. The first of these, the Revised *IRAS*-FSC Redshift Catalog (RIFSCz; Wang et al. 2014), consists of 60 303 galaxies selected from the *IRAS* Faint Source Catalog (FSC) that contains accurate redshifts and positions as well as some photometry for the galaxies therein. The second catalog, the Galaxy Zoo Data Release 1 (GZ1; Lintott et al. 2008, 2011), consists of almost 900,000 galaxies selected from the Sloan Digital Sky Survey (*SDSS*; Gunn et al. 1998, 2006; York et al. 2000; Doi et al. 2010). GZ1 galaxies were classified by the public into different categories including mergers. Our selection required that galaxies be at redshifts below $z = 0.06$,

and that the fraction of the public votes that the galaxy was a merger was greater than 0.33. These criteria produced the 453 interacting systems that make up the LSM parent sample.

The authors then inspected composite SDSS images of all 453 LSM systems and estimated the merger stages using the same criteria applied earlier to the SIGS systems, as defined in Sec. 5.2.1. In the full LSM sample, only a minority of 24.9% of the sources were classified as being in merger stages earlier than 3, i.e., our selection criteria successfully prioritized advanced mergers marked by obvious morphological distortions that signify a merger near coalescence. In the present work, we analyze all LSM galaxies having available *Herschel*/SPIRE imaging available in the archive. We excluded galaxies that were truncated by the edges of the SPIRE mosaics. We identified a total of 49 LSM objects with suitable *Herschel*/SPIRE imaging for the present work. The basic properties of 38 of them are given in Table 5.7; those for the remaining 12 LSM objects appear in table 1 of Dietrich et al. (2018, hereafter D18). NGC 2623 is a special case of those remaining 12 objects, we re-introduce their physical parameters as part of the SB sample.

In this work we present new SEDs for 188 galaxies. Adding the 11 galaxies from D18 brings the total sample size to 199 galaxies. For reasons fully described in Sec. 5.5, Ten of those 199 galaxies lack SEDs suitable for reliable inferences about the AGN contributions, so we subsequently analyse the implications of the SED fitting for only the 189 remaining galaxies.

5.3 SED assembly

In Secs. 5.3.1–5.3.4 we describe in detail how the SEDs were constructed. In Sec 5.3.5 we also describe additional analysis carried out to retrieve mid-infrared emission line strengths for galaxies in the SB sample.

5.3.1 Image Sources

To ensure well-constructed SEDs, our approach was first, to assemble all available archival imaging spanning the widest possible wavelength range in the thermal regime, and second, to photometer all galaxies in all images within matching apertures. Thus our resulting SEDs fully reflect all the relevant thermal emission mechanisms because they capture the totality of the galaxies' output at all thermal wavelengths, and they also have reliable colors, allowing us to accurately model the separate galaxy components that together comprise the SEDs.

We drew upon imaging data from the following space- and ground-based missions:

- *GALEX* (Martin et al. 2005, the Galaxy Evolution Explorer) for photometry in two ultraviolet bands, the far-ultraviolet (FUV) band centered at 0.152 μm, and the near-ultraviolet (NUV) band at 0.227 μm.

- SDSS DR12 (Gunn et al. 1998, the Sloan Digital Sky Survey) covering the u, g, r, i, and z bands, at 0.354, 0.477, 0.623, 0.762 and 0.913 μm, respectively.

- 2MASS (Skrutskie et al. 2006, Two Micron All-Sky Survey) covering the J, H, and K_s bands at 1.25, 1.65 and 2.17 μm, respectively.

- *Spitzer*/IRAC (Fazio et al. 2004, the Infrared Array Camera) providing mid-infrared coverage in up to four bands 3.6, 4.5, 5.8, and 8 μm.

- *Spitzer*/MIPS (Rieke et al. 2004, the Multiband Imaging Photometer) covering up to three far-infrared bands at 24, 70, and 160 μm.

- *WISE* (Wright et al. 2010, the Wide-Field Infrared Survey Explorer) which covered the full sky in four IR bands centred at 3.4, 4.6, 12, and 22 μm.

- *IRAS* (Neugebauer et al. 1984, the Infrared Astronomical Satellite), another all-sky survey mission that provides photometry in four broad bands at 12, 24, 60 and 100 μm. The *IRAS* photometry used in this work is treated differently in that it was drawn from the Revised IRAS-FSC Redshift Catalogue (RIFSCz, Wang et al. 2014), under the assumption that the IRAS data therein are mature and well-characterised, and the photometry is reliable for total galaxy measurements. We likewise adopted the photometric uncertainties corresponding to the catalogued quality flags for the IRAS bands. We did not use catalogued upper limits.

- *Herschel*/PACS (Poglitsch et al. 2010, Photoconductor Array Camera and Spectrometer) covering up to three far-infrared bands at 70, 100 and 160 μm.

- *Herschel*/SPIRE (Griffin et al. 2010, Spectral and Photometric Imaging Receiver) providing far-infrared imaging at 250, 350, and 500 μm.

For *GALEX*, *Spitzer*, *WISE*, and *Herschel* we relied on archived, publicly available mosaics. We verified the suitability of the available imaging for each galaxy and each band by inspection. Mosaics in which the galaxies were truncated by mosaic edges, and mosaics in which the galaxies were saturated, were not considered valid and were not used. Some archival IRAC mosaics for 20 of our galaxies were not suitable for photometry because of saturation of the galaxy nuclei. Where possible, for these objects we generated our own IRAC mosaics by combining only the short exposures (typically 0.6 sec) from archived IRAC high-dynamic range observations. These short-exposures mosaics were not, generally speaking, saturated, and were in most instances suitable for the photometric analysis described below.

For SDSS and 2MASS, we constructed our own mosaics centered at the positions of the sources listed in Tables 5.4- 5.7, ensuring that they were sufficiently large that the source-free celestial backgrounds could be reliably estimated.

5.3.2 Background Estimation

Accurate background subtraction is crucial for accurate photometry. In the present work, background calculation began with masking of mosaic pixels containing unphysical values, e.g., unexposed pixels not suitable for photometry. We also created a mask for potential contaminating foreground sources (Milky Way stars) by flagging all

pixels with a SNR higher than 3.0 for point sources. This step is crucial for accurate background estimation.

We tested two background estimation techniques on our masked science mosaics, both within the Python package *photutils*[*] (Bradley et al. 2018), an affiliated package of Astropy (Astropy Collaboration et al. 2013).[†] The first technique was *Local Background Subtraction*, where we used an external elliptical annulus of width equal to 10% of the elliptical aperture radius to estimate the background level around the galaxy. The second technique was *Global Background Subtraction*, where the image was analyzed using sigma-clipped statistics, and an overall background estimate of the image was obtained. We compared the global and local background calculations to those from our own custom calculation – also based on masked mosaics – within square regions far from the target galaxies. We found that *Global Background Subtraction* was significantly more accurate than the local technique, so we adopted it subsequently for all our photometry. This choice was validated when we found that our resulting *Herschel*/PACS+SPIRE photometry agreed with published values, within the uncertainties, for sources having published photometry. We speculate that the annuli used for the local background estimation were contaminated by low-level emission from the target galaxies at large radii.

5.3.3 Apertures, Inclinations, and Flux Densities

We used elliptical apertures to estimate total fluxes for all galaxies considered here. Specifically, for a given galaxy, the same aperture was used in every photometric band, to ensure accurate colors and thus reliable SEDs. Each aperture was sized to encompass the maximum apparent extent of each galaxy, as measured either in the *GALEX*/NUV or 3.6 μm IRAC mosaic (or, if the latter was unavailable, the 3.4 μm *WISE* mosaic). We inspected all mosaics of all galaxies with the apertures overlaid to ensure that no flux fell outside them. Based on those inspections, in some instances it was necessary to enlarge or shift the apertures and re-measure the photometry. Ultimately, all apertures were appropriately sized and located to enclose all of a galaxy's flux in all available bands.

We applied appropriate *Herschel*/PACS aperture and colour corrections to account for missing flux due to incomplete sampling of the point spread function (PSF) in each of the PACS bands.

The pixel values within apertures were summed and converted to flux densities using the flux calibrations in the instrument handbooks. We accounted for absolute calibration error by adding appropriate instrument-dependent uncertainties in quadrature to the measurement uncertainties calculated in the standard way. These were as follows: 10% for *GALEX* (Morrissey et al. 2007), 2% for SDSS (Doi et al. 2010), 2% for 2MASS (Cohen et al. 2003b), 3% for IRAC (Cohen et al. 2003a), 4% for MIPS (Engelbracht et al. 2007), 6% for *WISE* (Wright et al. 2010), 10% for PACS (Poglitsch et al. 2010), and 7% for SPIRE (Swinyard et al. 2010). Typically, the calibration errors were much larger than the measurement errors for these relatively bright objects. No additional uncertainties were added to those already adopted from the RIFSCz

[*] https://github.com/astropy/photutils

[†] Further documentation is at https://photutils.readthedocs.io/en/stable.

for the flux densities measured in the *IRAS* bands.

We present our *GALEX* and SDSS photometry in Table 5.9, our 2MASS and *Spitzer*/IRAC photometry in Table 5.10, our *WISE* and *Spitzer*/MIPS photometry in Table 5.11, and finally our *Herschel*/PACS+SPIRE photometry in Table 5.12. When the photometry was consistent with zero flux density (i.e. the estimated uncertainty was greater than the estimated flux density) we chose not to include it in our SED models.

5.3.4 Photometry Validation

We verified that our approach yields high-quality photometry by comparing our measurements to previously published photometry. Specifically, we compared our *Spitzer*/IRAC+MIPS 24 μm photometry to that published previously by B15. Overall we found good agreement. In the following we describe the comparison in detail.

Figure 5.1 compares our IRAC and MIPS 24 μm photometry with that of B15 for all systems common to both studies.

Outliers are apparent, however, and some are significant. To understand the causes of the discrepancies, we obtained from B15 their Source Extractor (SE; Bertin & Arnouts (1996)) output files and examined them in light of our own output from *photutils*. Our findings are listed below in order of significance.

NGC 4933: For two galaxies in this system we used significantly smaller apertures than B15 (the B15 aperture diameters were factors of roughly 15 and 4 times those we used for NGC 4933C and B, respectively), which allowed us to avoid the nearby potentially contaminating IR-bright source SSTSL2 J130402.66−112854.1. We also shifted our aperture center for NGC 4933A by 11″ relative to B15 to avoid potential contamination from NGC 4933B.

NGC 1253A: The B15 aperture is roughly 10.4 times the size of ours. It is a faint source compared to its companion NGC 1253; a nearby bright star (TYC 4711-231-1) lies within the B15 aperture.

IC 1801: The B15 aperture is roughly 4.6 times the size of ours, and contains part of the core of NGC 935 and a nearby source (2MASS J02281028+1934207).

NGC 4567: The B15 aperture is roughly 2.7 times the size of ours, and overlaps with the core of NGC 4568. We shifted our aperture center by 23″ to avoid potential contamination from NGC 4568.

NGC 2820A: The B15 aperture is roughly 5.5 times the size of ours, and therefore includes a nearby star (2MASS J09212802+6413442) that likely contaminates their 3.6 and 4.5 μm photometry.

NGC 5354: The B15 aperture is roughly 6 times the size of ours, and covers the core of NGC 5353. The B15 MIPS 24 μm photometry is only marginally lower (< 2 mJy) than our *WISE* band 4 photometry. Our photometry is however consistent with Vaddi et al. (2016). Due to contamination from the nearby NGC 5353, there appears to be considerable variation in the tabulated photometry of NGC 5354 in the literature (Zucker et al. 2016; Clark et al. 2018).

NGC 2444: The B15 aperture is roughly 2.1 times the size of ours. We offset our

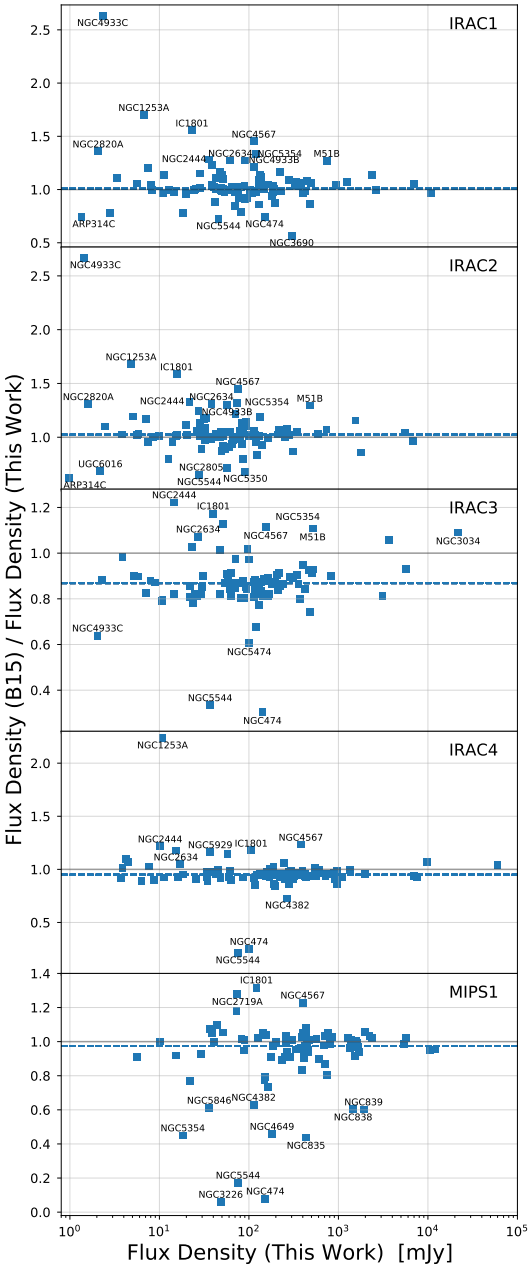

Figure 5.1 – *A comparison of our Spitzer/IRAC+MIPS global photometry to that of B15 for the SIGS galaxies. The photometry is consistent on average, but differences for individual galaxies differences are apparent. Most of the discrepancies are traced to different apertures, as described in Sec. 5.3.4.*

aperture center by 10″ relative to the galaxy center to avoid potential contamination from the nearby galaxy NGC 2445. NGC 2445 is faint at 24 μm so the contamination in the MIPS 24 μm band is not significant.

IC 694 and NGC 3690: It appears that the B15 aperture attributed to IC 694 actually corresponds to a portion of NGC 3690, and that the B15 aperture for the latter is undersized.

NGC 2634: The B15 aperture is roughly 4.2 times the size of ours, potentially admitting contaminating flux from several nearby sources.

M51B: The B15 aperture is roughly 2.7 times the size of ours, and encompasses part of one arm of M51A. This is significant only for the 3.6 and 4.5 μm bands because the relevant portion of that arm of M51A is relatively faint at longer wavelengths.

NGC 5544: In this work we treat NGC 5544 and 5545 as a single system because they are inseparable at *Herschel* spatial resolution, whereas B15 photometered them separately.

NGC 3034: Our aperture is roughly 2.4 times the size of that in B15, explaining the differences in IRAC3.

NGC 474: Our aperture is roughly five times the size of that in B15.

NGC 5474: The centroid of our aperture is offset from that of B15 by 32″ for this diffuse galaxy. Our 5.8 μm photometry is similar to that in Dale et al. (2005).

Arp 314C: Our aperture is roughly 1.4 times the size of that in B15. This is a faint galaxy, and is likely strongly affected by stars lying within the aperture, especially in the IRAC 3.6 and 4.5 μm bands.

UGC 6016: Our aperture is 1.3 times the size of that in B15, and is offset by 10″ to avoid potential contamination of this relatively faint galaxy from nearby bright stars in the 3.6 and 4.5 μm bands.

NGC 5929: The B15 aperture is roughly 1.5 times the size of ours, and our aperture is shifted relative to the galaxy center by 11″ to avoid potential contamination from the nearby NGC 5930.

NGC 4382: Our aperture is roughly 0.77 times the size of that in B15. Our MIPS 24 μm and *WISE* 22 μm photometry is consistent with Boselli et al. (2014). Our IRAC 8 μm photometry is consistent with Amblard et al. (2014).

NGC2719A: B15 aperture is roughly twice the size of ours.

NGC3226: Our aperture is less than half the size of that in B15. Our MIPS 24 μm and *WISE* 22 μm photometry is consistent with the *WISE* photometry reported in Vaddi et al. (2016) and Ciesla et al. (2014).

NGC 4649: Our MIPS 24 μm and *WISE* 22 μm photometry is consistent with the *WISE* photometry reported in Vaddi et al. (2016) and Ciesla et al. (2014).

NGC 835, NGC 838 and NGC 839: We obtain higher mid-infrared flux densities than B15 for these galaxies. Our MIPS 24 μm photometry is consistent with Johnson et al. (2007) and Bitsakis et al. (2011). Our *WISE* 22 μm photometry is consistent with

Zucker et al. (2016).

Having reached down to discrepancies of order 20% (specifically, 23, 23, 13, 19, and 22% in the IRAC 3.6, 4.5, 5.8, and 8.0 μm bands and the MIPS 24 μm band, respectively) relative to B15 without finding any serious faults with our photometry, we carried the comparison no further.

We also compared the *photoutils* photometry for the IRAC 3.6 and 4.5 μm bands to what we measured with the same aperture in the very similar *WISE* bands 1 and 2. In addition, we compared our MIPS 24 μm photometry to that obtained in the similar *WISE* band 4 at 22 μm. The results are shown in Fig. 5.2. In general, the agreement is excellent. A small systematic flux underestimation is present in *WISE* for low IRAC fluxes showing that the background level is overestimated, but this only affects a few galaxies. We were able to resolve most of the discrepant cases with small shifts in aperture centers or diameters, or (in a few cases) by correcting an erroneous background estimate. When we were unable to understand and resolve a pair of discrepant bands, we chose not to use either of them in the subsequent analysis.

On the basis of these two comparisons – of our photometry measured in similar bands and measured by B15 – and the fact that we visually inspected every mosaic for every galaxy with our *photutils* aperture overlaid, we are confident that our photometry is sound and that suitable for the SED modeling described in Sec. 5.4.1.

5.3.5 Mid-Infrared Spectroscopy of Galaxy Nuclei

Mid-infrared spectroscopy provides useful constraints on galaxy energetics because emission lines in the mid-infrared regime reveal the excitation conditions in the ISM nearly free of the usual complications from dust attenuation. For this reason we made use of *Spitzer*/IRS spectroscopy to help quantify the AGN contributions to our sample galaxies. Specifically, we used IRS short-high (SH) spectra of our SB sample galaxies to better understand their energetics via their neon and PAH features. In this Section we describe how we reduced and analysed those spectra. IRS spectra were taken for all our AGN sample galaxies as well, but we did not reduce them ourselves, we took the published neon line ratios from the literature to consistently fulfil the AGN sample selection criteria.

For each galaxy in the SB sample, we began with the SH basic calibrated data (BCD) produced by the IRS pipeline, covering wavelengths from 10 to 20 μm. We reduced the data in the standard way, first using IRSCLEAN to mask cosmic rays and bad pixels. We set the aggressive keyword to 0.5, so that a pixel which exceeds the sigma threshold could only be flagged as bad if it had no neighbours that also satisfied this criterion. We then used the CUbe Builder for IRS Spectra Maps (Smith et al. 2007a, CUBISM) to combine the spatial and spectral information of the datasets, perform background subtraction, and generate a one-dimensional spectrum for each galaxy. We then used PAHFIT (Smith et al. 2007b) to estimate the strengths of the emission features in our spectra.

In general, this procedure worked well, although there were some exceptions. As was also found by Brandl et al. (2006), the nuclei of NGC 520 and Mrk 52 were observed slightly off-center. A more severe mis-pointing was revealed for NGC 3310. Thus for

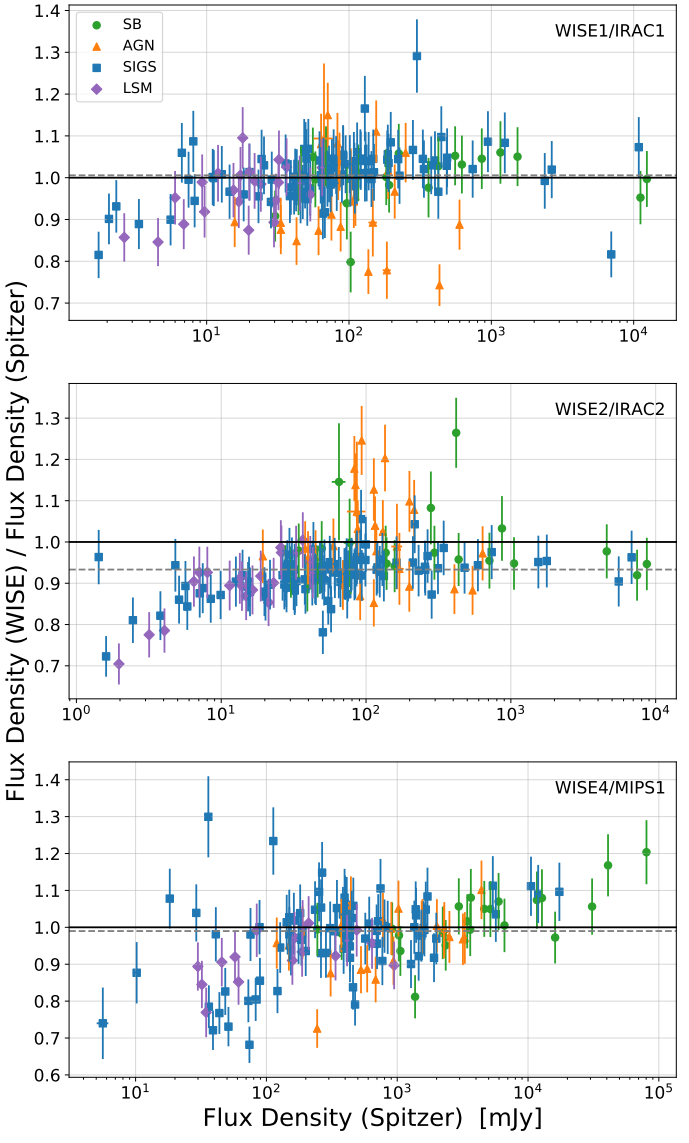

Figure 5.2 – *A comparison of our Spitzer/IRAC 3.6, 4.5, and MIPS 24 μm global photometry to that measured in the WISE 3.4, 4.6, and 22 μm bands for all galaxies in the SIGS, SB, AGN, and LSM samples. The gray dashed line shows the median value of the WISE-to-Spitzer ratio for all galaxies per panel.*

these three sources, our spectra do not represent all the emission from their nuclei. Results from these emission lines are bias to nuclear regions, so comparing with other galaxies not observed only in the nuclei can lead to different estimations. We assume that varying the physical scale of the systems will give similar results in terms of line ratios. The results of our IRS spectroscopy are described in Sec. 5.4.2 and the emission line strengths are tabulated in Table 5.8.

5.4 Analysis

This Section details how the SEDs compiled in Sec. 5.3.1 were modeled to estimate the contributions from young and old stellar populations, thermal emission from dust, and AGNs to the overall emission of each galaxy in our four study samples.

5.4.1 SED Modeling with CIGALE

This work relies primarily on CIGALE* as the means of interpreting galaxy SEDs. CIGALE is a widely used fitting code, based on an energy balance principle, that attempts to model galaxy SEDs in terms of a combination of a small number of separate components that overlap in wavelength. A detailed description of the mechanics of CIGALE are available from Ciesla et al. (2015) and Boquien et al. (2019); here we summarize only the main points relevant to our analysis.

CIGALE works by first populating a high-dimensional parameter grid of SED models consisting of all combinations of user-specified components that contribute to the emission, and then computes the goodness of fit for each model. CIGALE identifies the best-fit SED model by minimising the χ^2 statistic, and produces probability distribution functions for the model grid parameters by assuming Gaussian measurement errors (Burgarella et al. 2005; Noll et al. 2009; Serra et al. 2011). Most relevant to the present work is the fact that CIGALE implements convenient templates for emission from an obscured AGN, based on the models described in Fritz et al. (2006).

We used the parameters and values given in Table 5.2 to define the CIGALE grid of model galaxy SEDs. Except as noted below, the parameter settings were identical to those of D18. For all parameters not shown in Table 5.2, we adopted the CIGALE default settings. All fits were performed assuming the distances given in Tables 5.4-5.7.

We treated the galaxies' star formation histories (SFH) with a delayed SFH model, taking that as a reasonable approximation for the SF history during the last ~ 10 Myr. This approach assumes a single past starburst event (Ciesla et al. 2015). The parameters that control the delayed SFH model are the age of the oldest stars in the galaxy, and the folding time ($\tau_{\rm main}$) of the exponential decay in star formation after the starburst occurs. Depending in the combination of these two parameters, we can simulate ongoing or recent starburst events.

The stellar emission was modeled with the standard Bruzual & Charlot (2003) population synthesis libraries, weighted by the SFH. We used the default CIGALE nebular

* http://cigale.lam.fr/, version 0.12.1

Table 5.2 – CIGALE *grid parameter values adopted for the modeling described in Section 5.4.1*

Parameter	Values	Description
	Star formation history (SFH): Delayed	
$\tau_{\rm main}$	50, 500, 1000, 2500, 5000, 7500	e-folding time of the main stellar population model (Myr).
Age	500, 1000, 2000, 3000, 4000, 5000, 6000	Age of the oldest stars in the galaxy (Myr).
	Single-age stellar population (SSP): Bruzual & Charlot (2003)	
Separation Age	10	Age of the separation (to differentiate) between the young and the old star populations (Myr).
	Dust attenuation: Calzetti et al. (2000)	
$E(B-V)_{\rm young}$	0.1, 0.25, 0.4, 0.55, 0.7	Color excess of the stellar continuum light for the young population.
$E(B-V)_{\rm old\ factor}$	0.22, 0.44, 0.66, 0.88	Reduction factor for the $E(B-V)$ of the old population compared to the young one.
Power-law slope (δ)	0.0, 0.25, 0.5	Slope delta of the power law modifying the attenuation curve.
	Dust emission: Dale et al. (2014)	
α	1.0, 1.25, 1.5, 1.75, 2.0, 2.25, 2.5, 2.75, 3.0	Alpha from the power-law distribution in eq. 5.1.
	AGN model: Fritz et al. (2006)	
$R_{\rm max}/R_{\rm min}$	10.0, 30.0, 60.0, 100.0, 150.0	Ratio of the maximum to minimum radii of the dust torus.
τ	0.1, 0.6, 1.0, 6.0, 10.0	Optical depth at 9.7 μm.
β	−1.00, −0.75, −0.50, −0.25, 0.00	Beta from the power-law density distribution for the radial component of the dust torus (eq. 3 of Fritz 2006).
γ	0.0, 2.0	Gamma from the power-law density distribution for the polar component of the dust torus (eq. 3 of Fritz 2006).
Opening Angle (θ)	60.0, 100.0, 140.0	Full opening angle of the dust torus (Fig 1 of Fritz 2006).
ψ	30.1[a]	Angle between equatorial axis and line of sight.
$f_{\rm AGN}$	0.1 – 0.9 in steps of 0.05	Fraction of AGN torus contribution to the IR luminosity (fracAGN in Equ. 1 of Ciesla 2015)

Notes: [a] The apparent precision was adopted to accommodate an idiosyncrasy in CIGALE's mode of operation, fractional degree precision is not implied.

emission module. The module controlling UV attenuation followed Calzetti et al. (2000) and Leitherer et al. (2002). This module is parameterized by the young population color excess $E(B-V)_{\rm young}$ of stellar continuum light, the reduction factor of the color excess for the old population $E(B-V)_{\rm old\ factor}$ as compared with the young population, the UV bump central wavelength, FWHM, and amplitude (the CIGALE default values for these parameters of 2 175Å, 350Å, and 0 were used), and the power-law slope (δ) which modifies the attenuation curve.

The dust emission was modeled following Dale et al. (2014), implementing a modified blackbody spectrum with a power-law distribution of dust mass at each temperature,

$$dM \propto U^{-\alpha} dU \tag{5.1}$$

where U is the local heating intensity.

We adopted the same overall AGN model as D18 to estimate the AGN fraction $f_{\rm AGN}$ in our sample galaxies, i.e., the Fritz et al. (2006) model. Because one of our primary goals is to investigate the emission fraction coming from the obscured AGNs in our sample galaxies, we sampled the AGN fraction parameter $f_{\rm AGN}$ somewhat more finely than D18, in steps of 0.05 between 0.1 and 0.9, as well as at 0.0 (i.e., zero AGN contribution). We adopted a single value for the viewing angle into the AGN ($\psi = 30.1$), as intermediate between type 1 and 2 AGNs. We tested the effect of varying the viewing angle in the samples of this work. We run a similar grid as in Table 5.2 with half of the steps for α and β, and adding $\psi = 70.1$. In general, changing the angle does not usually make a significant difference in the output parameters. However, we also find that CIGALE can identify Type 1 AGNs: their output parameters, especially stellar mass, suddenly become sensitive with $\psi = 70.1$ and a lower χ^2 compared with $\psi = 30.1$. We detected six AGN galaxies that fall into this category, all of them are already known to be Type 1 Seyfert AGN. We present the derived parameters for those galaxies with $\psi = 70.1$ in Table 5.14. We use these lower χ^2 values in all the Figures of this work. A new version of CIGALE, "X-CIGALE" (Yang et al. 2020), has been recently released that is specifically designed to be more attentive to the angle and to the high-energy contributions to the SED. A study focusing on Type 1 AGN will benefit from both this new version and from a more detailed angle analysis, but is beyond the current work.

We also sampled α in increments of 0.25 between 1.0 and 3.0, and extended the values for the slope delta power-law modifying the dust attenuation curve (0.25 and 0.5 in addition to 0), the optical depth at 9.7 μm (including 0.1) and the density radial exponent of the torus (adding the values -0.5, -0.25, and 0). Our tests indicate that by choosing a compact grid of values for α and $f_{\rm AGN}$, and a single value for the viewing angle into the AGN $\psi = 30.1$, we can obtain well-behaved PDFs for these grid-parameters (i.e., they are well resolved probability distributions). We might therefore expect an improvement relative to the measurements of D18 because of the more finely sampled parameter space.

Our strategy was to run two different families of CIGALE models. The first family included AGNs parameterized according to Table 5.2, while the otherwise identical second family was run without. We adopted this approach because D18 found that $f_{\rm AGN}$ was typically uncertain by $\pm 10\%$. Thus, cases when $f_{\rm AGN} < 0.20$ are not

inconsistent with $f_{\rm AGN} = 0$, i.e., no AGN being present. We therefore chose to treat cases for which $f_{\rm AGN} < 0.20$ as if they had no AGN. We thereby avoid the pitfall noted by Ciesla et al. (2015) i.e., that the AGN contribution can be overestimated, an effect often seen when deriving low-valued parameters with truncated PDF analysis (Noll et al. 2009).

We present a example of the best SED fitting for each of the four samples presented in this work in Figure 5.16.

5.4.2 Neon emission lines

The MIR provides spectral features that are excited by the intense UV radiation from massive young stars. Among the most prominent infrared emission features are the PAHs bands that arise in the photon-dominated regions (PDR) around HII regions. Also, the forbidden nebular lines emitted by ionised atomic gas play an essential role in the characterisation of the gas physics.

Strong radiation fields such as those around AGNs are necessary to reach the ionization potential (IP) of the [Ne V] emission line at 14.3 μm (97.1 eV). Such radiation strength is unlikely to be produced by star formation (Sturm et al. 2002; Brandl et al. 2006). This line is therefore used as a tracer of AGN activity.

An additional advantage of using the [Ne V] emission line relates to the fact that dust extinction at 15 μm is small and typically independent of the orientation (Wu et al. 2011). Goulding & Alexander (2009) show that optical spectroscopic surveys, in contrast, can miss approximately half of the AGN population due to extinction through the host galaxy. Genzel et al. (1998) found a correlation between the strength of emission lines, higher stages of ionisation, and the level of AGN activity. Therefore, [Ne V] can be used to quantify AGN activity. The forbidden [O IV] at 25.9 μm is also used for similar reasons. This line has a lower IP (54.9 eV) and is detected in a more significant fraction of AGNs, but can also be produced in star-forming galaxies, mainly in the presence of WR stars and ionising shocks (Abel & Satyapal 2008; Pereira-Santaella et al. 2010; Weaver et al. 2010; Alonso-Herrero et al. 2012). Although [O IV] has proved to be useful as an AGN tracer by some authors (Veilleux et al. 2009; Gruppioni et al. 2016), here we have decided to use the [Ne V] emission only, in order to avoid any contamination from star formation.

The [Ne II] low ionisation line at 12.8 μm (IP = 21.6 eV) traces the thermal stellar emission in star-forming galaxies (Sturm et al. 2002). Therefore, comparing its strength to that of the [NeV] line provides a straightforward measurement of the relative contribution from star formation and AGN to the overall energy budget. The proximity of the two neon lines in wavelength implies that both of them are subject to similar extinction (Tommasin et al. 2010). A caveat is that the [Ne II] line blends with the PAH feature at 12.7 μm and the [Ne V] line blends with the [Cl II] line. We work under the assumption that the effect of this blending in the estimation of the lone strengths is not very significant. As noted in Goulding & Alexander (2009), it is safe to make this assumption when high signal-to-noise data is available, as is our case. In the present work the uncertainty in line strengths due to blending is smaller than the uncertainty due to instrumental and detection effects.

The [Ne V]/[Ne II] line ratio has been used to calibrate the relative AGN contribution to the total infrared luminosity of galaxies (Genzel et al. 1998; Sturm et al. 2002; Tommasin et al. 2008; Veilleux et al. 2009; Tommasin et al. 2010; Wu et al. 2011; Dixon & Joseph 2011; Díaz-Santos et al. 2017, among others). For example, Wu et al. (2011) estimate a 100% AGN contribution corresponding to a [Ne V]/[Ne II] ratio of $\approx$ 1.0 and a 0% AGN contribution corresponding to a [Ne V]/[Ne II] ratio of $\approx$ 0.01. As pointed out by Petric et al. (2011), discrepancies in the measured contribution of the AGN to the bolometric luminosity can be due to different calibrations of the line ratio. For instance, Pereira-Santaella et al. (2010) have argued that for pure AGN emission, the line ratio should be closer to [Ne V]/[Ne II] $\approx$ 2-3.

The fluxes of the [Ne V] lines for 19 of the 23 galaxies in the SB sample, are presented in Table 5.8 along with other useful lines in the Spitzer SH mode*: the [Ne III], [S III], [S IV], [Fe II] and H_2 S(2) and H_2 S(1). For Mrk 52, NGC 23, NGC 253 and NGC 7714 we took upper limits from the literature (Bernard-Salas et al. 2009; Pereira-Santaella et al. 2010) and we use those values for comparison.

Although the results for the SB sample fall in a region dominated by upper limits in the detection of the [Ne V]/[Ne II], some of the results could be affected by the adopted emission-line detection procedure and signal-to-noise ratio threshold used in weak cases, as noted by Goulding & Alexander (2009). Most of the spectroscopic data come from the nuclear region of the galaxies, so when we compare with the global values of the galaxies in the SED for the given apertures, we are comparing global characteristics with a measure of the central emission (most predominant) region of the samples.

5.5 Distributions of Derived Galaxy Properties

This section describes the CIGALE-based SED fitting results for the 199 objects in the SIGS, SB, AGN, and LSM subsamples described in Sect 5.2. The large overall sample size and the well-defined subsets facilitate some useful statistical comparisons. A total of 94 objects have $f_{\rm AGN} \geq 0.2$; for these objects we present the CIGALE parameters as computed. For the remaining objects we present the CIGALE parameters as computed with their AGN contribution set to zero. Ten galaxies (marked with an [b] in Tables 5.4-5.7) have sparse photometric coverage and consequently lack reliable SED fits; we omit these objects from further analysis.

Out of the $\sim$ 60 parameters that CIGALE estimates, we focus on those most relevant to star formation and AGN activity; the parameters we emphasize are listed in Table 5.13.

CIGALE treats the AGN as a composite object consisting of contributions from three elements in the context of the Fritz model, namely: 1) the primarily mid-IR emission arising from the molecular torus, 2) the emission from the accretion disk in the optical and near-IR, and 3) light scattered from the torus. The CIGALE parameter $f_{\rm AGN}$ is typically used to mean the ratio of the mid-IR emission from the torus only to the total infrared luminosity (see for example D18). However, our investigations show that

* With a slit size of 4.7″× 11.3″

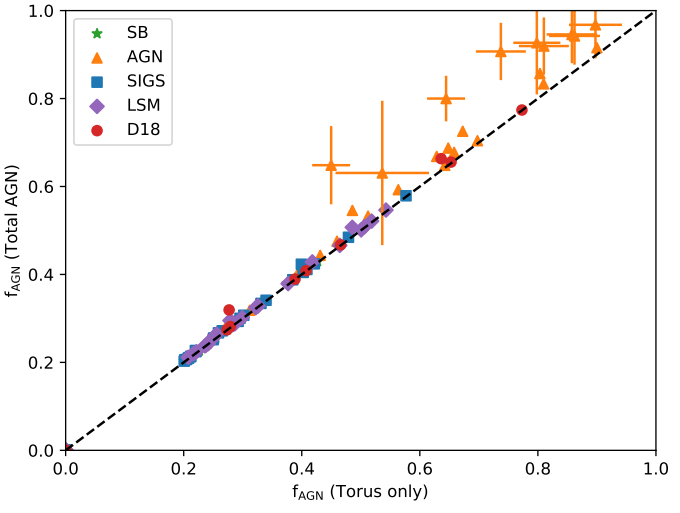

Figure 5.3 – *The total AGN contribution to the infrared luminosity, $f_{\rm AGN}$, compared to the fractional contribution from the molecular torus alone in* CIGALE*'s implementation of the Fritz AGN model. For clarity, error bars are only shown for galaxies that deviate significantly from the line of equality. The uncertainties for points without drawn error bars are similar. Symbols indicate which of the four subsamples the objects belong to (Sec. 5.2); the D18 galaxies are also shown. Here and in the main text we use $f_{\rm AGN}$ to indicate the total AGN fraction, which differs from the popular convention that considers emission only from the torus (see Sec. 5.5). The distinction matters for a significant number of objects in the AGN sample.*

the emission from the torus does not accurately account for the total AGN output for some of our galaxies (Fig. 5.3). In most cases this makes very little difference, as demonstrated by the near one-to-one correlation between $f_{\rm AGN}$ (TOTAL) hand the fractional contribution arising only from the torus ($f_{\rm AGN}$ (Torus Only), in Fig. 5.3). Therefore, throughout this work we define $f_{\rm AGN}$ as the contribution coming from the torus. We find that nine objects (NGC 3516, NGC 5548, ESO 141-55, Mrk 771, Mrk 841, Mrk 1383, Mrk 1513, Mrk 335, and 2XMM J141348.3+440014) are significant outliers of this correlation. All of them are characterized by an accretion-disk luminosity that exceeds that of the torus, including both the thermal and the scattering components, as calculated by CIGALE using the Fritz model. They all have good wavelength coverage in their SEDs and reliable CIGALE fits, with reduced-χ^2 between 1 and 3. No other CIGALE parameters single out these objects as having high accretion or identify them as unusual in other ways. The SFRs in this set, for example, vary from about $30.5\,{\rm M}_\odot\,{\rm yr}^{-1}$ (Mrk 1513) to $0.04\,{\rm M}_\odot\,{\rm yr}^{-1}$ (NGC 3516). The most extreme outlier, Mrk 771, has an accretion luminosity almost five times larger than its torus emission. This object is noted for having soft X-ray excess emission of $0.15 \times 10^{-11}\,{\rm erg\,cm^{-2}\,sec^{-1}}$ (Boissay et al. 2016). The excess in the soft band and the high accretion luminosity favors an interpretation in which UV photons from the accretion disk are comptonized by the electrons in the hot plasma (comptonization) as the cause for excess soft emission(Boissay et al. 2016).

For purposes of qualitative illustration, we collect the CIGALE model SEDs in

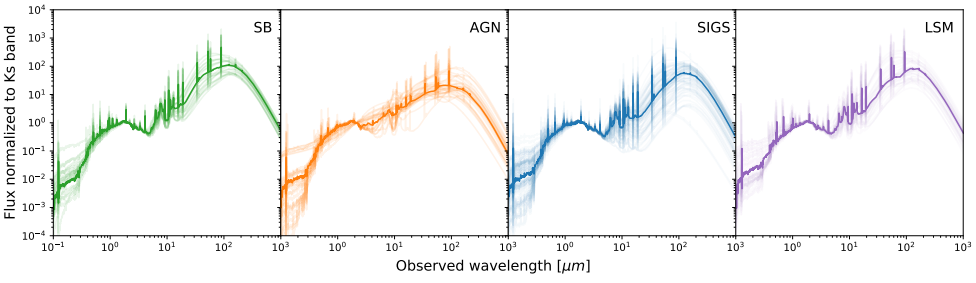

Figure 5.4 – *Model spectra normalized in the K_s band for all galaxies in this work, separated by sample. From left to right: the SB sample (in green), the AGN sample (in orange), the SIGS sample (in blue), and finally the LSM sample (in violet). Faint spectra are those of individual galaxies. The bold lines indicate the median-averaged spectra for the full subsamples. This comparison highlights the differences seen among the subsamples on average, especially those of the AGN sample in the mid-infrared relative to the other samples.*

Fig. 5.4. We indicate the median-averaged SEDs with bold lines, normalized to their K_s flux densities, for each of our four subsamples, together with the most likely fitted SEDs. Some aspects of the SEDs are immediately evident. For example, the SB and LSM subsamples show qualitatively similar overall behavior, which suggests that star formation dominates the SEDs of the LSM galaxies. There is however a weak bump in the median SB and LSM SED at about 50 μm of uncertain origin; it may reflect the presence of warm dust. The AGN sample has a higher median ratio of NIR to FIR flux than do any of the other samples. It also, unsurprisingly, has a higher median ratio of MIR to FIR flux, reflecting the presence of the hot dust associated with the AGN component. Finally, the SIGS sample shows the greatest variety in individual galaxy SEDs. The latter can be understood as reflecting the much larger variety of star formation activity present throughout the merger sequence as compared with our other samples.

We calculated IR luminosities (L_{IR}) by integrating the best fitted SEDs from 5 to 1000 μm. We chose this definition to conform to that in Fritz et al. (2006), to account for PAH features between 5 and 8 μm, and to avoid the near-infrared stellar emission that enters into the 1-5 μm window. When the contribution of the AGN to the SED drops below about 20% it becomes increasingly difficult to use the SED to reliability determine the Fritz parameter values.

To illustrate the effect of low AGN fractions in the accuracy of our results, in Figure 5.5 we include all fits with f_{AGN} fraction values larger than 15%, i.e., we include objects that are below the 20% threshold of what we consider reliable AGN fractions. The uncertainties of these points in the plots are not any larger than those of higher AGN fraction, but a closer look at the SED fits and their reduced-χ^2 in all samples prompt us to use the 20% cutoff in the remaining figures so that genuine effects can be highlighted. (A galaxy whose AGN contribution is less than 20% is then reanalyzed with CIGALE with the AGN parameters set equal to zero, and the other resultant parameter values are the ones listed in the Tables.) The left panel of Fig. 5.5 shows no clear relation between f_{AGN} and L_{IR} for any of the subsamples considered here, but it does clearly reveal the tendency of the AGN sample galaxies to host strong and

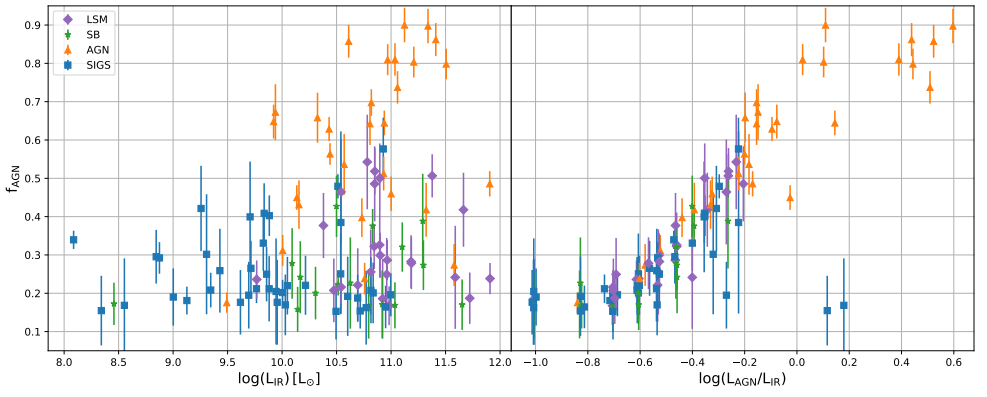

Figure 5.5 – CIGALE-*estimated total fractional AGN contribution to the IR luminosity versus* CIGALE-*estimated total luminosities for the SIGS (blue squares), SB (green stars), AGN (orange triangles) and LSM (violet squares) galaxies. Left panel: AGN fraction as a function of* L_{IR}. *Unsurprisingly, the AGN sample behavior is distinct from that of the other three subsamples. Right panel: AGN fraction as a function of the ratio of AGN luminosity and* L_{IR}. *AGN emission at UV wavelengths can in some cases drive the ratio of AGN luminosity to IR luminosity to values greater than unity.*

5

in some cases dominant AGNs. Measurable AGN contributions are only present at IR luminosities above $\sim 10^{9.5} L_\odot$.

f_{AGN} is plotted as a function of L_{AGN}/L_{IR} (L_{AGN} from CIGALE output) for all modeled galaxies in the right panel of Figure 5.5. The expected relationship is apparent, and at the smallest values of f_{AGN} the flattening confirms our decision to limit further analyses to values exceeding 20%. Toward large ratios, the trend in f_{AGN} flattens and becomes more scattered. The flattening is the result of a larger fraction of the AGN luminosity being emitted at UV wavelengths for the brightest AGNs. The scattering, on the other hand, is explained by the different levels of obscuration in each case, related to geometrical (i.e., inclination) effects, and the contribution of the f_{AGN} coming only from the torus (Fig. 5.5).

5.5.1 Galaxy Properties by Merger Stage

We segregated the LSM+SIGS galaxies by merger stage and compared them to the SB and AGN subsets using a Kolmogorov-Smirnov (KS) test to determine how statistically different the derived parameters are between samples. The KS statistic is a measure of how likely it is that two distributions are consistent with being two realizations of the same underlying distribution. The higher the KS probability between two parameter distributions, the more likely it is that they are coming from the same parent distribution.

Lanz et al. (2013) reported KS tests on a smaller sample of mergers in the original SIGS program (before the availability of Herschel data) and tentatively did not find statistically significant correlations between SED shape, merger stage, and star-forming properties. With our enlarged study sample, analyzed with an SED code that does take into account the AGN contribution, we are more successful at finding

Table 5.3 – *Results of KS tests comparing fitted parameters by subsample.*

Samples Compared	$f_{\rm AGN}$	L(AGN)	E($B-V$) (Old Stars)	E($B-V$) (Young Stars)	α (Dust)	L(dust)	SFR	age(stars)	Mgas	M_*	sSFR
2-3	0.44	0.42	0.95	0.96	0.24	0.88	0.91	0.62	0.59	0.92	0.24
2-4	0.3	0.05	0.01	0.01	0.38	<0.01	0.01	0.07	0.42	0.26	0.5
2-5	0.12	0.01	0.03	0.03	0.34	0.02	0.05	0.75	0.49	0.49	0.07
2-AGN	<0.01	<0.01	<0.01	0.66	<0.01	0.02	0.84	0.19	<0.01	<0.01	0.09
3-4	0.43	0.18	<0.01	0.01	<0.01	0.01	0.02	0.26	0.63	0.49	0.1
3-5	0.21	0.02	0.03	0.08	0.06	0.05	0.17	0.8	0.22	0.22	0.14
3-AGN	<0.01	<0.01	<0.01	0.86	<0.01	<0.01	0.76	0.03	<0.01	<0.01	<0.01
4-5	0.61	0.42	0.51	0.97	0.84	0.83	0.98	0.92	0.51	0.51	0.38
4-AGN	<0.01	<0.01	0.31	0.06	0.08	0.43	0.10	0.01	0.01	0.04	<0.01
5-AGN	<0.01	0.02	0.56	0.15	0.38	0.56	0.15	0.17	0.07	0.07	<0.01
SB-2	0.31	0.03	0.01	0.01	0.11	<0.01	0.01	0.2	0.43	0.41	0.05
SB-3	0.61	0.05	<0.01	0.01	<0.01	<0.01	0.02	0.44	0.35	0.35	0.1
SB-4	>0.99	0.85	0.22	0.43	0.42	0.43	0.43	0.76	0.98	0.99	0.18
SB-5	0.53	0.38	0.08	0.64	0.93	0.26	0.48	0.85	0.58	0.64	0.96
SB-AGN	<0.01	<0.01	0.07	0.01	0.60	0.07	0.04	0.03	0.04	0.09	<0.01
noAGN-AGN	<0.01	0.01	0.18	0.03	<0.01	0.28	0.04	0.01	0.02	0.02	<0.01

Note. KS-derived probabilities indicating the likelihood that the CIGALE-derived parameter distributions (column headings) for two galaxy subsamples were drawn from the same parent sample. The left-hand column indicates the two samples tested. For this test, the SIGS and LSM samples were combined and then divided into merger-stage-based subsamples. The numbers in the left column indicate the merger stage used. The 9th column [age(stars)] refers to the mean age of the stellar population. Because the subsamples were relatively small (e.g., 29 AGN galaxies), we made no attempt to refine probabilities below 1% (> 99% confidence that the two samples differ).

meaningful statistical differences.

We summarize the results of this comparison for selected parameters in Table 5.3 and present the distributions for SF and SFR in Fig. 5.6. Analogous results for the other derived parameters are presented in Fig. 5.17.

The KS tests reveal a number of trends:

- First, the parameter distributions of advanced merger stages (4,5), especially $f_{\rm AGN}$, M_*, and $\alpha_{\rm dust}$, are most similar to those of the starburst sample, and are less statistically correlated with the parameters of the AGN sample.
- Second, KS scores for consecutive stages (2-3, 3-4, 5-6) are higher than for non-consecutive stages, with the smallest correlation occurring for stages that are farther apart along the merger evolution (e.g. 2-5), which is expected if the properties of the system evolve gradually as the merger progresses.
- Third, apart from their dust luminosities and $E(B-V)$ values, the AGN sample parameters have a very small statistical correlation with the parameters of any other samples, but the KS scores are slightly larger between the AGN galaxies and the advanced merger stages than between AGN and the early merger stages.
- Fourth, lowest statistical correlation occurs between the SB and the AGN samples, even for parameters that tend to be correlated between all the other samples.

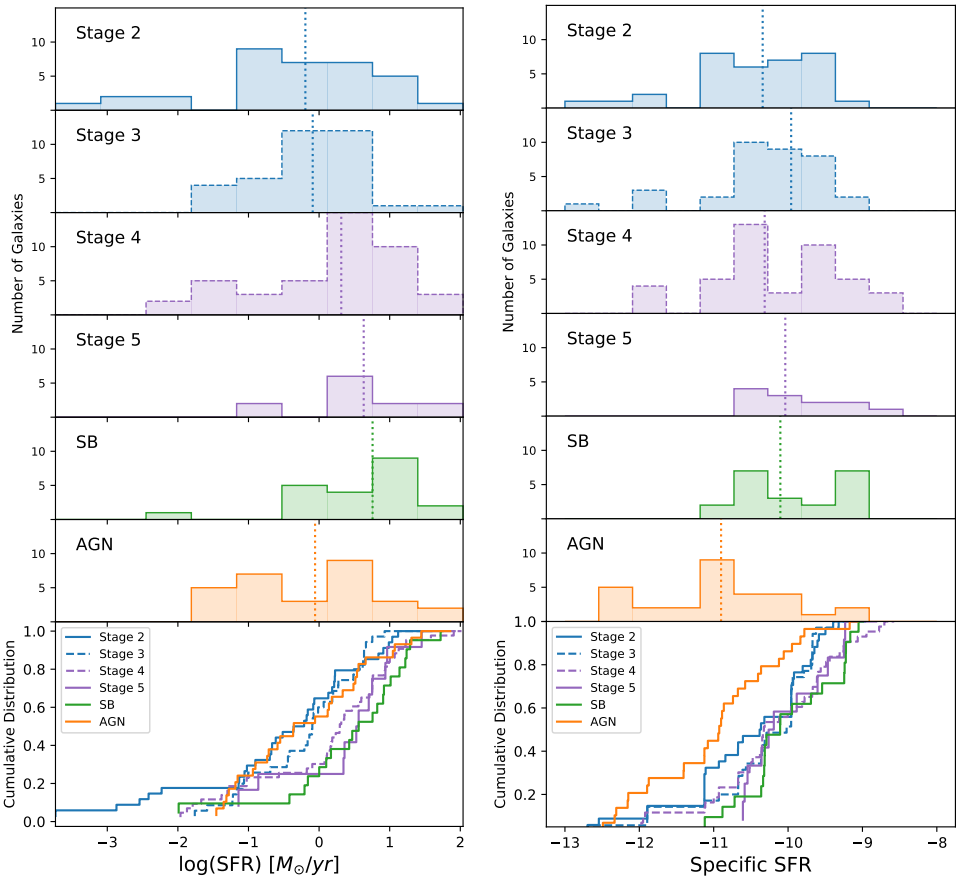

Figure 5.6 – *Histograms (top) for the stages of the combined SIGS+LSM sample and the normalized cumulative distributions (bottom) of SFR (Left panel) and sSFR (Right panel). In the histograms we present the stage 2 (blue), 3 (dashed blue), 4 (dashed purple) and 5 (purple) from the SIGS+LSM sample, and the SB (green) and AGN (orange) samples. The respective median value is represented in the histograms by dotted vertical lines. In the cumulative distributions we present all the previous samples.*

- Finally, the parameters that show more dispersion between samples are $f_{\rm AGN}$ and $\alpha_{\rm dust}$, which implies that they are the most useful parameters to discriminate between galaxy types.

The picture that emerges from these results (and from the overall SED shapes in Fig. 5.4) is in agreement with a classical interpretation: in the local Universe, mergers trigger starburst activity in galaxies (Hopkins et al. 2006; Pearson et al. 2019a). They also trigger AGN activity, but to a lesser extent (Hopkins et al. 2006; Satyapal et al. 2017; Blecha et al. 2018; Dietrich et al. 2018).

5.5.2 Dust spectral slope (α) and the star formation main sequence

Fig. 5.7 shows how galaxies from the four samples populate the M_*-SFR plane, using M_* and SFR results from CIGALE. The so-called star formation main sequence (MS) is usually defined in terms of a positive correlation followed by star-forming galaxies between star formation rate and stellar mass. Both Figs. 5.6 and 5.7 show that the SB and stage 5 galaxies indeed lie in a narrow and relatively high range of SFRs. We have color coded the symbols according to their CIGALE-derived estimates for α, the exponent of the power law defined in Eq. 5.1, which parametrizes the average dust temperature. For the same range of masses, the AGN and SIGS samples extend to lower ($< 10^{-1} M_\odot\,{\rm yr}^{-1}$) SFRs compared to the other two samples.

The bulk of the SIGS galaxies follows the MS over three orders of magnitude in stellar mass, and over four orders of magnitude in SFR. There are some outliers at low SFR, consistent with these being quiescent galaxies. The LSM and SB galaxies also seem to follow the MS, but they are more massive, than the SIGS sample and consequently show higher values of SFR. The AGNs in our sample have masses limited to a narrow range between 10^{10} $M_\odot$ and 4×10^{11} $M_\odot$ and a broad range of SFRs.

The behavior of the AGN galaxies is notable especially when we consider dust temperatures as parametrized by α. For all the other samples, dust temperature positively correlates with sSFR, that is, for a given stellar mass, dust becomes hotter (alpha decreases) as SFR increases. Using simulations and observations of SIGS galaxies, Martínez-Galarza et al. (2016) have shown that this evolution of the dust temperature as galaxies depart the MS is linked to the interaction stage: initially (at early interaction stages) galaxies have low SFRs and relatively cool dust temperatures, but SFR and dust temperature both increase as the systems approach coalescence. This is related to an increase in the *compactness* of the ISM, i.e, the average distance of the dust to the heating sources, normalized by the luminosity of the source.

We observe a similar evolution of dust temperature with distance from the MS for the SB and LSM samples. The AGN sample, however, is different. For AGN sample galaxies the α parameter is completely uncorrelated with the location of the system relative to the MS, and moreover, SEDs compatible with hot dust are observed at very low SFRs. The average dust temperature is therefore not controlled by star formation in AGN-dominated galaxies, and the concept of compactness should be interpreted from a different perspective for these systems.

Perhaps more relevant here is that by disentangling AGN and SF activity we can obtain more reliable SFR estimates for these systems unbiased by the thermal emission

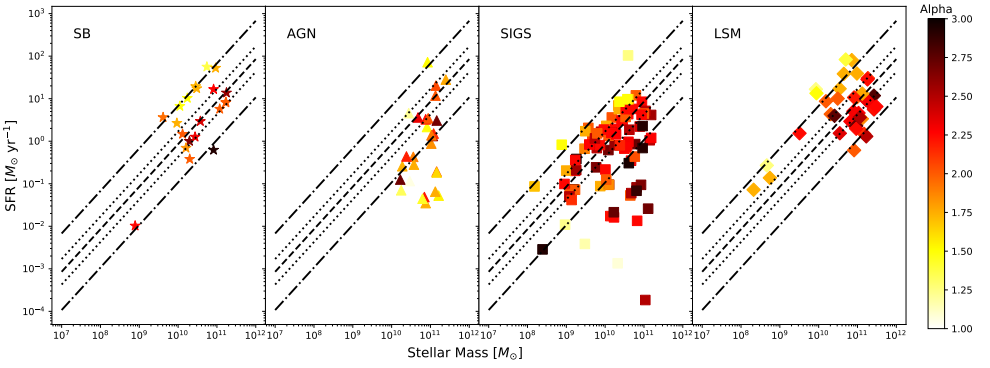

Figure 5.7 – *SFR as a function of stellar mass with selected galaxies color-coded by the derived parameter α (Dale et al. 2014) obtained from* CIGALE. *The dashed, dotted, and dashed-dotted lines in every panel indicate respectively the $z = 0$ MS and the 0.3 and 0.9 dex limits above and below it. We compare the SB (left panel), AGN (middle-left panel), SIGS (middle-right panel) and LSM samples (right panel). All the SB galaxies follow the MS and the gradient of α seems to agree with Martínez-Galarza et al. (2016) with the parameter C, as we discuss in Section 5.5.2. As expected, the AGN sample galaxies do not follow the MS, and no trend with α is evident. The SIGS sample shows a large scatter across the MS and some SIGS galaxies appear in the region where SF is quenched.*

from the AGN. We also corroborate that mergers can be an important factor in contributing to the scatter of the MS, since galaxies move away from the MS as they evolve into later phases of the merger. One additional note has to do with quenching. Although the SIGS galaxies lying below the MS (as indicated by the dot-dashed lines in Fig. 5.7) are most likely quenched systems with small or negligible gas reservoirs, some of the galaxies that we would infer to be actively forming stars might actually be recently quenched systems where the stars formed right before the quenching are still dust-enshrouded, as demonstrated in Hayward et al. (2014). Additionally, the fact that we find AGN systems below the MS suggests AGN activity persists after the quenching, even at very low levels of SF.

Figure 5.8 shows how galaxy morphology affects location within the star-forming main sequence by using the method of de Vaucouleurs et al. (1991), which assigns numerical values to the Hubble stages (usually called T values, de Vaucouleurs 1977). We were unable to classify four galaxies (2XMM J141348.3+440014, 4U 0557-385, LEDA 68751, Mrk 1383), which was unsurprising given that these are among the most distant galaxies in this work. For three galaxies with morphologies not available in de Vaucouleurs et al. (1991) we use the NASA/IPAC Extragalactic Database (NED). The three galaxies classified using NED are ESO 033-02 (a SB0; T= −3), Mrk 841 as (E; T= −5), and MCG-03-34-064 (S0/a and S0+ are both given; we assigned it T= −0.5). Not all galaxies have T values in the range from −6 to 11. Six are classified as Non-Magellanic irregulars (NGC 2820A, NGC 2968, NGC 3034, NGC 3077, NGC 3448, M51B) and three as peculiar irregulars (NGC 2623, NGC 3256, NGC 520). For close mergers (NGC 4038/4039 and NGC 5614/5615) we only use the information of the most prominent galaxy. We were unable to determine morphologies for seven galaxies with NED or a literature search; these object were ESO 141-55, ESO 383-35, Mrk 1502, Mrk 1513, Mrk 335, Mrk 771, and IC 694). They do not appear in Fig-

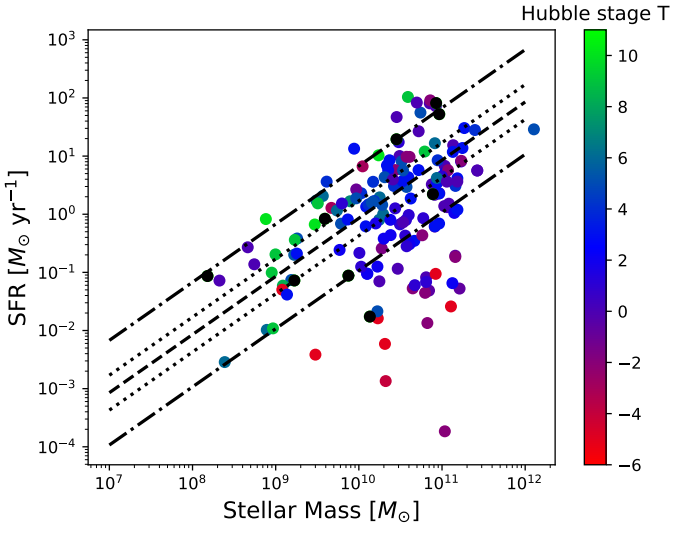

Figure 5.8 – *Main sequence of star-forming galaxies showing the galaxies in this work, but color-coded by Hubble stage (T). Peculiar galaxies outside of range of colors shown for T in the color bar are plotted in black. The dashed lines are identical to those in Figure 5.7. Early types (elliptical and lenticular galaxies; T< 0) tend to have low SFRs while the spirals follow the MS. Irregular galaxies scatter over in several order of magnitude in SFR for a given stellar mass.*

ure 5.8.

Unsurprisingly, most of early type galaxies lie in the quiescent region (below the lowest diagonal line) of the MS diagram, while most of the spiral galaxies follow the MS. No obvious trend is apparent for the irregular galaxies, which scatter widely in SFR for a given stellar mass. This is in harmony with the demographics of disk-dominated ("blue cloud") and spheroid-dominated ("red sequence") systems (Kauffmann et al. 2003b; Brinchmann et al. 2004; Somerville & Davé 2015).

The emission from warm dust is an essential contributor to the SED in most stages of mergers. Martínez-Galarza et al. (2016) found that the compactness parameter C that relates the distribution of dust temperatures with the geometry of the environment is correlated with the sSFR. The position of our galaxies in the MS, (Figure 5.7 and 5.8) supports a picture in which the dust within star-forming galaxies evolves as those galaxies evolve and transform their morphologies.

5.5.3 Comparisons with other AGN indicators

In this Section we compare our fitted $f_{\rm AGN}$ and other CIGALE-derived AGN-related parameters to widely used AGN indicators.

IRAS 60/25 μm and Neon emission-line ratios

The IRAS 60 μm to 25 μm flux ratio f_{60}/f_{25} is an indicator of hot dust content (Wu et al. 2011), and thus suggestive of the strength of the AGN relative to ongoing star

formation because dust in the AGN torus is on average hotter than in star formation regions. In combination with mid-infrared emission-line ratios, f_{60}/f_{25} can be quite effective at separating AGNs from starbursts (Tommasin et al. 2008; Veilleux et al. 2009; Tommasin et al. 2010). In the top panel of Fig. 5.9 the f_{60}/f_{25} ratio is plotted as a function of the [Ne V]/[Ne II] integrated intensity ratio for our galaxies. There is a clear separation between star formation-dominated galaxies ([Ne V]/[Ne II] $\lesssim 0.7$, $f_{60}/f_{25} \gtrsim 3$) and AGNs, in agreement with Tommasin et al. (2010) and Higuera-G. & Ramos P. (2013). The result, together with Fig. 5.4, supports the common assumption that the shape of the mid-IR continuum of galaxies with significant emission from AGNs can be approximated by a power-law and thus that the continuum in this region is a good discriminator between galaxies with and without strong AGN emission (Brandl et al. 2006; Tommasin et al. 2008; Dixon & Joseph 2011). Veilleux et al. (2009, figures 24-26) found a similar relationship with the analogous [O IV]/[Ne II] ratio and less dramatically with the [Ne V]/[Ne II] ratio. They find a progression of these line ratios from low to high starting with star-forming galaxies, followed by Seyferts 2, Seyferts 1, ULIRGs, and finally QSOs.

The SIGS sample unfortunately only possesses a few published measurements or upper limits for these lines. Most of them have ([Ne V]/[Ne II] $\lesssim 0.1$ and $f_{\rm AGN} < 0.2$. We would expect the interacting systems in the SIGS sample to have a different fraction of AGN emission as they move from early on in the interaction towards the coalescence phase, to fall in between the two regimes of [Ne V]/[Ne II] ratio presented here.

We observe that those SIGS galaxies for which we were able to collect line emission and that are not upper limits, do fall in between the SB and the AGN galaxies, with intermediate cases. The AGN galaxies with the highest *IRAS* ratio (cooler dust) are NGC 3281, ESO428-14, NGC 4941, NGC 4388 and NGC 7674, and the SB with the most elevated [Ne V]/[Ne II] are NGC 2623, NGC 1365, NGC 4088, NGC 4194 and NGC 4676. There is one upper limit published for a LSM galaxy (2MASX J10591815+2432343, with [Ne V]/[Ne II]$<$ 0.02), so we do not include that galaxy in Fig. 5.9.

Comparing galaxy parameters with their emission line ratios

The fine structure Ne-lines can help us discriminate between the SB and AGN samples as well as with the SIGS galaxies because they flag the presence of an AGN even in galaxies otherwise classified as a star-forming (Abel & Satyapal 2008). One of the most useful outputs of CIGALE is the fraction of AGN derived from the SED (Sect. 5.4.1). The bottom panel of Figure 5.9 offers confirmation of the CIGALE estimated AGN contributions, compared with other estimations not using SED from Veilleux et al. (2009, table 12) and Díaz-Santos et al. (2017, table 2, column 7). We can see that the AGN sample and SB sample separate very well in Figure 9. However, the estimated AGN fraction can in some cases have a value near to 40% in the SB sample galaxies (NGC 1365 and NGC 660). In the same way, three AGN galaxies have a AGN fraction below 0.3 (NGC 4941, NGC 7674 and NGC 4388).

Most of the SIGS sample galaxies have weak or no AGN contributions (Lanz et al. 2013). A particular outlier of this behaviour is NGC 3034, with the highest $f_{\rm AGN} = 0.48 \pm 0.03$ in the sample but yet a very low [Ne V]/[Ne II]. Lanz et al. (2013) show that

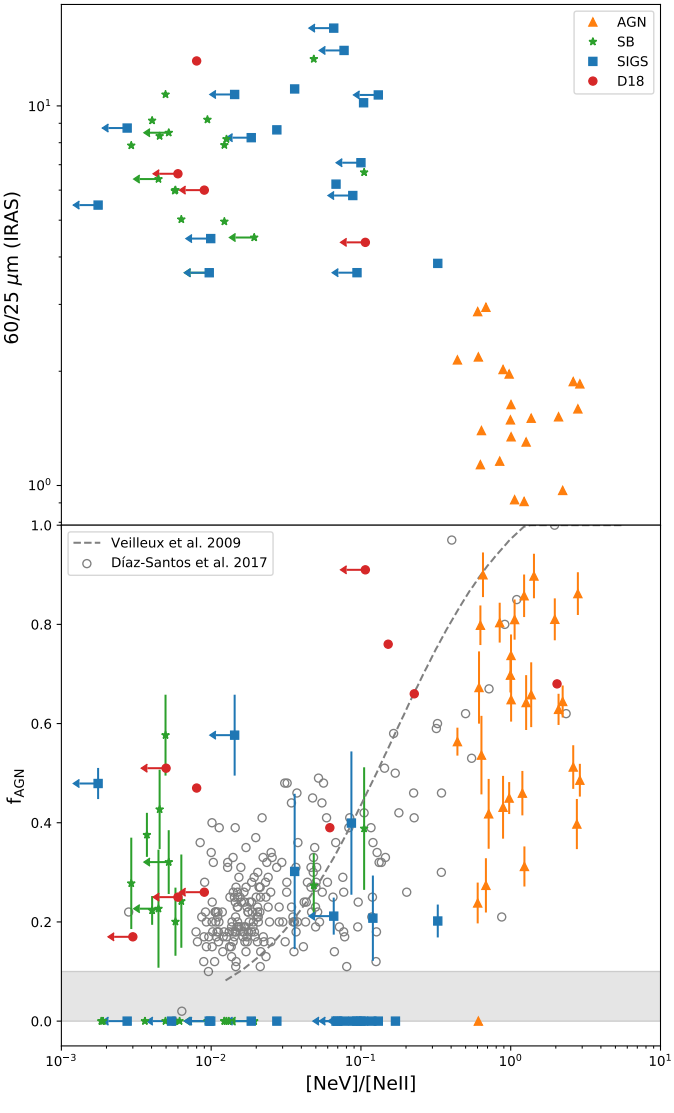

Figure 5.9 – *Top panel: The ratio of integrated IRAS flux densities at 60 and 25 μm f_{60}/f_{25} as a function of the [Ne v]/[Ne ii] integrated line intensity ratio. The symbols are the same as in Figure 5.5 with the addition of values from D18. Upper limits are indicated with arrows. Bottom panel: Total AGN fraction as a function of the [Ne v]/[Ne ii] integrated line intensity ratios. The markers are the same as in the top panel, with the errorbars on $f_{\rm AGN}$ indicating the uncertainty estimates taken from* CIGALE. *We added the estimations from Veilleux et al. (2009) (dashed line) and Díaz-Santos et al. (2017) (empty gray circles) to compare with literature results. The grey region indicates where the $f_{\rm AGN}$ estimates are not believed to be reliable; we treat objects in this region as if they have $f_{\rm AGN} = 0$. In both panels the SB and AGN samples are well-separated. The SIGS galaxies tend not to have significant AGN contributions.*

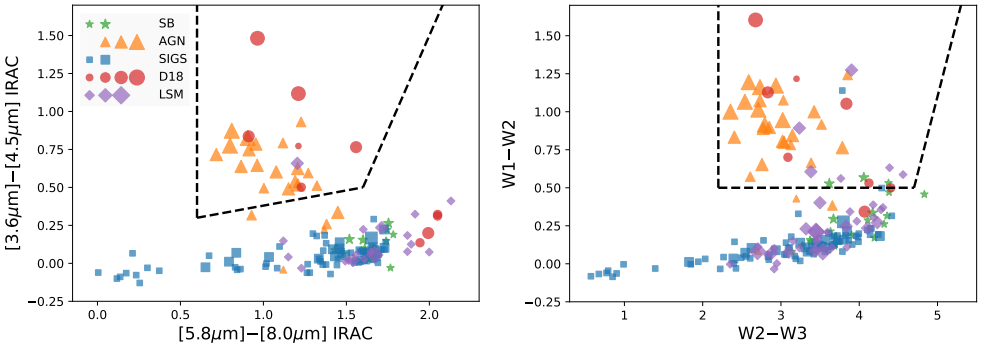

Figure 5.10 – *Mid-infrared color-color classification diagrams showing the SIGS (blue squares), SB (green stars), LSM (purple diamonds) and AGN (orange triangles) galaxies together with late-stage mergers (red circles) from D18. The markers sizes are divided in four ranges for the $f_{\rm AGN}$ <0.25,0.25-0.5,0.5-0.75 and >0.75, where not all the samples cover the four ranges. Left panel:* $[5.8]-[8.0]$ *vs.* $[3.6]-[4.5]$ *color-color diagram following Stern et al. (2005) for all sample galaxies with available IRAC photometry. Sources in the region enclosed by the dashed lines (commonly referred to as the Stern wedge) are dominated by AGN emission in the mid-infrared, while sources below this region are dominated by star-formation. Right panel:* $W2-W3$ *vs.* $W1-W2$ *color-color diagram, with an AGN wedge (dashed lines) per Blecha et al. (2018). Most late-stage mergers and AGN fall in this wedge with some galaxies from the SIGS and SB sample. The separation of these wedges with the* $f_{\rm AGN}$ *is in agreement with the expected behaviour of AGN activity onset.*

this galaxy was very difficult to fit with MAGPHYS (da Cunha et al. 2008), possibly due to the high obscuration, the presence of an outflow, or some other unaccounted for activity. We were also unable to obtain a good fit with CIGALE except when we include the AGN component model; in that case our reduced-χ^2 was 3.29, which is low enough to be considered a reliable fit. Therefore, even when strong star formation is present, $f_{\rm AGN}$ estimates with CIGALE could reveal a hidden AGN, that is invisible in optical wavelengths. This estimate can only be checked by the emission spectra of high IP lines like [Ne V].

Infrared color diagnostics

Infrared colors are well-known diagnostics of the energy sources powering infrared-luminous galaxies; two salient examples are the color-color diagrams developed by Stern et al. (2005) and Lacy et al. (2004) to discriminate between galaxies dominated by star formation and AGN emission. More strict criteria can be applied (e.g. Donley et al. 2012), but they depend on other factors, as luminosity. We can use our more precise AGN and SFR measurements to test the reliability of these diagrams. Fig. 5.10 shows *Spitzer*/IRAC and *WISE* color-color diagrams for our galaxies. Galaxies that lie within the wedge enclosed by the dotted lines in left panel are expected to be AGN-dominated. Only five of the 29 galaxies from our AGN sample (ESO428-14, NGC 3516, NGC 4941, NGC 4388, IC5063) lie outside the wedge. This is not unexpected: Petric et al. (2011) find that faint AGNs with measurable PAH 6.2 μm EW fall outside the wedge.

AGN- and star formation-dominated galaxies are also efficiently segregated in a com-

plementary *WISE* color-color diagram (Fig. 5.10, right panel). All but three of our AGN sample galaxies lie in the AGN selection region described by Blecha et al. (2018); two of the outliers are located close to the boundary. However, some SB populate the AGN wedge in this plot. We include the late-stage mergers from (D18) in Fig. 5.10. Many of the (D18) late-stage mergers have weak or undetected AGNs; they appear to populate both the AGN- and star formation-dominated regions of both panels of Fig. 5.10.

Stern et al. (2012) proposed that *WISE* $W1-W2 \geq 0.7$ colour is a robust indicator of AGN emission. The majority (22) of our 29 AGN sample galaxies meet this criterion. A less conservative $W1-W2 \geq 0.5$ colour cut (Ashby et al. 2009) is similar to the lower boundary of the wedge in the left panel of Fig. 5.10, and is identical to that shown in the right panel. The two AGN galaxies which fail to satisfy the less stringent criterion are ESO 428-14 and NGC 4388. Our most-likely CIGALE models for them yield total AGN fractions of $f_{\rm AGN} = 0.31 \pm 0.04$ and 0.24 ± 0.04 respectively, which are significant although not large enough to make them AGN-dominated, and help indicate the reliability limits of these diagrams.

NGC 4941 is the only AGN having a blue $W1-W2$ colour comparable to those of the SIGS galaxies, with $f_{\rm AGN}<20\%$ yet also possessing a high [Ne V]/[Ne II] ratio (Sec. 5.5.3). Its blue mid-IR color and high Ne line ratio is consistent with NGC 4941 being a heavily absorbed low-luminosity AGN (Kawamuro et al. 2013) and illustrates how the [Ne V] emission can help identify the AGN contribution in highly obscured cases. Overall, we confirm that mid-IR color diagnostics in general do identify AGNs, and with $f_{\rm AGN}$ we quantify their contribution to the total galaxy output.

Some of the SB and SIGS galaxies lie close to the AGN wedge in the right-hand panel of Fig. 5.10. We decided to test whether their location in Fig. 5.10 could be interpreted straightforwardly to mean that they are composite objects hosting significant AGN and star formation but not necessarily dominated by either. We examined galaxies having intermediate *WISE* colours $W1-W2 > 0.3$. This included five SIGS galaxies (NGC 838, NGC 839, NGC 3034, NGC 3227 and NGC 3690) and eight SB galaxies (NGC 660, NGC 1222, NGC 1365, NGC 1614, NGC 2146, NGC 2623, NGC 3256 and NGC 4194).

We consider the SIGS galaxies first. For NGC 838 and 839 the most likely CIGALE fits give $f_{\rm AGN} \sim 0.0$ for both galaxies. B15 classify NGC 839 as a low-ionization nuclear emission-line region (LINER, e.g., Kewley et al. 2006). The most likely $f_{\rm AGN}$ estimate for NGC 3690 is 0.0, but D18 notice different classification for this galaxy, as LINER, AGN and star-forming. The optical spectroscopic classifications (Pereira-Santaella et al. 2010) for NGC 3034 and 3227 are HII and Seyfert 1, these galaxies have an SED that is consistent with an AGN contribution of $f_{\rm AGN} = 0.48 \pm 0.03$ and 0.20 ± 0.03, respectively.

Next we consider the SB galaxies. For NGC 660 we estimated $f_{\rm AGN} = 0.43 \pm 0.08$. This object is usually classified as star-forming (Petric et al. 2011; Wu et al. 2011), and there are also signs of interaction. NGC 1365 has one of the highest [Ne V]/[Ne II] ratios in the SB sample. Its calculated $f_{\rm AGN}$ estimate is 0.39. Hernán-Caballero et al. (2015) obtained 0.6. NGC 1614 has $f_{\rm AGN}$=0.0; Hernán-Caballero et al. (2015) obtained 0.3. This object is classified as an "uncertain AGN" by Asmus et al. (2014).

NGC 1222, NGC 2146 and NGC 3256 have $f_{\rm AGN} \sim 0.0$, 0.37 and 0.0, respectively. Finally, NGC 4194 has a high ratio of [Ne V]/[Ne II] and seems to be undergoing a merger, but the CIGALE results show a $f_{\rm AGN} = 0.0$, consistent with there being no AGN contribution. It is nonetheless classified as a Seyfert 2 galaxy (Pereira-Santaella et al. 2010).

Lastly, for NGC 2623 the [Ne V]/[Ne II] ratio is among the highest in the SB sample (see Sect. 5.2.4), with $f_{\rm AGN} = 0.27 \pm 0.06$. Unfortunately, we obtained a relatively poor fit to the SED of this source, so our CIGALE model for it may be unreliable. It has been classified as starburst (e.g. Wu et al. 2011), composite (Asmus et al. 2014) and AGN (e.g. Hernán-Caballero et al. 2015) where a fraction of AGN of 0.44 is obtained, comparable to the AGN fraction of 0.39 ± 0.05 from D18. These are intermediate cases in which both an AGN presence and also intense star-formation are underway, and color alone is insufficient to categorize the source unambiguously.

The last sample we examine in Figure 5.10 is the LSM. Twelve galaxies are above the cut at $W1 - W2 > 0.3$ and six of them with values of $W1 - W2 > 0.5$. Galaxies 2MASX J01221811+0100262 and 2MASX J08434495+3549421 have the highest $f_{\rm AGN}$ of the LSM sample, 0.42 ± 0.10 and 0.48 ± 0.10 respectively. The former was found to be a tentative dual AGN with mixed signs of star-formation by Satyapal et al. (2017). The latter is classified as Seyfert 2 by Véron-Cetty & Véron (2010). The rest of the galaxies have values for $f_{\rm AGN}$ below $\lesssim$0.25.

In summary, these five SIGS and eight SB galaxies that lie close to the AGN wedge in Fig. 5.10 appear to be composite systems, and the LSM systems inside the wedge are classified as AGN as one would expect: the CIGALE models indicate varying fractional contributions of AGNs to their emission, but in none of them there is a dominant AGN that was somehow 'missed' by the color-color diagram. We can identify systems like the ones presented here, which do have AGN contribution, but not large enough to make the Stern cut, and quantify the amount of the contribution in terms of $f_{\rm AGN}$. Likewise, our SED analysis is consistent with the implications of the canonical infrared color-color diagnostic diagrams.

sSFR and stellar mass estimates in light of prior results

B15 and Lanz et al. (2013) use the MAGPHYS SED fitting code (a version that did not include an AGN component) plus DECOMPIR (Mullaney et al. 2011) to characterize the SIGS galaxies. In Fig. 5.11 we compare our results with those from B15, focusing on the stellar mass and specific star formation rate, parameters only indirectly influenced by the presence of an AGN.

The left panel of Fig. 5.11 shows that our stellar mass estimates agree on average with those from B15, but with a large scatter. No obvious dependence on the total AGN fraction is apparent, which can be related to the selected sample. SED models with AGN and without AGN reproduce similar stellar masses and specific star formation rates, although objects with high AGN fraction show a small trend for being outliers in stellar mass. IC 694 is a conspicuous outlier because of aperture issues (Sec. 5.3.4). Even more scatter is observed in the comparison for derived sSFR (right panel of Fig. 5.11), especially below log(sSFR)= −11.5, where the accuracy

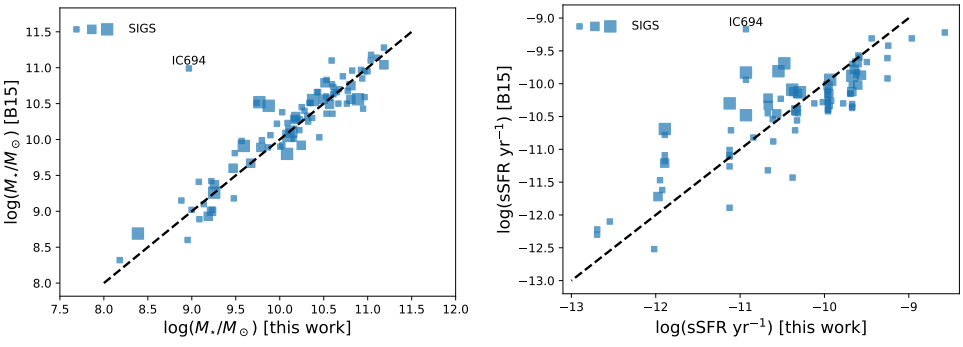

Figure 5.11 – *Stellar mass (Left panel) and specific star formation rate (Right panel) estimates derived in this work compared to those obtained with MAGPHYS by B15 for the SIGS galaxies. Symbol sizes indicate the total AGN contributions $f_{\rm AGN}$: the smallest symbols indicate an AGN contribution below 20%, the intermediate symbols indicate an AGN contribution between 20 and 30%, and the largest symbols indicate an AGN contribution above 30%. The line of equality is indicated with the dashed line in both panels. The stellar masses are in good overall agreement, with large scatter. The sSSR estimates also show approximate overall agreement, but also with significant scatter, especially below log(sSFR)= −11.5.*

of SED-based methods in measuring the sSFR significantly decreases (Eales et al. 2017). The coarse CIGALE input parameter grid below log(sSFR)= −11.5 produces the entirely artificial grouping of galaxies along discrete sSFR values. NGC 4933A, NGC 5353 and NGC 5481 are not shown because their estimated sSFRs fall in an extremely low, likely unreliable, sSFR regime ($< 10^{-13}$).

5.6 Discussion

In this Section we discuss the implications of the SED modeling for star formation and AGN emission in interacting galaxies.

5.6.1 Interaction Stage

B15 note that almost all their galaxies with log(sSFR) < −11.0 are at early interaction stages. Morphologically disturbed systems lie along a broad range of sSFRs, with the range occupied by stage 4 galaxies extending to higher sSFRs than the earlier stages (Fig. 5.6). Our results are equivocal with respect to stage 4 systems and sSFR, however, because the KS tests applied in Sec. 5.5.1 do not provide compelling evidence, even with our enlarged sample, that stage 4 systems differ significantly from stage 2 or 3 systems (Table 5.3). However, the outcome is different for the stage 5 systems. They differ significantly in the aggregate from the AGN sample, marginally from the stage 2 and 3 systems (7% and 14% chance of being drawn from the same parent sample, respectively), and there is no evidence that they differ from the starburst systems. Collectively, the evidence thus favors a picture in which sSFRs are greatest in stage 5 systems, i.e., in or approaching coalescence. This is in agreement with hydrodynamical simulations (Hayward et al. 2014; Martínez-Galarza et al. 2016), which show a steep increase of the SFR very close to coalescence (more for massive

systems).

As discussed in Sect. 5.5.1, Lanz et al. (2013) did not find statistically significant correlations between SED shape, merger stage, and star-forming properties. In the present work we have enlarged the study sample and the available photometry, and analyzed these data with CIGALE, i.e., a code that explicitly accounts for AGN emission. With these enhancements, we observe that $f_{\rm AGN}$ does show a correlation with luminosity (see Sect. 5.6.4). Our results also point to a weak correlation between $f_{\rm AGN}$ and interaction stages, with a larger fraction of late stage mergers showing a higher $f_{\rm AGN}$. SED analysis can therefore be used to infer the physical conditions associated with different stages.

5.6.2 The Schmidt-Kennicutt Relation

Star-forming galaxies form a relatively narrow distribution in the two-dimensional parameter space defined by total stellar mass and SFR, commonly known as the star-forming main sequence (MS). This scaling relation has been widely used to study the relationship between galaxy morphology, star formation, and SED shapes (e.g., Elbaz et al. 2011). However, the SFR is not a directly measured quantity, as it is indirectly derived from different observables; total stellar mass estimations, although generally robust, also suffer from being model-dependent. As an example, the vast majority of papers discussing the MS rely on a version of Schmidt-Kennicutt (S-K) relation to infer SFRs from the infrared luminosity of galaxies, that is first converted to a dust mass, and then, via a gas to dust ratio, to a gas mass. In this sub-section we re-examine the reliability of that relation for LIRGs. For the analysis, we rely on parameters derived from our full CIGALE modeling of the SEDs.

In order to estimate the obscured SFR, the 8 μm or 24 μm luminosities are often used (see Lanz et al. 2013, for a more detailed discussion), but when possible, it is convenient to use the integrated infrared luminosity between 5–1 000 μm, which is related to the thermal emission from dust heated by star formation (at wavelengths shorter than about 5 μm the SED is dominated by emission from stellar photospheres rather than dust heated by star formation). The Schmidt-Kennicutt relation, often formulated as the relationship between gas surface density and SFR, can also be formulated as a relationship between the total SFR and the infrared luminosity (Lanz et al. 2013). This has been a very useful SFR diagnostic ever since the infrared was first made accessible by the *IRAS* and *ISO* satellites. With panchromatic datasets now at hand, incorporating photometry from *Spitzer*, *WISE*, *AKARI*, and *Herschel*, there is room for considerable improvement.

An important caveat in this conventional approach is that the infrared emission is often interpreted as originating from thermal emission from dust heated only by star formation. However, other heating mechanisms are often in place and need to be accounted for, such as AGN activity and older stellar populations. (Hopkins et al. 2010; Lanz et al. 2013; Wilson et al. 2019). Although for most star-forming galaxies this additional contribution is small or negligible, in some cases it can be significant or even dominate the infrared luminosity of the entire galaxy. For the galaxies included in this work, which by construction emphasize infrared-luminous systems and mergers at various stages, we show that the AGN alone can contribute up to $\sim$ 80% of the

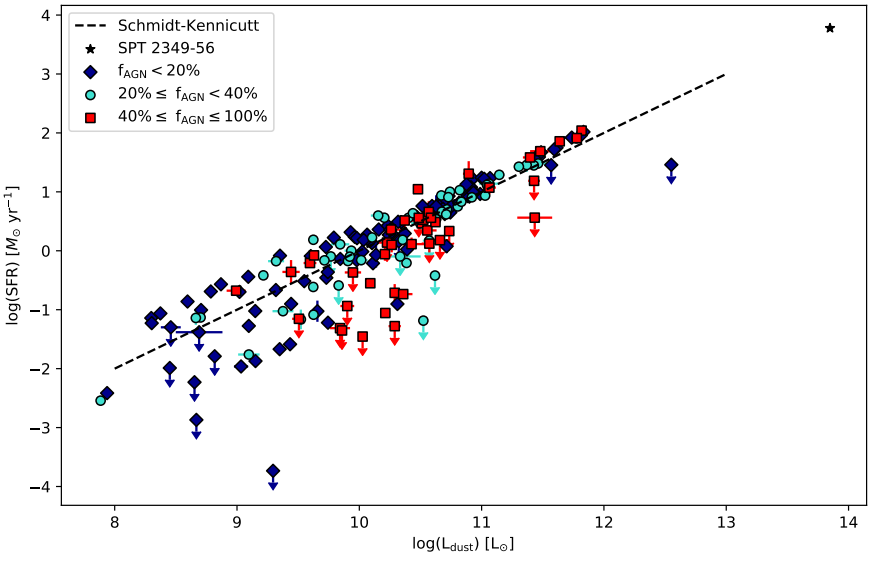

Figure 5.12 – *Total derived SFR as a function of total dust luminosity for all galaxies modeled in this work. Red squares, cyan circles, and blue diamonds respectively indicate galaxies with total AGN fraction > 40%, in the range 20–40%, and below 20%. The dashed line is the S-K relation. Also plotted for comparison is an IR-luminous cluster SPT2349−56 at z = 4.6 (black star, Miller et al. 2018). Arrows indicate the galaxies for which the SFR are only estimated to within a factor of roughly two (see Table 5.13).*

infrared luminosity for these systems.

Figure 5.12 plots derived SFR as a function of dust luminosity for all galaxies in Tables 5.4–5.7, color-coded by the estimated $f_{\rm AGN}$. We observe that for dust luminosities above roughly $10^{10.5}$ $L_\odot$, the S-K diagnostic provides a good measure of the amount of star formation taking place. In particular, the fact that at those luminosities there is so little scatter at all levels of $f_{\rm AGN}$ indicates that our approximation correctly accounts for other sources of dust heating, allowing a better estimation of the SFR. At lower luminosities we observe a significant scatter in the SFR at a given luminosity. This indicates that the S-K diagnostic might not provide a reliable measure of the SFR at these lower dust luminosities. A plausible explanation for this is that low dust luminosities also implies a larger relative amount of unobscured star formation that is not accounted for by the infrared diagnostics. Less luminous, less morphologically disturbed systems are less optically thick to UV radiation, and therefore a pan-chromatic approach such as the one we have adopted here is more likely to provide reliable estimates.

Figure 5.12 also indicates that galaxies that are less luminous in the infrared have a broader range of AGN contributions skewed towards smaller contributions, i.e., very few galaxies with infrared luminosities below 10^9 $L_\odot$ have $f_{\rm AGN} > 20\%$. This supports an scenario in which significant AGN emission occurs preferentially in highly disturbed/obscured systems, and is in agreement with hydrodynamical simulations. Finally, our results are also consistent with a wider dynamical range of SFR at lower

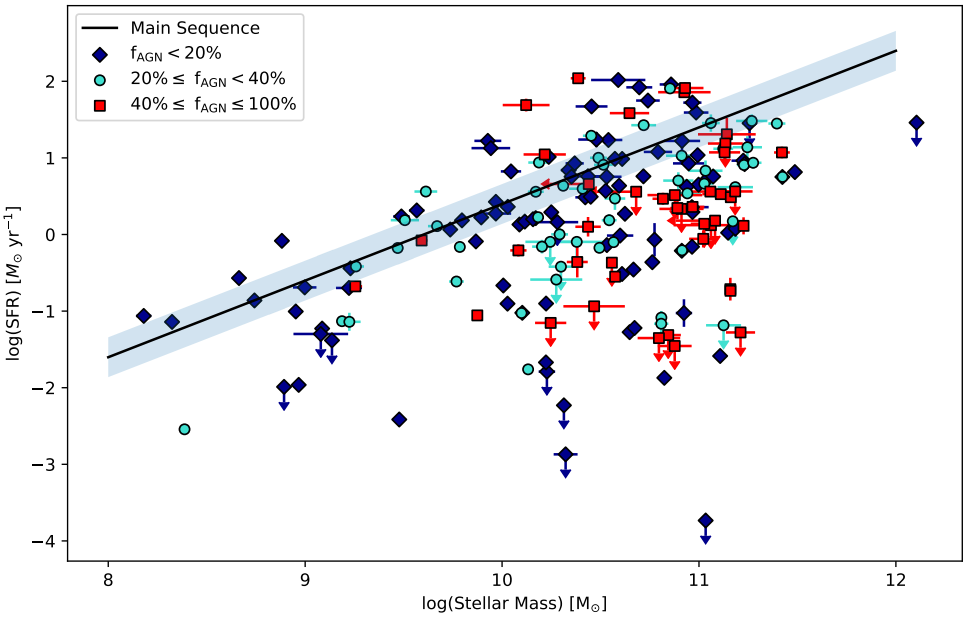

Figure 5.13 – *Integrated derived SFR as a function of total stellar mass populated with all galaxies modeled in this work. Red squares, cyan circles, and blue diamonds respectively indicate galaxies with total AGN fraction > 40%, in the range 20–40%, and below 20%. The star forming main sequence per Elbaz et al. (2011) is indicated with the black line; the shaded region extends to ±0.26 dex about it. Values of SFR below −1 in the log are relatively uncertain. Arrows indicate the galaxies for which the SFR or stellar masses are only estimated to within a factor of roughly two (see Table 5.13)*

luminosities, in absolute terms. At low luminosties, star formation can range from being a negligible to being a significant factor in the galaxy evolution, but at high infrared luminosities, the dust heating from SFR and AGN completely dominates the galactic evolution. This also has to do with the timescales of this evolution. Luminous systems are morphologically disturbed with starburst-like, short-term episodes of star formation and AGN accretion, whereas low luminous systems evolve more secularly, impacting the range of possible fractional SFR contributions.

5.6.3 The Galaxy Main-Sequence: New Subtleties and Issues

In Fig. 5.13 we plot the CIGALE-derived stellar masses and SFRs for our galaxies, together with the location of the MS as derived by Elbaz et al. (2011) for comparison. Several interesting inferences follow from the way our sample galaxies populate this two-dimensional space. First, only a minority of galaxies are located on the nominal MS locus. Some lie above it, in the zone associated by Elbaz et al. (2011) with starburst-like star formation, whereas a significant amount lie below it, even by a few orders of magnitude. At high stellar masses, the majority of those systems above the MS have higher AGN contributions, which indicates that not only the SFR per unit stellar mass in enhanced, but also the AGN activity.

What is more puzzling is the significant amount of systems that we observe below the MS. Significant divergences from the MS have been reported both in observations of high redshift galaxies within protoclusters, due to environmental quenching (e.g. Zavala et al. 2019), and in cosmological simulations that relate the growth of galactic halos to that of stellar mass, in which the MS scatter depends on the timescale of star formation variability (e.g. Hahn et al. 2019).

In our case, the large scatter is probably due to the way we assembled our sample. We have selected preferentially galaxies that are luminous and that are morphologically disturbed through mergers. In some of these cases, it is impossible to tell from the morphology alone whether the system has undergone coalescence, and it is likely that in some of those systems star formation has been suppressed due to negative feedback from the AGN, after coalescence. This interpretation is consistent is supported by the results shown in Figure 5.8, with early type galaxies showing significantly lower SFRs. Our sample is therefore not representative of the secular stages of star formation in galaxies. Instead it represents systems with enhanced star formation through the effect of mergers, and systems where AGN feedback has probably quenched star formation. The fact that both high and low $f_{\rm AGN}$ values are similarly represented below the MS indicates that quenching takes place very raplidly after the onset of the AGN. In an upcoming paper (Della Costa et al. in prep.), we discuss this latter conclusion in more detail.

In Fig. 5.14 we compare the stellar mass to the total infrared luminosity from dust heated by stars and AGN in the left panel and to the AGN luminosity only in the right panel. We note that out of a total of 188 galaxies, 42 galaxies have $f_{\rm AGN} \geq 40\%$ whereas 51 of them have $f_{\rm AGN} \geq 20\%$. We observe a mild correlation between stellar mass and luminosity for systems with low contribution from AGN. Presumably, in these systems the infrared luminosity is dominated by star formation, and the correlation confirms that more massive systems tend to have more dust heating, but not always and there are wide variations. A similar correlation is found for systems with a significant contribution form the AGN, but notably the scatter is much smaller. For galaxies with AGN, we also observe a correlation that implies that most luminous AGNs tend to love in the most massive galaxies regardless of the fraction of total luminosity that the AGN contributes, as long as it is above 20%. The apparent scatter for systems above $10^{10} M_{\odot}$ is most likely due to larger uncertainties in the determination of stellar masses for these systems. In the context of galaxy assembly, this positive correlation supports a joint evolution of super massive black holes and their hosts.

Fig. 5.15 shows the correlation between SFR and AGN luminosity for those systems with $f_{\rm AGN} > 20\%$. We observe that systems with intermediate AGN contributions (cyan dots) show a tight correlation in this plane over 5 orders of magnitude in SFR. For systems with high AGN contributions (red dots), we observe a similar correlation, but there is a larger scatter, and also, the luminosity of the AGN at a given SFR is higher in comparison with the intermediate systems. Similar correlations between AGN lumonosity and SFR have been found in more uniform samples, such as the COSMOS field (Lanzuisi et al. 2017); such correlations support scenarios proposed in recent galaxy evolution models in which black hole accretion and star formation are correlated due to the compression of large amounts of gas in nuclear regions.

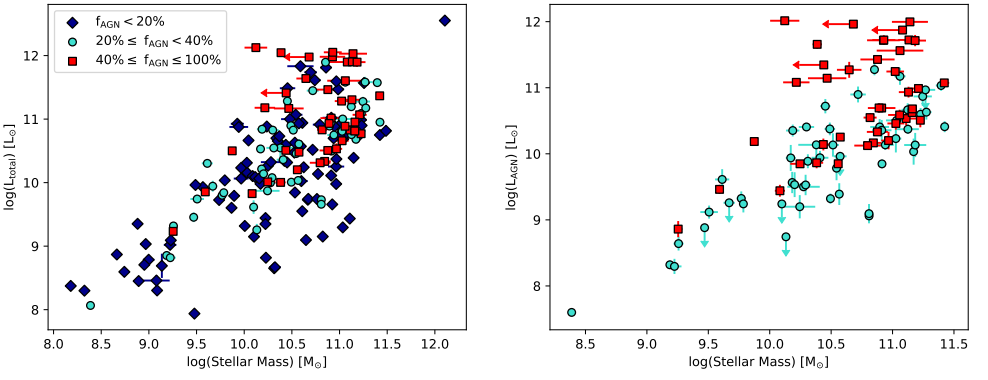

Figure 5.14 – *Left panel: The total dust luminosity as a function of stellar mass. Note that sources with a* CIGALE*-estimated AGN fraction below 20% are recalculated and plotted here assuming its AGN contribution is zero. Right panel: The AGN luminosity only as a function of stellar mass. Arrows indicate the galaxies for which the luminosities or stellar masses are only estimated to within a factor of roughly two (see Table 5.13)*

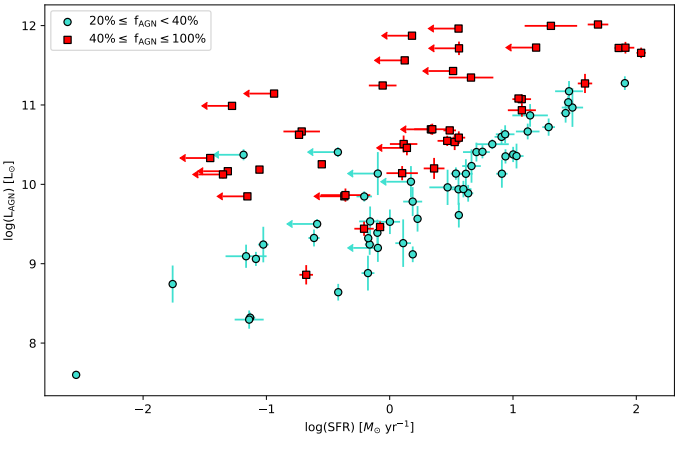

Figure 5.15 – *Total AGN luminosity as a function of SFR. Red squares indicate objects with* $f_{\rm AGN} > 40\%$*, whereas cyan dots indicate those with* $f_{\rm AGN}$ *in the range of 20–40%. Arrows indicate the galaxies for which the luminosities or SFR are only estimated to within a factor of roughly two (see Table 5.13)*

The fact that the correlation is less tight at higher AGN contributions, at evolutionary stages closer to coalescence, can be interpreted in terms of star formation quenching: as hydrodynamical simulations show, the AGN reaches a peak in luminosity right after coalescence, and star formation gets quenched very rapidly. This supports a real effect of AGN feedback on SFR, as opposed to recent studies (Harrison et al. 2021) that suggest that AGN activity does not quench galaxy wide AGN. One possible explanation of this discrepancy could be the difference between galaxy-wide star formation and the nuclear, merger-induced star formation that we are measuring in the present work.

5.6.4 The AGN as crucial ingredient in galaxy evolution

We have shown (e.g., Figs. 5.13-5.15 and related discussions), that accounting for the AGN emission is a necessary step in order to gain a better understanding of the physics and energy budget of infrared-luminous galaxies, specially in the late stage of mergers. In particular, we have provided evidence that dust heating by the AGN can be a dominant factor in the galaxy SED, and that the latter can provide hints as to the specific stage of the merger. We have also shown that the rapid evolution of AGN accretion an SFR right before and right after coalescence creates significant spread in the so-called Main Sequence, partly due to star formation quenching. This conclusion echoes that by earlier works (e.g., Alonso-Herrero et al. 2012; Hayward et al. 2014). Here we elucidate some implications of these findings.

Alonso-Herrero et al. (2012) suggest that the prevalence of galaxies with $f_{\rm AGN} > 0.05$ correlates with IR luminosity, increasing from $f_{\rm AGN} \sim 0.3$ for $10^{11}{\rm L}_{\odot} < {\rm L}_{\rm IR} < 10^{12}{\rm L}_{\odot}$ to $f_{\rm AGN} \gtrsim 0.5$ in ${\rm L}_{\rm IR} > 10^{12}{\rm L}_{\odot}$. Our Figure 5.5 points to a similar trend. They also found that the AGN contribution is close to 50% for LIRGs, and that it reaches 80% in the case of ULIRGs. Our results are consistent with those numbers, with galaxies in the AGN sample reaching the highest values of fractional AGN contribution. We also see an increase in the fractional AGN contribution for bins of increasing luminosity, For bins centered at 10^{9}, 10^{10}, 10^{11}, 10^{12} in ${\rm L}_{\rm IR}$, the respective means in $f_{\rm AGN}$ values are 5%, 14%, 20% and 36%. Similar results were found by Dai et al. (2018) in a larger sample of U(LIRGS) with a measured AGN contribution. Including the AGN contribution in the SED analysis of luminous infrared galaxies is therefore a required step in any reliable study of their physical properties.

There are a number of diagnostics that can be better interpreted if a reliable determination of the merger or AGN onset stages can be made from SED analysis, as we are doing in this work. For example, Tommasin et al. (2010) notice that if the AGN ionising continuum is switched off, the photoionized narrow line region (NLR) could still be detected due to its large extension and long line recombination time ($\sim$ 300yr, depending on the density). They also suspect that this delay might be related to the existence of many different types of AGNs as classified by their line emission properties. Our work provides a sample that can be used as a proxy to study how line emission properties change as a function of AGN fractional contribution and amount of star formation quenching. A combination of $f_{\rm AGN}$ estimations in addition to fine structure lines ratios (as Fig. 5.9) can lead to a reliable analysis of AGN stages and their relation with interacting galaxies, as with the $f_{\rm AGN}$-[Ne V]/[Ne II] plane. We

have confirmed that line ratios are a more reliable proxy for AGN presence that the SED analysis alone.

Funneling of large amounts of gas into the nuclear regions during a merger can trigger large episodes of star-formation and AGNs activity (Weaver et al. 2010; Lanzuisi et al. 2017). The resulting obscuration can significantly attenuate the AGN optical emission (Blecha et al. 2018). Our results support a co-evolution of star formation and AGN activity during a merger, while correcting by obscuration effects by using a pan-chromatic, energy-conserving approach.

In their study of post-starbursts, Alatalo et al. (2017) find that, well before coalescence, merging galaxies are generally located in the "green valley" and show bluer $W1 - W2$ *WISE* colours, characteristic of AGN activity (see also our Fig. 5.10). These galaxies thus may contain buried AGNs that emit in the infrared and are better traced by infrared emission lines (see Figure 5.9). They suggest that the AGNs do not radiatively dominate the post-starburst phase, and that a better census of these post-starbursts can be constructed if there are reliable tracers of the AGN activity during the early phases of quenching. Our work provides an example of such search, and the fact that we see a significant number of less luminous AGNs below the main sequence in Fig. 5.7 agrees with their results.

The need for more extensive samples and better indicators of the interaction stage specially during the obscured and morphologicallly disturbed phase is critical for an improved understanding of the evolution of mergers. Our work provides a pilot study of what JWST and the SPace IR Telescope for Cosmology and Astrophysics (SPICA) (Spinoglio et al. 2017) will be able to do with their improved sensitivity and spectral resolution.

5.7 Conclusions

We have re-reduced and re-analysed photometric observations from the UV to the FIR on 199 luminous and ultra-luminous galaxies in four different sets of objects including mergers and AGN, analyzed their physical properties using the CIGALE SED modeling code, and presented an analysis of the results. This is the largest systematic, wide, multi-band SED analysis program yet done on an ULIRG sample. In particular, our approach included galaxies over a broad range of AGN activity as reported in the literature. Our goal was to accurately measure the fractional AGN contribution to the total luminosity in these systems, and to assess how this contribution impacts popularly used SED diagnostics of star formation and ISM properties. We also aimed at examining the evolving effects of AGN activity across the merger sequence. From the original sample, we excluded ten objects that either had limited or uncertain datasets and/or unreliable SEDs. Our primary conclusions apply to the remaining 189 objects. Here are our major findings:

1. A reliable measure of the fractional contribution of AGN emission to the total luminosity of galaxies is essential in the understanding of galaxy-wide physics, such as star formation evolution and total energy output. About half of the galaxies in our sample have more that 20% AGN contribution to their total

luminosities, and about a quarter of the systems had contributions over 40%. This results in warmer dust temperatures that can be wrongly associated to star formation if AGN is not included in the modelling. Overall, we find only a weak correlation between the merger stage and the AGN fractional contribution, in agreement with other studies (Lanz et al. 2013, e.g.,).

2. AGN activiy can be responsible for a significant displacement of galaxies across the so-called "Main Sequence" of star formation. Outliers of this correlation must therefore be interpreted in terms of their AGN activity, and not only in terms of their star formation properties. We have produced a carefully remeasured SFR-M_* plane that shows significant deviations from the MS correlation, both above and below it. These deviations are only partially explained by increased star formation, as increased AGN activity and feedback-driven star formation quenching can have a significant role in the emission properties, specially at merger stages just before and just after coalescence.

3. As a result of the previous statement, infrared galaxies at intermediate and high redshifts should not have their physical properties, specially those related to star formation, interpreted as if they were local MS infrared galaxies, without first accounting for their merger stage and AGN fractional contribution. Possible diagnostics to do this from their SED and spectra include their location in the $f_{\rm AGN}$-[Ne V]/[Ne II] plane.

4. At high (>40%) fractional AGN contributions, both the star formation luminosity and the AGN luminosity independently correlate with the total stellar mass of the galaxy. This is in agreement with findings in large uniform surveys such as COSMOS, and supports scenarios in which both black hole accretion and star formation are driven by gas compression in the nuclear regions during the merger. The lack of correlation between the total luminosity and the stellar mass at low AGN fractional contributions calls into question the use of infrared diagnostics alone to estimate SFR in the early stages of mergers.

5. The SFR-M_* plane for our 189 luminous galaxies reveals significant outliers from the Main Sequence, specially among systems larger than about 10^{10}M$_*$. In particular, many systems in late stages of the merger fall up to a few orders of magnitude below the MS. This suggests that the MS paradigm stops being valid for luminous merging systems near to coalescence, due to the rapid quenching of star formation by the AGN feedback. This is in agreement with recent evidence of quenching in intermediate redshift galaxies affecting the MS, and implies that the MS paradigm needs to be evaluated carefully for samples without a though SED analysis that includes AGN emission. A similar study with a much larger sample of galaxies might be more conclusive in this respect.

6. Our sample is significantly larger than many other studies and (we argue) the SED method is more accurate; using it we support and refine earlier conclusions that $f_{\rm AGN}$ correlates with L$_{\rm IR}$, with the average AGN contribution to a galaxy's L$_{\rm IR}$ increasing from about 5% to 36% as L$_{\rm IR}$ increases from 10^9 to 10^{12}L$_\odot$.

CIGALE was in many cases able to identify Type 1 AGNs by varying the viewing angle Psi and looking for a minimum χ^2 value (and other parameters changes); Type

1's had optimum fits with $\psi \gtrsim 70$. This feature may be of particular value in studies of high-z objects whose morphology is unknown but whose line strengths cannot be properly analyzed without attention to the possible extinction corrections. A new version of CIGALE has recently been released which we plan to use in a more detailed analysis of viewing angle effects.

In the near future, planned and proposed facilities such as *JWST* and *SPICA*, and ground-based telescopes, will provide better resolution and new insights into the physical processes at work in galaxies and their evolution. In particular, they will begin to piece together the cosmic history of galaxies in the universe. The method of meticulous SED modelling, as presented in this work, can play an important role in the interpretation of these new datasets.

Acknowledgements: The authors gratefully thank Steve Willner, Andreas Zezas, Lingyu Wang, Elizabeth Hora, Madison Hora, John Della Costa III, and Antonio Frigo for their assistance and advice. They particularly thank Denis Burgarella and Laura Ciesla from the CIGALE team for their expert guidance and assistance. We acknowledge the anonymous referee for a careful reading of the manuscript and very helpful questions and comments. HAS, MA, and JRM-G acknowledge partial support from NASA Grants NNX14AJ61G and NNX15AE56G.

The SAO REU program is funded by the National Science Foundation REU and Department of Defense ASSURE programs under NSF Grant AST-1659473, and by the Smithsonian Institution.

This publication makes use of data products from the Two Micron All Sky Survey, which is a joint project of the University of Massachusetts and the Infrared Processing and Analysis Center/California Institute of Technology, funded by the National Aeronautics and Space Administration and the National Science Foundation.

This research is based on observations made with the Galaxy Evolution Explorer, obtained from the MAST data archive at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-26555.

Funding for SDSS-III has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, and the U.S. Department of Energy Office of Science. The SDSS-III web site is `http://www.sdss3.org/`. SDSS-III is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS-III Collaboration including the University of Arizona, the Brazilian Participation Group, Brookhaven National Laboratory, Carnegie Mellon University, University of Florida, the French Participation Group, the German Participation Group, Harvard University, the Instituto de Astrofisica de Canarias, the Michigan State/Notre Dame/JINA Participation Group, Johns Hopkins University, Lawrence Berkeley National Laboratory, Max Planck Institute for Astrophysics, Max Planck Institute for Extraterrestrial Physics, New Mexico State University, New York University, Ohio State University, Pennsylvania State University, University of Portsmouth, Princeton University, the Spanish Participation Group, University of Tokyo, University of Utah, Vanderbilt University, University of Virginia, University of Washington, and Yale University.

This work is based in part on observations made with the *Spitzer Space Telescope*, which is operated by the Jet Propulsion Laboratory, California Institute of Technology under a contract with NASA. This publication makes use of data products from the *Wide-field Infrared Survey Explorer*, which is a joint project of the University of California, Los Angeles, and the Jet Propulsion Laboratory/-California Institute of Technology, funded by the National Aeronautics and Space Administration." Herschel is an ESA space observatory with science instruments provided by European-led Principal Investigator consortia and with important participation from NASA.

This research has also made use of results from NASA's Astrophysics Data System. This research has made use of the NASA/IPAC Extragalactic Database (NED), which is operated by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration. This research has made use of the SIMBAD database, operated at CDS, Strasbourg, France. This research made use of Astropy, a community-developed core Python package

for Astronomy (Astropy Collaboration et al. 2013). We thank the people that support photutils (Bradley et al. 2018).

Table 5.4 – *Basic data for the SIGS sample galaxies.*

Group ID	Galaxy ID	RA (J2000)	Dec (J2000)	Redshift (z)	Sample	Interaction Stage	Size ($'' \times ''$)	Angle[a] (°)
1	NGC 274[b]	00:51:01.6	−07:03:22.7	0.0058	A	4	33.8 × 23.0	130.0
	NGC 275	00:51:04.8	−07:03:59.8	0.0058	A	4	38.2 × 28.1	25.0
2	NGC 470	01:19:44.9	03:24:35.6	0.0079	A	2	90.0 × 55.1	65.0
	NGC 474	01:20:06.7	03:24:55.4	0.0077	A	2	225.0 × 175.0	165.0
3	NGC 520	01:24:35.1	03:47:32.7	0.0076	A	5	147.3 × 97.2	235.0
4	IC 195[b]	02:03:44.6	14:42:33.5	0.0122	A	3	37.8 × 21.2	39.8
	IC 196	02:03:49.8	14:44:20.8	0.0122	A	3	95.0 × 55.1	62.0
5	NGC 833	02:09:20.8	−10:07:59.2	0.0129	A	4	42.5 × 23.0	175.0
	NGC 835	02:09:24.6	−10:08:09.2	0.0136	A	4	42.8 × 35.3	125.0
	NGC 838	02:09:38.5	−10:08:48.1	0.0128	A	3	45.0 × 25.9	175.0
	NGC 839	02:09:42.9	−10:11:02.8	0.0129	A	2	45.0 × 28.0	5.0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Note: Group IDs, Redshifts, Sample, and Interaction Stages are taken from B15, as described in Sec. 5.2.1. A Sample of C or A indicates objects belonging to the Keel-Complete sample or the Arp sample, respectively. The RA, Dec, Size, and Angle columns correspond to the centroids, semi-axis lengths, and position angles of the elliptical apertures used for the photometry as described in Sec. 5.3.3.

[a] Angles are given in degrees from the East as measured by *photutils* (Sec. 5.3.2), so Angle=PA−90 degrees.

[b] This galaxy was not analyzed in the SED models described in Sec. 5.4.1 because the photometry was too sparse to support reliable SED models.

Appendix

Here we present all the information for the samples, photometry values, line emissions, examples of the CIGALE SED fitting, CIGALE derived parameters and histograms of the parameters analysed for all the 188 galaxies presented in this work. We include a table with the derived parameters for six AGN galaxies where a different viewing angle in CIGALE give different output parameters. In addition, as online material, we provide the SED fits of the 178 galaxies with good fits. For the remaining 10 galaxies where the SED fit is not good enough, we provide the AGN and no-AGN SED fits.

The following tables and figures show small examples of what is presented in the original article.

Table 5.5 – *Basic data for the SB sample galaxies.*

Galaxy ID	RA (J2000)	Dec	Redshift (z)	Sample	Size ($'' \times ''$)	Angle[a] (°)
NGC 23	00:09:53.4	25:55:25.6	0.0152	S	86.3 × 58.2	90.0
NGC 253	00:47:32.4	−25:17:44.0	0.0008	S	820.4 × 226.5	140.0
NGC 660	01:43:02.4	13:38:42.2	0.0028	SB	304.8 × 124.2	75.0
NGC 1222	03:08:56.7	−02:57:18.5	0.0081	B	73.4 × 60.0	70.0
NGC 1365	03:33:36.4	−36:08:28.2	0.0055	B	353.8 × 221.6	128.0
IC 342	03:46:48.5	68:05:46.9	0.0001	B	716.5 × 598.5	0.0
NGC 1614	04:33:59.8	−08:34:44.0	0.0159	B	82.9 × 54.9	114.7
NGC 1797	05:07:44.9	−08:01:08.7	0.0149	S	66.7 × 41.0	162.9
NGC 2146	06:18:37.7	78:21:25.3	0.003	B	174.8 × 125.9	210.0
NGC 2623	08:38:24.0	25:45:16.1	0.0185	B	76.4 × 45.8	160.0
⋮	⋮	⋮	⋮	⋮	⋮	⋮

Note: The RA, Dec, Size, and Angle columns define the centroids, semi-axis lengths, and angles of the elliptical apertures used for the photometry of the starburst sample galaxies. Redshifts were taken from NED. Samples are B for Brandl et al. (2006) and S for added well known local starbursts.
[a] Angles are given in degrees from the East as measured by *photutils* (Sec. 5.3.2), so Angle=PA−90 degrees.

Table 5.6 – *Basic data for the AGN sample galaxies.*

Galaxy ID	RA (J2000)	Dec	Redshift (z)	Sample	Size ($'' \times ''$)	Angle[a] (°)
Mrk 335	00:06:19.5	20:12:10.5	0.0258	S	26.9 × 26.8	84.2
Mrk 1502	00:53:34.9	12:41:36.2	0.0589	S	29.0 × 28.5	45.0
NGC 931	02:28:14.5	31:18:42.0	0.0167	S	100.9 × 35.7	165.7
NGC 1068	02:42:40.7	−00:00:47.8	0.0038	HRG	215.1 × 174.7	170.0
NGC 1194	03:03:49.1	−01:06:13.5	0.0136	S	100.7 × 42.7	50.6
NGC 1320	03:24:48.7	−03:02:32.2	0.0089	S	63.5 × 33.9	47.2
ESO 33−2	04:55:59.0	−75:32:28.2	0.0181	S	31.0 × 29.0	45.0
4U 0557−385	05:58:02.0	−38:20:04.7	0.0339	S	25.0 × 23.0	229.3
Mrk 3	06:15:36.4	71:02:15.1	0.0135	S	48.6 × 42.9	55.0
ESO 428−14	07:16:31.2	−29:19:29.0	0.0057	S	46.2 × 32.0	230.0
⋮	⋮	⋮	⋮	⋮	⋮	⋮

Note: The RA, Dec, Size, and Angle columns define the centroids, semi-axis lengths, and angles of the elliptical apertures used for the photometry of the AGN sample galaxies. Redshifts were taken from NED. The Sample column indicates whether objects belong to the GOALS sample (G), Higuera-G. & Ramos P. (2013) (HR) or taken from SIMBAD (S).
[a] Angles are given in degrees from the East as measured by *photutils* (Sec. 5.3.2), so Angle=PA−90 degrees.

Table 5.7 – *Basic data for 38 of the 49 LSM sample galaxies.*

Galaxy ID	RA (J2000)	Dec	Redshift (z)	$\log(L_{IR})$ ($L_\odot$)	Stage	Size ($'' \times ''$)	Angle[a] (°)
NGC 0078	00:20:26.6	00:49:46.7	0.0183	9.98	2.0	60.0 × 35.0	145.0
UM 246	00:29:45.1	00:10:09.0	0.0594	10.83	4.5	60.0 × 30.0	45.0
2MASX J01221811+0100262	01:22:17.8	01:00:27.5	0.0555	11.54	4.0	39.0 × 36.0	30.0
CGCG 087-046	07:54:31.8	16:48:26.3	0.0463	11.28	4.0	45.0 × 35.0	115.0
UGC 04383	08:23:33.5	21:20:34.7	0.0179	10.46	3.0	55.0 × 38.0	125.0
2MASX J08343370+1720462	08:34:33.7	17:20:46.4	0.0479	10.86	4.0	30.0 × 28.0	115.0
2MASX J08381760+3054533	08:38:17.6	30:54:53.5	0.0477	10.62	4.0	28.0 × 22.0	110.0
2MASX J08434495+3549421	08:43:45.0	35:49:42.0	0.054	10.45	5.0	28.0 × 25.0	55.0
UGC 05044	09:27:44.0	12:17:12.3	0.029	10.52	4.0	50.0 × 45.0	110.0
Arp 142	09:37:44.0	02:45:15.1	0.0233	10.89	3.5	65.0 × 55.0	130.0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Note: The RA, Dec, Size, and Angle columns define the centroids, semi-axis lengths, and angles of the elliptical apertures used for the photometry of the LSM sample galaxies. This table presents data for 38 galaxies; corresponding quantities for the remainder of the 49-galaxy sample appear in Table 1 of Dietrich et al. (2018). Redshifts were taken from NED. Infrared luminosities were taken from the Revised IRAS-FSC Redshift Catalog (RIFSCz, Wang et al. 2014). The stages were determined by the entire team as described in Sec. 5.2.1.

[a] Angles are given in degrees from the East as measured by *photutils* (Sec. 5.3.2), so Angle=PA−90 degrees.

[b] This galaxy was not analyzed in the SED models described in Sec. 5.4.1 because the photometry was too sparse to support reliable SED models.

Table 5.8 – *Integrated IR line intensities measured for SB sample galaxies using PAHFIT in units of* 1×10^{-21} W cm^{-2}.

Galaxy ID	[Ne II] 12.81 μm	[Ne III] 15.56 μm	[Ne V] 14.32 μm	[S III] 18.71 μm	[S IV] 10.51 μm	[Fe II] 12.64 μm	H_2 S(2) 12.28 μm	H_2 S(1) 17.03 μm
NGC 23	49.37±0.14	6.95±0.04	···	16.08±0.14	1.62±0.06	0.48±0.06	5.71±0.11	6.65±0.07
NGC 253	2689.15±1.29	183.27±0.57	···	576.23±1.08	···	77.21±0.83	93.53±0.80	79.39±0.67
NGC 520	46.18±0.21	6.84±0.09	0.23±0.08	6.71±0.11	0.13±0.09	···	5.24±0.12	8.89±0.10
NGC 660	286.20±0.34	27.60±0.19	1.30±0.14	54.78±0.24	1.47±0.22	3.40±0.14	6.30±0.13	17.82±0.20
NGC 1222	82.21±0.16	78.18±0.13	0.47±0.06	49.33±0.17	25.16±0.12	0.69±0.09	4.79±0.09	7.92±0.08
NGC 1365	178.44±0.37	55.36±0.23	18.74±0.20	55.25±0.30	28.31±0.29	4.35±0.17	12.61±0.18	20.18±0.18
IC 342	639.63±0.38	34.58±0.16	2.31±0.16	255.47±0.39	5.20±0.21	11.80±0.21	12.01±0.26	9.30±0.18
NGC 1614	265.07±0.30	66.59±0.19	1.63±0.10	70.91±0.30	11.65±0.17	1.97±0.14	6.84±0.15	10.05±0.14
NGC 1797	67.82±0.19	5.33±0.07	0.34±0.12	18.63±0.16	1.11±0.09	0.53±0.10	3.50±0.14	3.92±0.10
NGC 2146	803.04±1.52	121.59±0.43	2.99±0.31	253.62±1.17	21.55±0.82	10.90±0.30	10.07±0.28	37.91±0.35
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Table 5.9 – *GALEX and SDSS DR12 photometry for the four study samples.*

Galaxy	*GALEX*		SDSS DR12				
ID	FUV (mJy)	NUV (mJy)	u (mJy)	g (mJy)	r (mJy)	i (mJy)	z (mJy)
Photometry for the SIGS sample							
NGC274	···	···	4.53 ± 0.13	21.15 ± 0.42	43.97 ± 0.88	64.74 ± 1.30	81.14 ± 1.66
NGC275	3.30 ± 0.33	4.56 ± 0.46	10.80 ± 0.24	23.35 ± 0.47	36.91 ± 0.74	43.79 ± 0.88	50.87 ± 1.08
NGC470	3.18 ± 0.32	4.67 ± 0.47	16.59 ± 0.37	44.43 ± 0.89	78.42 ± 1.57	107.23 ± 2.15	131.44 ± 2.67
NGC474	1.44 ± 0.15	1.98 ± 0.20	20.66 ± 0.61	67.87 ± 1.36	125.69 ± 2.52	179.48 ± 3.61	243.94 ± 5.05
NGC520	1.69 ± 0.17	3.12 ± 0.31	4.25 ± 0.22	62.00 ± 1.25	113.32 ± 2.27	160.68 ± 3.22	192.31 ± 3.92
IC195	0.05 ± 0.01	0.10 ± 0.01	1.96 ± 0.06	9.48 ± 0.19	19.99 ± 0.40	29.45 ± 0.59	37.98 ± 0.77
IC196	0.65 ± 0.06	0.82 ± 0.08	4.33 ± 0.14	18.03 ± 0.36	36.51 ± 0.73	53.72 ± 1.08	68.04 ± 1.41
NGC833	···	0.29 ± 0.03	3.33 ± 0.09	16.49 ± 0.33	35.86 ± 0.72	53.35 ± 1.07	68.14 ± 1.39
NGC835	···	1.67 ± 0.17	7.74 ± 0.17	29.07 ± 0.58	57.24 ± 1.15	81.33 ± 1.63	102.36 ± 2.07
NGC838	···	2.04 ± 0.20	6.39 ± 0.14	16.92 ± 0.34	30.38 ± 0.61	39.34 ± 0.79	48.92 ± 1.02
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Note: Photometry expressed in mJy in the UV and optical bands for the SIGS, SB, AGN, and LSM samples described in Sec. 5.2. The full table is available in the online version of this paper. A portion is shown here for guidance regarding its form and content.

Table 5.10 – *2MASS and Spitzer/IRAC photometry for the four study samples.*

Galaxy	2MASS			*Spitzer*/IRAC			
ID	J (mJy)	H (mJy)	K_s (mJy)	3.6 μm (mJy)	4.5 μm (mJy)	5.8 μm (mJy)	8.0 μm (mJy)
Photometry for the SIGS sample							
NGC274	111.29 ± 2.28	132.95 ± 2.76	102.24 ± 2.28	50.25 ± 1.51	31.65 ± 0.95	23.19 ± 0.70	17.06 ± 0.52
NGC275	64.31 ± 1.41	73.18 ± 1.68	59.58 ± 1.68	39.06 ± 1.17	27.29 ± 0.82	70.69 ± 2.13	170.52 ± 5.12
NGC470	186.08 ± 3.96	199.56 ± 4.58	184.86 ± 4.47	108.22 ± 3.25	73.47 ± 2.20	163.96 ± 4.93	417.70 ± 12.53
NGC474	296.05 ± 7.02	314.31 ± 8.98	271.30 ± 8.35	149.83 ± 4.50	86.50 ± 2.60	141.57 ± 4.34	100.29 ± 3.14
NGC520	288.70 ± 6.23	357.34 ± 8.33	312.68 ± 7.30	182.73 ± 5.48	138.65 ± 4.16	357.79 ± 10.74	916.85 ± 27.51
IC195	54.38 ± 1.22	66.50 ± 1.60	53.22 ± 1.47	25.29 ± 0.76	15.80 ± 0.47	10.82 ± 0.34	6.30 ± 0.21
IC196	95.34 ± 2.39	129.52 ± 3.48	105.90 ± 3.34	49.53 ± 1.49	32.17 ± 0.97	29.94 ± 0.92	42.85 ± 1.30
NGC833	101.56 ± 2.08	125.16 ± 2.62	106.44 ± 2.32	48.47 ± 1.45	30.78 ± 0.92	25.01 ± 0.76	25.52 ± 0.77
NGC835	155.86 ± 3.17	194.36 ± 4.01	166.35 ± 3.51	89.23 ± 2.68	60.95 ± 1.83	117.65 ± 3.53	291.88 ± 8.76
NGC838	73.96 ± 1.57	93.24 ± 2.04	86.14 ± 1.97	67.69 ± 2.03	50.08 ± 1.50	212.29 ± 6.37	589.03 ± 17.67
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Note: Photometry in seven near- and mid-IR bands for the SIGS, SB, AGN, and LSM samples described in Sec. 5.2. The full table is available in the online version of this paper. A portion is shown here for guidance regarding its form and content.

Table 5.11 – *WISE and Spitzer/MIPS photometry for the four study samples.*

Galaxy	*WISE*				*Spitzer*/MIPS		
ID	3.4 μm (mJy)	4.6 μm (mJy)	12 μm (mJy)	22 μm (mJy)	24 μm (mJy)	70 μm (Jy)	160 μm (Jy)
Photometry for the SIGS sample							
NGC274	52.6 ± 3.2	28.3 ± 1.7	11.3 ± 0.7	···	···	···	···
NGC275	38.1 ± 2.3	24.5 ± 1.5	146.5 ± 8.8	384.6 ± 23.1	459.32 ± 18.39	5.40 ± 0.22	7.34 ± 0.32
NGC470	112.5 ± 6.8	69.5 ± 4.2	334.5 ± 20.1	802.2 ± 48.1	799.55 ± 32.02	9.59 ± 0.39	13.72 ± 0.55
NGC474	151.8 ± 9.1	83.3 ± 5.0	···	···	152.39 ± 8.99	···	···
NGC520	183.0 ± 11.0	131.4 ± 7.9	738.3 ± 44.3	2233.4 ± 134.0	2347.72 ± 93.94	33.44 ± 1.34	···
IC195	26.0 ± 1.6	13.9 ± 0.8	4.2 ± 0.3	3.5 ± 0.3	···	···	···
IC196	52.6 ± 3.2	28.4 ± 1.7	37.5 ± 2.3	36.0 ± 2.2	···	···	···
NGC833	51.8 ± 3.1	29.2 ± 1.7	26.7 ± 1.6	49.6 ± 3.0	···	···	···
NGC835	89.9 ± 5.4	56.5 ± 3.4	242.0 ± 14.5	412.1 ± 24.7	434.53 ± 17.43	6.82 ± 0.27	9.33 ± 0.38
NGC838	62.1 ± 3.7	46.2 ± 2.8	459.2 ± 27.6	1344.6 ± 80.7	1458.64 ± 58.35	11.50 ± 0.46	9.60 ± 0.41
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Note: Photometry in seven mid-IR bands for the SIGS, SB, AGN, and LSM samples described in Sec. 5.2. The full table is available in the online version of this paper. A portion is shown here for guidance regarding its form and content.

Table 5.12 – *Herschel/PACS and SPIRE photometry for the four study samples.*

Galaxy	*Herschel*/PACS			*Herschel*/SPIRE		
ID	70 μm (Jy)	100 μm (Jy)	160 μm (Jy)	250 μm (Jy)	350 μm (Jy)	500 μm (Jy)
Photometry for the SIGS sample						
NGC274	···	···	···	···	···	···
NGC275	5.96 ± 0.60	8.71 ± 0.87	7.85 ± 0.79	···	···	···
NGC470	···	···	···	···	···	···
NGC474	···	···	···	···	···	···
NGC520	41.68 ± 4.17	51.86 ± 5.19	39.13 ± 3.91	···	···	···
IC195	···	···	···	···	···	···
IC196	···	···	···	···	···	···
NGC833	0.39 ± 0.07	0.79 ± 0.10	1.14 ± 0.12	0.58 ± 0.04	0.27 ± 0.02	0.10 ± 0.01
NGC835	7.44 ± 0.75	11.72 ± 1.17	10.58 ± 1.06	4.04 ± 0.28	1.49 ± 0.10	0.46 ± 0.03
NGC838	14.98 ± 1.50	17.84 ± 1.78	12.99 ± 1.30	3.91 ± 0.27	1.37 ± 0.10	0.41 ± 0.03
⋮	⋮	⋮	⋮	⋮	⋮	⋮

Note: Photometry in the FIR bands for the SIGS, SB, AGN, and LSM samples described in Sec. 5.2. The full table is available in the online version of this paper. A portion is shown here for guidance regarding its form and content.

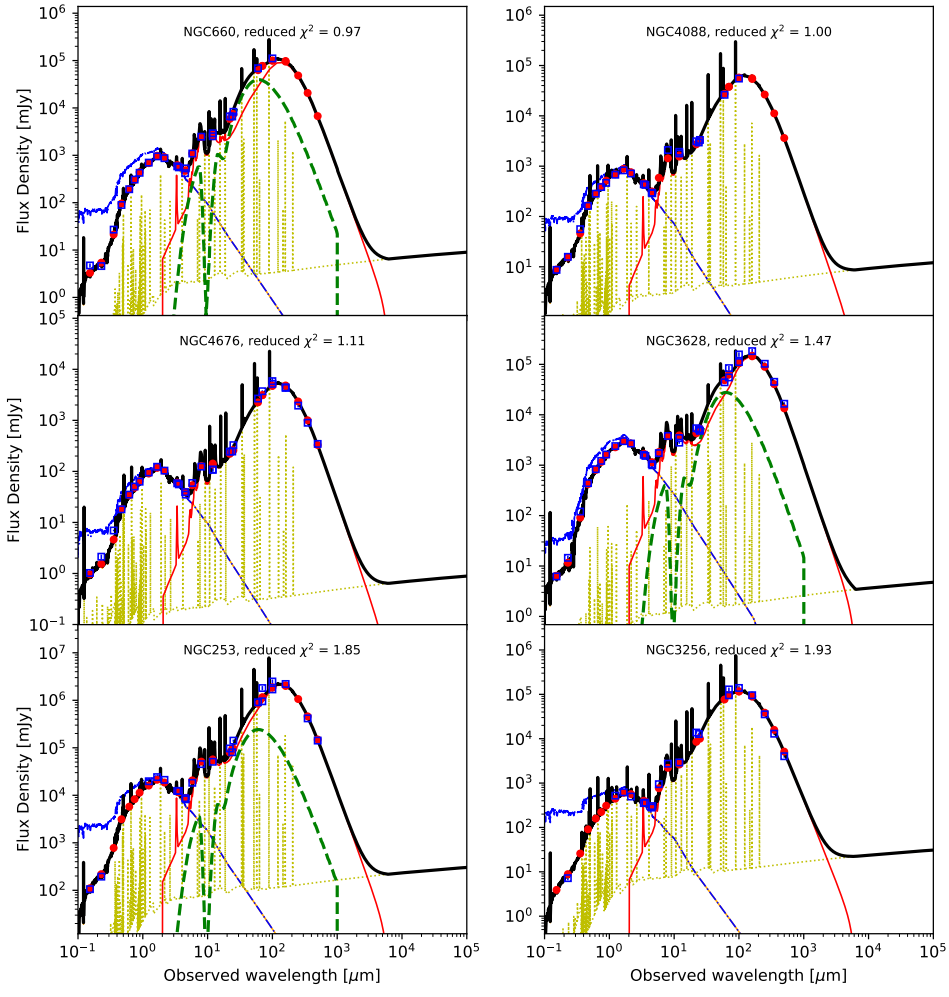

Figure 5.16 – *Best-fit SED models for 6 galaxies in the SB sample containing the nebular emission (gold dotted lines), both attenuated stellar emission (orange) and non-attenuated stellar emission (blue dot-dashed), dust emission (red solid), and AGN emission (green dashed). The red dots are the best model flux densities and the blue squares mark the observed flux densities with 1σ error bars.*

Table 5.13 – CIGALE-*derived parameters for the AGN sample.*

Galaxy ID	$f_{\rm AGN}$	$L_{\rm AGN}$ $\log(L_\odot)$	Old Att. $\log(L_\odot)$	Young Att. $\log(L_\odot)$	Dust α	$L_{\rm dust}$ $\log(L_\odot)$
Mrk335	0.77 ± 0.04	11.03 ± 0.02	9.42 ± 0.15	9.78 ± 0.07	1.03 ± 0.08	9.99 ± 0.05
Mrk1502	0.49 ± 0.03	11.72 ± 0.02	11.28 ± 0.05	11.31 ± 0.04	1.52 ± 0.08	11.64 ± 0.03
NGC931	0.51 ± 0.04	10.68 ± 0.04	10.50 ± 0.04	9.95 ± 0.07	2.68 ± 0.21	10.62 ± 0.04
NGC1068	0.42 ± 0.07	10.93 ± 0.08	10.86 ± 0.07	10.56 ± 0.11	2.01 ± 0.10	11.06 ± 0.05
NGC1194	0.66 ± 0.07	10.17 ± 0.05	9.83 ± 0.08	8.06 ± 0.44	2.20 ± 0.24	9.84 ± 0.08
NGC1320	0.43 ± 0.06	9.85 ± 0.05	9.87 ± 0.06	9.08 ± 0.30	1.89 ± 0.12	9.95 ± 0.07
ESO33-2	0.63 ± 0.03	10.33 ± 0.02	10.02 ± 0.03	7.99 ± 0.59	1.75 ± 0.09	10.03 ± 0.03
4U0557-385	0.90 ± 0.04	11.25 ± 0.02	10.14 ± 0.04	9.32 ± 0.19	1.84 ± 0.12	10.21 ± 0.02
Mrk3	0.64 ± 0.06	10.62 ± 0.03	10.34 ± 0.08	8.75 ± 0.04	1.63 ± 0.13	10.36 ± 0.08
ESO428-14	0.31 ± 0.04	9.50 ± 0.06	9.77 ± 0.07	8.89 ± 0.44	1.74 ± 0.09	9.83 ± 0.03
⋮	⋮	⋮	⋮	⋮	⋮	⋮

Note: Galaxy ID is the common identifier used in the same order as in Table 5.6, $f_{\rm AGN}$ is the fraction of AGN contribution (from both torus and accretion) to the IR, or $L_{\rm IR}^{\rm AGN} = f_{\rm AGN} \times L_{\rm IR}^{\rm Tot}$ as defined by Ciesla et al. (2015), $L_{\rm AGN}$ is the AGN luminosity of the three AGN components by Fritz et al. (2006) , Old Att. is the attenuation from the old stellar population, Young Att. is the attenuation from the young stellar population, Dust α is the parameter that defines the contribution of the local heating intensity in the dust (eq. 5.1) and $L_{\rm dust}$ is the dust luminosity.

Galaxy ID	SFR $\log(M_\odot/yr)$	$\tau_{\rm main}$ $\log(yr)$	Stellar Age $\log(yr)$	$M_{\rm gas}$ $\log(M_\odot)$	$M_\star$ $\log(M_\odot)$
Mrk335	0.49 ± 0.09	9.60 ± 0.28	8.34 ± 0.44	8.29 ± 0.64	8.94 ± 0.37
Mrk1502	1.86 ± 0.03	9.56 ± 0.33	8.92 ± 0.20	10.40 ± 0.17	10.93 ± 0.13
NGC931	0.49 ± 0.05	8.99 ± 0.04	9.61 ± 0.02	10.77 ± 0.02	11.16 ± 0.02
NGC1068	1.07 ± 0.11	8.95 ± 0.10	9.36 ± 0.07	10.70 ± 0.09	11.13 ± 0.08
NGC1194	-1.31 ± 0.42	8.47 ± 0.33	9.55 ± 0.09	10.45 ± 0.07	10.84 ± 0.06
NGC1320	-0.37 ± 0.25	8.81 ± 0.15	9.52 ± 0.06	10.15 ± 0.03	10.56 ± 0.02
ESO33-2	-1.46 ± 0.58	8.39 ± 0.40	9.56 ± 0.12	10.48 ± 0.10	10.88 ± 0.09
4U0557-385	-0.06 ± 0.11	8.72 ± 0.12	9.48 ± 0.04	10.61 ± 0.07	11.02 ± 0.07
Mrk3	-0.74 ± 0.03	8.70 ± 0.02	9.60 ± 0.02	10.77 ± 0.02	11.16 ± 0.02
ESO428-14	-0.59 ± 0.43	8.72 ± 0.39	9.40 ± 0.27	9.85 ± 0.18	10.28 ± 0.13
⋮	⋮	⋮	⋮	⋮	⋮

Note: Galaxy ID is the common identifier used in the same order as in Table 5.6, SFR is the star formation rate, $\tau_{\rm main}$ is the e-folding time of the main stellar population model, Stellar Age is the age of the oldest stars in the galaxy, $M_{\rm gas}$ is the gas mass and $M_\star$ is the stellar mass.

Table 5.14 – CIGALE-*derived parameters for six AGN galaxies, where a Type 1 AGN ($\psi = 70$) give better χ^2. Units as Table 5.13.*

Galaxy ID	$f_{\rm AGN}$	$L_{\rm AGN}$ $\log(L_\odot)$	Old Att. $\log(L_\odot)$	Young Att. $\log(L_\odot)$	Dust α	$L_{\rm dust}$ $\log(L_\odot)$
Mrk335	0.86 ± 0.04	11.14 ± 0.02	9.88 ± 0.07	8.53 ± 0.90	1.06 ± 0.16	9.90 ± 0.06
Mrk771	0.81 ± 0.04	11.43 ± 0.02	10.18 ± 0.13	9.84 ± 0.23	2.24 ± 0.29	10.37 ± 0.09
2XMMJ141348.3+440014	0.90 ± 0.04	11.96 ± 0.02	10.30 ± 0.18	9.95 ± 0.35	2.33 ± 0.25	10.49 ± 0.04
Mrk1383	0.80 ± 0.04	12.00 ± 0.04	10.52 ± 0.25	10.59 ± 0.20	2.05 ± 0.16	10.89 ± 0.05
ESO141-55	0.74 ± 0.04	11.56 ± 0.02	10.52 ± 0.10	9.55 ± 0.64	2.56 ± 0.21	10.57 ± 0.06
Mrk1513	0.86 ± 0.04	11.87 ± 0.02	10.61 ± 0.11	9.65 ± 0.63	2.13 ± 0.23	10.66 ± 0.07

Galaxy ID	SFR $\log(M_\odot/yr)$	$\tau_{\rm main}$ $\log(yr)$	Stellar Age $\log(yr)$	$M_{\rm gas}$ $\log(M_\odot)$	$M_\star$ $\log(M_\odot)$
Mrk335	-0.94 ± 0.91	8.51 ± 0.53	9.58 ± 0.14	10.08 ± 0.17	10.47 ± 0.16
Mrk771	0.51 ± 0.24	8.94 ± 0.22	9.47 ± 0.12	10.47 ± 0.16	10.88 ± 0.14
2XMMJ141348.3+440014	0.56 ± 0.40	9.31 ± 0.52	9.34 ± 0.30	10.27 ± 0.45	10.68 ± 0.41
Mrk1383	1.31 ± 0.21	9.00 ± 0.39	9.24 ± 0.22	10.68 ± 0.19	11.14 ± 0.15
ESO141-55	0.12 ± 0.68	8.60 ± 0.47	9.49 ± 0.21	10.66 ± 0.22	11.06 ± 0.19
Mrk1513	0.18 ± 0.65	8.72 ± 0.67	9.51 ± 0.19	10.68 ± 0.24	11.08 ± 0.22

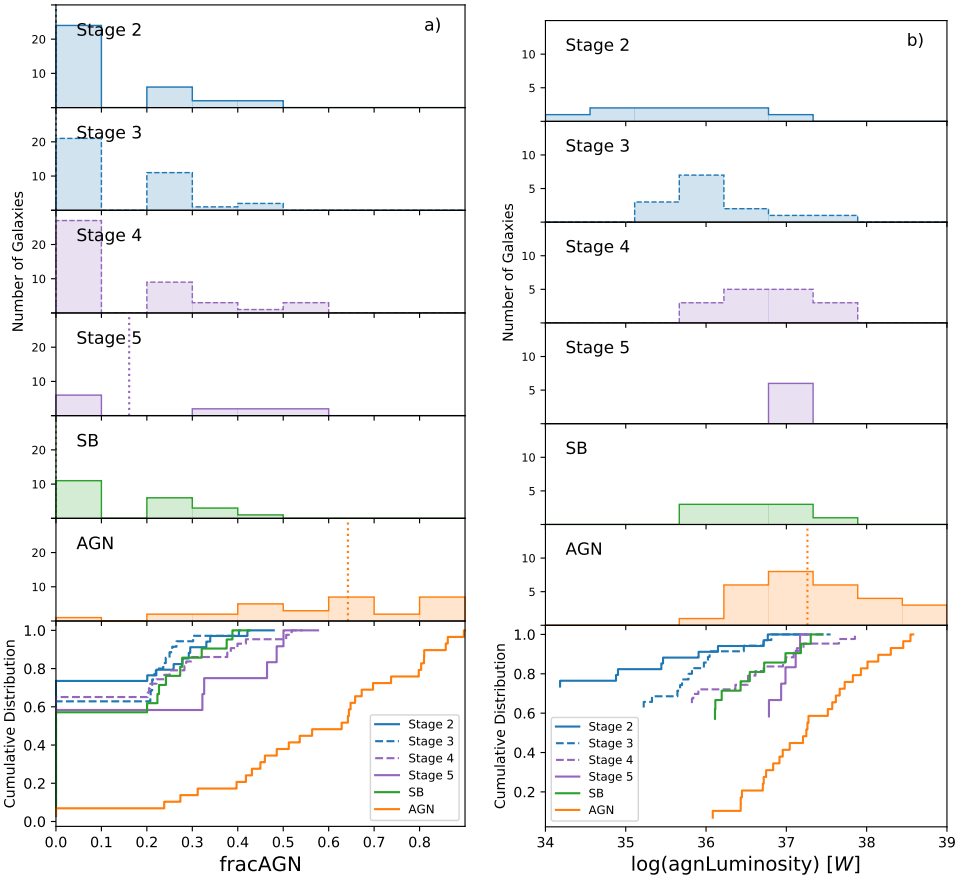

Figure 5.17 – *Histograms (top) for the stages of the SIGS+LSM, AGN and SB samples and the normalized cumulative distributions (bottom) for the AGN fraction (a) and luminosity (b). The colors and lines are identical to Figure 5.6.*

Life is like a box a chocolate, you never know what your going to get.

Forrest Gump (Tom Hanks) - Forrest Gump

Hey! I got another chocolate from the university. Wiiii!

Andrés checking the mail

6

The viewing angle in AGN SED models: a data-driven analysis

A. F. Ramos Padilla, Lingyu Wang, Katarzyna Małek, Andreas Efstathiou and Guang Yang

Published in Monthly Notices of the Royal Astronomical Society, Vol. 510, Issue 1, Pages 687–707

Highlights

- The unified model of AGN considers that the dust torus is the main responsible for some AGN types classified because of the viewing angle.
- We assess the importance of the viewing angle in the classification of AGN galaxies, in particular Seyfert galaxies, in SED AGN models.
- To make the results of this research reproducible, we use data from public astronomical databases and publish all data analysed in this work.
- AGN disc luminosity is the best discriminator of the AGN Types in Seyfert galaxies, rather than the viewing angle which is the second discriminator when assessing AGN type.
- We found an evolutionary path between Type-1 and Type-2 AGN (i.e. their intermediate classifications) in terms of the accretion rates inferred by the AGN models.

Abstract

Context: The validity of the unified active galactic nuclei (AGN) model has been challenged in the last decade, especially when different types of AGNs are considered to only differ in the viewing angle to the torus.
Aims: We aim to assess the importance of the viewing angle in classifying different types of Seyfert galaxies in spectral energy distribution (SED) modelling.
Methods: We retrieve photometric data from publicly available astronomical databases: CDS and NED, to model SEDs with X-CIGALE in a sample of 13 173 Seyfert galaxies located at redshift range from $z = 0$ to $z = 3.5$, with a median redshift of $z \approx 0.2$. We assess whether the estimated viewing angle from the SED models reflects different Seyfert classifications. Two AGN models with either a smooth or clumpy torus structure are adopted in this paper.
Results: We find that the viewing angle in Type-1 AGNs is better constrained than in Type-2 AGNs. Limiting the viewing angles representing these two types of AGNs do not affect the physical parameter estimates such as star-formation rate (SFR) or AGN fractional contribution ($f_{\rm AGN}$). In addition, the viewing angle is not the most discriminating physical parameter to differentiate Seyfert types.
Conclusions: We suggest that the observed and intrinsic AGN disc luminosity can: i) be used in $z < 0.5$ studies to distinguish between Type-1 and Type-2 AGNs, and ii) explain the probable evolutionary path between these AGN types. Finally, we propose the use of X-CIGALE for AGN galaxy classification tasks. All data from the 13 173 SED fits are available at Zenodo*.

Keywords: Methods: data analysis, statistical – astronomical data bases: miscellaneous – Techniques: photometric – galaxies: Seyfert

6

6.1 Introduction

The presence of an optically thick structure in a Seyfert galaxy (Antonucci & Miller 1985) led to the creation of the unified model of active galactic nuclei (AGN), where an obscuring torus explains the variety of AGN types due to the orientation with respect to the line of sight (e.g. Antonucci 1993; Urry & Padovani 1995). This simple model uses the viewing angle to separate AGN galaxies into two types: unobscured (Type-1 AGN) and obscured (Type-2 AGN). Type-1 AGNs have broad emission lines, in terms of the full width at half maximum (FWHM), while Type 2 AGNs do not, due to differences in the viewing angle i. We observe narrow-line regions (NLR, with FWHM $\lesssim 1000\,{\rm km\,s^{-1}}$) or broad line regions (BLR, with FWHM $\gtrsim 1000\,{\rm km\,s^{-1}}$) depending on i. In Type-1 AGN, the BLR and NLR are viewed directly because the galaxies are viewed at small angles with respect to the line of sight ($i \approx 0 - 30°$), while for Type-2 AGN, only the NLR is visible because the galaxies are viewed at high angles with respect to the line of sight ($i \approx 70 - 90°$) and obscuration hides the BLR (e.g. Antonucci 1993; Kauffmann et al. 2003a). However, in the last decade, the "zoo" of AGNs has become more complex and difficult to explain with this simple toroidal structure model (Padovani et al. 2017) and the viewing angle has not been easy to estimate (e.g Marin 2016). Besides, obscuration is not static and may depend on different physical conditions that may vary (e.g. Hönig & Kishimoto 2017; Hickox

* https://doi.org/10.5281/zenodo.5221764

& Alexander 2018). In addition, the "changing look" AGNs cannot be explained with the unified model, but rather with the accretion state of the AGN (LaMassa et al. 2015; Elitzur et al. 2014). Therefore, updates to the unified model of AGNs have been proposed describing new AGN scenarios such as clumpy structures (e.g. Krolik & Begelman 1988; Nenkova et al. 2002; Dullemond & van Bemmel 2005), radiation-pressure modes (e.g. Fabian et al. 2008; Ricci et al. 2017; Wada 2015), polar dust (e.g. Braatz et al. 1993; Cameron et al. 1993; Efstathiou et al. 1995; Efstathiou 2006) and disc winds (e.g. Emmering et al. 1992; Netzer 2015; Elitzur & Shlosman 2006; Hönig 2019).

The study of Seyfert galaxies can help to understand the nature of the AGNs in these scenarios. Seyferts are moderate luminosity AGN galaxies that possess high excitation emission lines (Padovani et al. 2017) which can be used to classify these galaxies in Type-1 AGN (Seyfert 1, hereafter Sy1) and Type-2 AGN (Seyfert 2, hereafter Sy2) in catalogues (e.g. Véron-Cetty & Véron 2010). In addition, Seyfert sub-classes, like the narrow line Sy1 (NLSy1, Osterbrock & Pogge 1985; Rakshit et al. 2017) or the intermediate Seyfert types (Osterbrock 1981; Winkler 1992), could be ideal to understand the new AGN scenarios (Elitzur et al. 2014). Nevertheless, large samples of spectroscopically classified Seyfert galaxies are mainly limited to $z < 1$ (e.g. Véron-Cetty & Véron 2010; Koss et al. 2017).

One solution to increase the number of Seyfert galaxies at higher redsfhits is to identify AGNs through colour selections in IR broad-bands (a compilation of these selection criteria is presented by Padovani et al. 2017, table 2) and then observe their spectrum in optical wavelengths for the classifications. However, these photometric broad-bands can also be used in spectral energy distribution (SED) analysis, which allows us to obtain a more reliable estimation of the contribution of the AGN than using only the IR colours (Ciesla et al. 2015; Dietrich et al. 2018; Ramos Padilla et al. 2020; Mountrichas et al. 2021; Pouliasis et al. 2020). The contribution of the AGN in SED models comes from AGN templates (e.g. Mullaney et al. 2011; Bernhard et al. 2021) or AGN models (e.g. Pier & Krolik 1992; Granato & Danese 1994; Efstathiou & Rowan-Robinson 1995; Fritz et al. 2006; Nenkova et al. 2008; Stalevski et al. 2012, 2016; Siebenmorgen et al. 2015; Tanimoto et al. 2019), which are fitted together with dust emissions and stellar populations (e.g. Calistro Rivera et al. 2016; Leja et al. 2018; Boquien et al. 2019) in different configurations (check Thorne et al. 2021; Pérez-Torres et al. 2021, for an overview of the most popular SED fitting codes). SED models without an AGN contribution do not provide the adequate physical properties of AGNs (e.g. Leja et al. 2018; Dietrich et al. 2018), which could lead to over-estimations in star-formation rate (SFR) and stellar masses, especially in X-ray selected AGNs (Florez et al. 2020).

When the AGN is included in the SED modelling, it is shown to be possible to identify Type-1 and Type-2 AGN (e.g. Calistro Rivera et al. 2016; Ramos Padilla et al. 2020). These AGN types tend to show differences not only in the spectrum, but also in colours from photometric bands. For example, Type-1 AGNs tend to have typically bluer colours than Type-2 because of their higher brightness and lower extinction in the UV and optical bands (Padovani et al. 2017). In addition, in Type-1 the contribution from the AGN seem to be more dominant in UV and NIR–MIR bands, while for Type-2 this contribution is dominant in MIR–FIR bands (Ciesla

et al. 2015), which can explain why it is possible to differentiate Type-1 and Type-2 AGNs according to their fractional contribution of the AGN to the IR (Fritz et al. 2006). Therefore it should also be possible to identify Sy1 and Sy2 galaxies with SED analysis when observing broad-band emissions.

In this work, we aim to assess the importance of the estimated viewing angle in classifying AGN galaxies, and highlight its implications in high-redshift studies. Particularly, we gather a sample of Seyfert galaxies with available photometry in astronomical databases to develop a data-driven approach with easily accessible data. We use X-CIGALE (Yang et al. 2020), a modified version of CIGALE (Boquien et al. 2019), one of the most popular SED tools to obtain physical parameters in host galaxies and AGN itself. X-CIGALE has several AGN-related improvements compared to CIGALE, ideal for this work. In addition, we test the two different AGN models inside X-CIGALE to see how the classifications depend on the selected model. We compare two popular machine learning techniques, random forest (Breiman 2001) and gradient boosting (Chen & Guestrin 2016), with individual physical parameters when classifying unclassified and discrepant cases.

We present the Seyfert sample selection, the description of the SED models and the verification of the estimations with a similar model in Sect. 6.2. Then, we select our main physical parameters, compare the estimated galaxy physical parameters from different AGN SED setups, and we compare different classifications in Sect. 6.3. After that, we present the discussions about the role of the viewing angle, its implications and possible bias of these results (Sect. 6.4). Finally, we present our conclusions (Sect. 6.5).

6.2 Data and Analysis

6.2.1 Seyfert Sample

Seyfert galaxies are a good starting point to differentiate between Type-1 and Type-2 AGN, as described by the AGN unification model. Thus, we selected a sample of Seyfert galaxies by combining the *SIMBAD* astronomical database (Wenger et al. 2000, hereafter SMB)* and the dedicated catalogue of AGNs by Véron-Cetty & Véron (2010, hereafter VCV). SMB is widely used for retrieving basic information of galaxies in an homogeneous manner, while VCV is one of the most popular catalogues for AGN studies. From SMB, we picked galaxies whose main type was Seyfert, including: Seyfert 1 (Sy1), Seyfert 2 (Sy2) and, unclassified Seyfert galaxies. From VCV, we selected all Seyfert types galaxies, including all intermediate numerical classifications (e.g. Sy1.5). We cross-matched SMB and VCV samples using a cross-matching radius of $2''$. We removed galaxies where the difference in redshifts between the catalogues ($|\Delta z|$) was higher than 0.01, which is the limit of the reported numerical accuracy between the catalogues, as shown in Figure 6.1. This decision help us to avoid misidentification and uncertain redshifts in the sample of Seyfert galaxies.

* Data and classification types of the galaxies were retrieved on 2020 December 3.

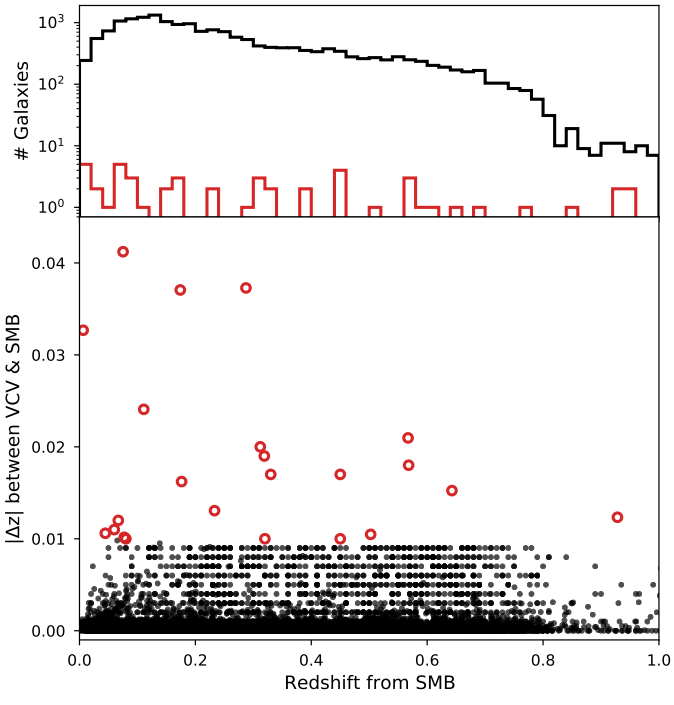

Figure 6.1 – *Redshift distribution for the matches between SMB and VCV catalogues. Upper-panel: Histogram of the redshift distribution for galaxies where the difference in redshift between the catalogues was below (black line) or above (red line) the threshold at 0.01. Only a few galaxies were discarded using this threshold. Bottom-panel: Absolute difference in redshift ($|\Delta z|$) between the catalogues with respect to the SMB redshift. Galaxies with redshifts above 1, or with a large difference in redshift in the catalogues are not shown. Discarded galaxies are indicated as red circles.*

Classification type

We used the classification type from both SMB and VCV samples. Classifications types in VCV come from spectroscopic measurements with SDSS data (Abazajian et al. 2009), while classification types in SMB are a compendium of the literature. The information gathered in *SIMBAD* was manually added by documentalist till the 90's, and now is done semi-automatically with *COSIM* (Brunet et al. 2018). Unfortunately, the source of the classifications was not recorded until 2006, therefore almost half of the Seyfert classifications in SMB are marked as coming from *SIMBAD*. A small fraction of our Seyfert galaxies ($\sim 5\%$) still have an unknown source, as the object type classification is still under development (Oberto et al. 2020)*. Therefore, classifications in SMB should be taken with caution. If the classification source is unknown and the main Seyfert classification in SMB matches VCV, we assume that the classification source is VCV. If the main Seyfert classification source is unknown and the classification in VCV is Seyfert 3 (also known as LINERs) we remove the galaxies from the sample. We re-classified the remaining unknown sources, 49 galaxies, as unclassified Seyfert to study them further. These decisions led us to a sample of

* We use the 2018 August 2 classification version.

18 921 Seyfert galaxies.

For the classifications in SMB, we found: i) Almost half of the classifications (45%) came from the basic data of the galaxy (asigned by the astronomical database); ii) Toba et al. (2014) work contributed to 21% of the Sy1 and Sy2 classifications iii) Zhou et al. (2006), Oh et al. (2015), and Rakshit et al. (2017) together contributed to 25% of SMB classifications, all of them in Sy1 galaxies. The SMB sample contains in total: 13 760 Sy1, 5 040 Sy2, and 121 unclassified Seyfert galaxies.

In VCV, we found 17 different Seyfert type classifications. In this work, we focus on the typical Sy1, Sy2 and unclassified Seyferts which all together account for 71% of the sample. We added the narrow-line Sy1 (NLSy1) galaxies (e.g. Zhou et al. 2006; Rakshit et al. 2017) to the Sy1 classification because most of the NLSy1 galaxies in VCV are classified as Sy1 in SMB. However, some differences in the estimates may indicate that the total accretion power in NLSy1 galaxies is higher than in normal Sy1 galaxies, as we verified in Appendix 6.5. Three of the NLSy1 galaxies (2MASX J10194946+3322041, 2MASS J09455439+4238399 and 2MASX J23383708-0028105) were classified as Sy2 in SMB, so we reclassified them as unclassified Seyfert for further study. In addition, we checked the subgroups between Sy1 and Sy2 as divided by Osterbrock (1977, 1981) and Winkler (1992) which account for 5% of the sample. We denoted a small fraction of the galaxies from VCV (∼1%), which do not fall in the classifications described before (e.g. LINERS, NLSy1.2 and polarised classifications), as alternative Seyfert galaxies. VCV sample contains in total: 13 180 Sy1, 4 567 Sy2, 84 unclassified Seyfert galaxies, 920 in the intermediate numerical subgroups between Sy1 and Sy2, and 170 alternative Seyfert galaxies.

Photometry

We used 31 bands in the UV-FIR wavelength range to get a well-sampled SED for our sample of Seyfert galaxies. We list the selected bands for the SED modelling in Table 6.1 with their respective effective wavelength and number of galaxies detected in that band. We retrieved photometric values of these bands available in CDS* and NED†. CDS and NED photometric data points are ideal for this data-driven work as they are published and curated by other researchers, saving time in the photometric reduction. However, we needed to make sure that the retrieved data were good enough for our purpose.

We keep in mind that the use of heterogeneous measurements may lead to some systematics in the analysis. For example, for galaxies in the local Universe, or where the instrument resolution is good enough to resolve the galaxies, measurements will come from specific regions within the galaxies, like their centres. In contrast, for galaxies at higher redshifts or instruments where the resolution is not high enough to resolve them spatially the measurements will correspond to the whole galaxy as we observe the galaxies as unresolved point sources. Fortunately, in terms of spatial resolution, most of the galaxies in this sample could be treated as point sources for most of the instruments operating at different wavelengths, therefore we expect

* http://cdsportal.u-strasbg.fr/

† The NASA/IPAC Extragalactic Database (NED) is funded by the National Aeronautics and Space Administration and operated by the California Institute of Technology.

measurements at different wavelengths to be consistent with each other. When this is not the case, discrepant apertures at different wavelengths will lead to unphysical jumps in the SED models, which will give us erroneous fittings that we can ignore before going further with the analysis.

Hence, we followed a series of steps to obtain a set of galaxies with useful photometry. First, we decided not to use upper or lower limits from published values in CDS or NED. Second, we remove duplicate data photometry values between the CDS and NED, keeping the value reported in CDS. Third, we used the mean value when more than one measurement was available per band. These measurements can also come from the same apertures but from different works or methods. Fourth, we selected photometric data points with a relative error (after propagating the initial reported errors) below 1/3. Finally, we accounted for the absolute calibration error for each band as in Ramos Padilla et al. (2020), where instrument-dependent uncertainties were added to the measurement uncertainties.

We constrained the galaxies to have good coverage over the optical and IR wavelengths. We only include sources satisfying both criteria: i) more than five photometric data points in wavelengths between $0.1 - 3\mu$m (*GALEX*, SDSS and 2MASS), and ii) more than three photometric data points in wavelengths between $3 - 500\mu$m (*Spitzer*, *WISE*, *IRAS* and *Herschel*). With these criteria, we ended up with 13 173 Seyfert galaxies for which we carry out the following SED modelling analysis.

We also looked for X-ray and radio photometric data points. However, the coverage at those wavelengths was not homogeneously tabulated in CDS or NED as in the selected bands in Table 6.1. We decided not to use X-ray and radio wavelengths as this will require more computational and time efforts for a few number of galaxies (only $\sim 0.01\%$ of the sample). In addition, currently X-CIGALE does not include a AGN radio component. We discuss the implications of not using X-ray data in Sect. 6.4.5.

6.2.2 SED Models

Parameter grids

We modelled the SEDs of the Seyfert galaxies with X-CIGALE (Yang et al. 2020). X-CIGALE is a modified version of CIGALE (Boquien et al. 2019), a SED fitting code based on an energy balance principle. The difference between X-CIGALE and CIGALE is the addition of i) an X-ray photometry module and, ii) a polar dust model in AGNs. These two enhancements help to connect the X-ray emission to the UV-to-IR SED, and account for dust extinction in the polar angles, respectively. The X-ray emission is helpful to constrain AGN intrinsic accretion power in the SED (Lyu & Rieke 2018; Toba et al. 2021), and the polar dust follow observational results from MIR interferometry (López-Gonzaga et al. 2016; Ramos Almeida & Ricci 2017).

We included six modules which account different galactic emission processes to fit the SEDs. The first module defines the star-formation history (SFH). We used a delayed SFH model for our sample of galaxies because this has shown a good agreement in different types of galaxies with ongoing or recent starburst events (Dietrich et al. 2018; Ramos Padilla et al. 2020), and can provide better estimates for physical pa-

Table 6.1 – *Photometric bands used in the SEDs modelling. The last column shows the number of galaxies detected in a given band.*

Mission or Survey	Band	Effective Wavelength [μm]	Number of galaxies
GALEX	FUV	0.152	6456
	NUV	0.227	9266
SDSS	u	0.354	12024
	g	0.477	12542
	r	0.623	12326
	i	0.762	12274
	z	0.913	11604
2MASS	J	1.25	7018
	H	1.65	6566
	Ks	2.17	8215
Spitzer	IRAC-1	3.6	4063
	IRAC-2	4.5	4048
	IRAC-3	5.8	458
	IRAC-4	8.0	447
	MIPS1	24.0	809
	MIPS2	70.0	225
	MIPS3	160.0	110
WISE	W1	3.4	13170
	W2	4.6	13165
	W3	12.0	12361
	W4	22.0	8295
IRAS	IRAS-1	12.0	462
	IRAS-2	25.0	634
	IRAS-3	60.0	979
	IRAS-4	100.0	722
Herschel	PACS-blue	70.0	265
	PACS-green	100.0	178
	PACS-red	160.0	303
	SPIRE-PSW	250.0	840
	SPIRE-PMW	350.0	476
	SPIRE-PLW	500.0	233

rameters such as star-formation rate (SFR) and stellar mass (Ciesla et al. 2015). The second module defines the single-age stellar population (SSP). We selected the standard Bruzual & Charlot (2003) model taking into account the initial mass function (IMF) from Chabrier (2003) and a metallicity close to solar. The dust attenuation law from Calzetti et al. (2000) is our third module. This module helps us control the UV attenuation with the colour excess E(B-V), and also the power-law slope (δ) that modifies the attenuation curve. The fourth module takes the dust emission in the SED into account. We modelled the dust emission following Dale et al. (2014), implementing a modified blackbody spectrum with a power-law distribution of dust mass at each temperature,

$$dM \propto U^{-\alpha} dU, \tag{6.1}$$

where U is the local heating intensity. We also included the nebular emission module although we did not change the default parameters.

The sixth and most important module for this work is the module that describes the AGN SED. For our experiments setups, we selected the two AGN modules available in X-CIGALE: A simple smooth torus (Fritz et al. 2006), and a two-phase (smooth and clumpy) torus (Stalevski et al. 2016, also known as SKIRTOR). For both models, we covered a larger sample of parameters for the viewing angle i and the fraction of AGN contribution to the IR luminosity $f_{\rm AGN}$ (Ciesla et al. 2015, eq. 1),

$$\mathrm{L}_{\rm IR}^{\rm AGN} = \mathrm{f}_{\rm AGN} \times \mathrm{L}_{\rm IR}^{\rm total}, \tag{6.2}$$

to investigate the effect of i in Seyfert galaxies. In addition, we set the extinction law of polar dust to the SMC values (Prevot et al. 1984), with a temperature of polar dust to 100 K (Mountrichas et al. 2021; Buat et al. 2021) and the emissivity index of polar dust to 1.6 (Casey 2012). The values for the colour excess of polar dust go from no extinction ($\mathrm{E}(B-V) = 0$) to $\mathrm{E}(B-V) = 1.0$ because $\mathrm{E}(B-V)$ cannot be well constrained only from the SED shape (Yang et al. 2020). However, adding $\mathrm{E}(B-V)$ as a free parameter can improve the accuracy of the classification type (Mountrichas et al. 2021).

The ratio of the outer to inner radii of the dust torus $\mathrm{R_{out}/R_{in}}$ and the optical depth at 9.7 μm τ are parameters that both AGN models share. The selection of these values changes in studies similar to this one on AGN galaxies depending on the AGN model used. When using the Fritz model it is common to use $\tau = 6.0$ and $\mathrm{R_{out}/R_{in}} = 60$ (e.g. Vika et al. 2017; Małek et al. 2018; Wang et al. 2020), while for SKIRTOR $\tau = 7.0$ and $\mathrm{R_{out}/R_{in}} = 20$ are often used (e.g. Yang et al. 2020; Mountrichas et al. 2021). We adopted the same values as in the literature, even though the τ values can be considered large, with the difference that we used $\mathrm{R_{out}/R_{in}} = 30$ for the Fritz model to make it more similar to SKIRTOR. We used the default geometrical parameters (power-law densities) in both models to focus on i, $f_{\rm AGN}$ and $\mathrm{E}(B-V)$. Finally, we tested two angle configurations: i) with viewing angles between 0°and 90°, and ii) using typical viewing angles of Type-1 and Type-2 AGNs of 30°(unobscured) and 70°(obscured). This comparison helps us to understand how important the viewing angle input parameter is in X-CIGALE.

In summary, we used the parameters and values given in Table 6.2 to define the grid of X-CIGALE SED models for the sample of Seyfert galaxies. For the remaining

parameters not shown in Table 6.2, we adopted the X-CIGALE default settings. We decided not to include the X-ray or radio modules due to the lack of homogeneous information for the selected sample of Seyfert galaxies (see Sect. 6.4.5). We assumed the redshifts from SMB in the SED fits.

Cleaning X-CIGALE fits

We ran another setup of X-CIGALE without the AGN module (hereafter No-AGN) in addition to the X-CIGALE setups with AGN models described in Table 6.2. The No-AGN setups helped us to identify bad-fittings and ambiguous cases where an AGN is not dominant in the SED, even though the galaxies are classified as Seyfert.

X-CIGALE minimises the χ^2 statistic and produces probability distribution functions for the grid parameters by assuming Gaussian measurement errors (Burgarella et al. 2005; Noll et al. 2009; Serra et al. 2011). In Fig. 6.2, we show an example of the SED fitting in one of the galaxies (Mrk 662) using the five different setups: smooth torus (Fritz setup from now on), smooth and clumpy torus (SKIRTOR setup from now on), smooth and clumpy torus with only two viewing angles (Fritz 30/70 and SKIRTOR 30/70 setups), and a model without AGN (No-AGN setup). The No-AGN setup (upper-right panel) shows a significant difference with the AGN setups in terms of reduced χ^2 ($\chi^2_{\rm red}$), which is expected as an AGN model is needed for most of our Seyfert galaxies. However, in some cases, No-AGN setup have a lower or equal $\chi^2_{\rm red}$ than AGN setups, meaning a worse fit with the AGN setup and/or a non-dominant AGN. Another slight difference in these SEDs is the contribution from the AGN (green dashed line) at 10μm, which varies depending on the best-fitting to a given setup. These differences can play a role in some physical parameters (e.g. attenuation), that may affect the classification type.

We compare the $\log(\chi^2_{\rm red})$ distribution for the SED setups in Fig. 6.3. There is a small difference in $\chi^2_{\rm red}$ between AGN setups (Fritz and SKIRTOR), with and average value of $\Delta\chi^2_{\rm red} = 0.147$, which shows that both setups fit the data similarly. Besides, we found that AGN and No-AGN setups have an average difference in $\log(\chi^2_{\rm red})$ of $\sim$0.4 dex, favouring AGN setups. For the No-AGN setup, if we use the same $f_{\rm AGN}$ value as in the AGN setups (which by construction have a $f_{\rm AGN} = 0$), then we can compare the difference in $\chi^2_{\rm red}$ when adding an AGN model in the SED. We found the smallest $\chi^2_{\rm red}$ differences at $f_{\rm AGN}$ below 0.2, while the largest differences are at $f_{\rm AGN} \sim 0.7$, when comparing the setups with and without AGN. For the SED setups with AGN, we found that galaxies with $f_{\rm AGN}$ between 0.2 and 0.8 have $\chi^2_{\rm red}$ values close to one. Therefore, for most galaxies outside this $f_{\rm AGN}$ range have poorer fittings. This differences shows the importance of adding the AGN model in the SED fitting and how $\chi^2_{\rm red}$ changes for non-dominant AGNs ($f_{\rm AGN} < 0.2$) and highly-dominant AGNs ($f_{\rm AGN} > 0.85$).

To better compare the No-AGN and AGN setups, we use the Bayesian Inference Criterion (BIC) to see if the AGN module is preferred for the fits, as in other CIGALE works (e.g. Buat et al. 2019). The BIC is defined as BIC $= \chi^2 + {\rm k} \times \ln({\rm N})$, with k the number of free parameters and N the number of data points used for the fit (Ciesla et al. 2018) and works as an approximation of the Bayes factor (Kass & Raftery 1995). Then, the difference between the setups can be calculated as ΔBIC =

Table 6.2 – X-CIGALE *grid parameter values adopted for the modelling described in Section 6.2.2*

Parameter	Values	Description
	Star formation history (SFH): Delayed	
$\tau_{\rm main}$	50, 500, 1000, 2500, 5000, 7500	e-folding time of the main stellar population model (Myr).
Age	500, 1000, 2000, 3000, 4000, 5000, 6000	Age of the oldest stars in the galaxy (Myr).
	Single-age stellar population (SSP): Bruzual & Charlot (2003)	
IMF	1	Initial Mass Function from Chabrier (2003).
Metallicity	0.02	Assuming solar metallicity.
	Dust attenuation: Calzetti et al. (2000)	
$E(B-V)$	0.2, 0.4, 0.6, 0.8	Color excess of the nebular light for the young and old population.
$E(B-V)_{\rm factor}$	0.44	Reduction factor for the $E(B-V)$ to compute the stellar continuum attenuation.
Power-law slope (δ)	-0.5, -0.25, 0.0, 0.25, 0.5	Slope delta of the power law modifying the attenuation curve.
	Dust emission: Dale et al. (2014)	
α	1.0, 1.5, 2.0, 2.5, 3.0	Alpha from the power-law distribution in Eq. 5.1.
	AGN models:	
i	0 – 90[a]	Viewing angle (face-on: $i = 0°$, edge-on: $i = 90°$).
$f_{\rm AGN}$	0.1 – 0.9 in steps of 0.05	Fraction of AGN torus contribution to the IR luminosity in Eq. 6.2
$E(B-V)_{\rm polar}$	0.0, 0.03, 0.1, 0.2, 0.3, 0.4, 0.6, 1.0	$E(B-V)$ of polar dust (fig 4 of Yang et al. (2020)).
$T_{\rm pd}$	100	Temperature of polar dust (eq. 10 of Yang et al. (2020))..
$\beta_{\rm pd}$	1.6	Emissivity index of polar dust (eq. 10 of Yang et al. (2020)).
	Fritz model (Fritz et al. (2006))	
$R_{\rm out}/R_{\rm in}$	30.0	Ratio of the outer to inner radii of the dust torus.
τ	6.0	Optical depth at 9.7 μm.
β	-0.50	Beta from the power-law density distribution for the radial component of the dust torus (eq. 3 of Fritz et al. (2006)).
γ	4.0	Gamma from the power-law density distribution for the polar component of the dust torus (eq. 3 of Fritz et al. (2006)).
Opening Angle (θ)	100.0	Full opening angle of the dust torus (fig 1 of Fritz et al. (2006)).
	SKIRTOR model (Stalevski et al. (2016))	
$R_{\rm out}/R_{\rm in}$	20.0	Ratio of the outer to inner radii of the dust torus.
τ	7.0	Optical depth at 9.7 μm.
p	1.0	Power-law exponent of the radial gradient of dust density (eq. 2 of Stalevski et al. (2012)).
q	1.0	Angular parameter for the dust density (eq. 2 of Stalevski et al. (2012)).
Δ	40	Angle between the equatorial plane and edge of the torus (half opening angle).

Notes: [a] We covered viewing angles between 0°and 90°in steps of 10°. We used the values closest to the predefined angle grid in X-CIGALE. For other setups, we used only 30°and 70°, taking into account that $i = 90 - \psi$ (with ψ the angle between equatorial axis and line of sight).

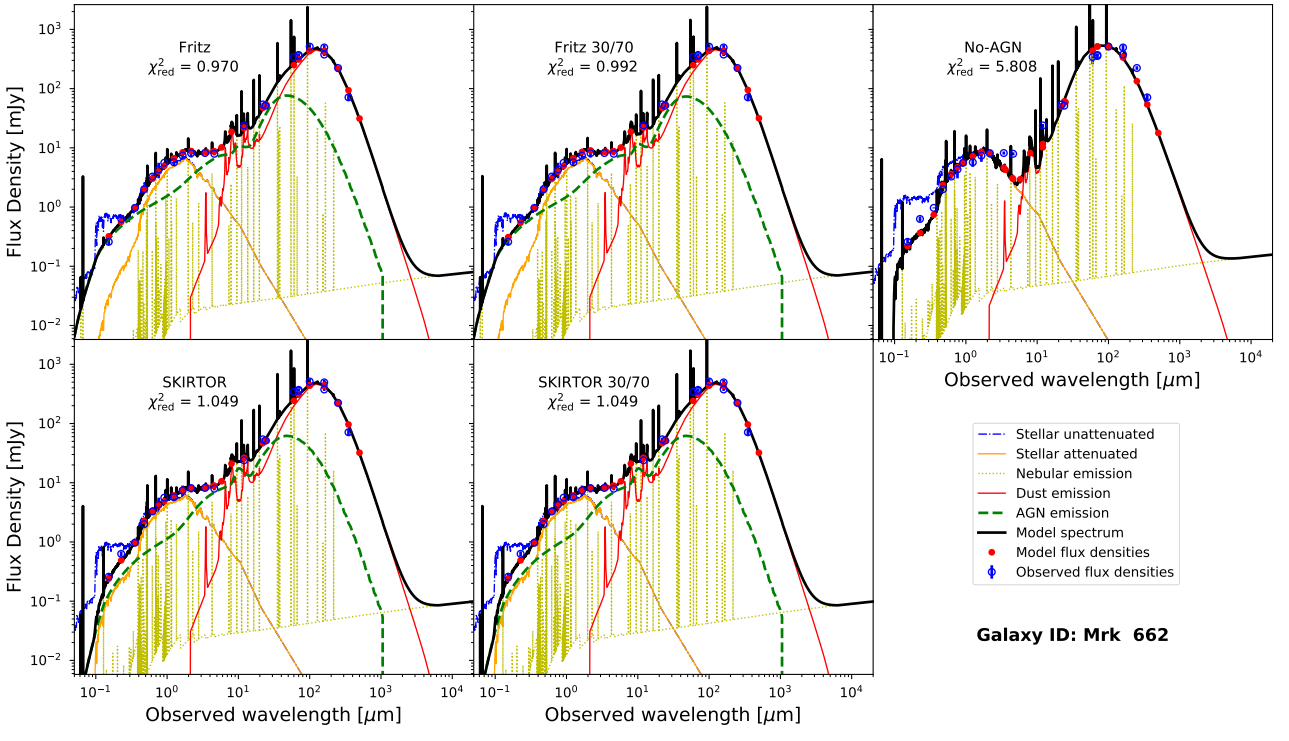

Figure 6.2 – *Example SEDs of the five SED fitting setups used in this work for the galaxy Mrk 662 at $z = 0.05$. No-AGN setup (upper-right panel) usually have higher reduced χ^2 ($\chi^2_{\rm red}$) values than AGN setups (other four panels). Each plot contains the contribution to the model spectrum (black line) of: nebular emission (gold dotted lines), attenuated (orange) and non-attenuated stellar emission (blue dot-dashed), dust emission (red solid), and AGN emission (green dashed). The red dots are the best model flux densities and the blue squares mark the observed flux densities with 1σ error bars.*

Table 6.3 – *Summary of the Seyfert samples used in this work. The original samples of Seyfert galaxies are in columns 2 (VCV) and 3 (SMB). Galaxies with photometry from the samples fulfilling our criteria are in column 4. The last four columns show the final counts for well-constrained SEDs in* X-CIGALE. *We show the counts from VCV and SMB classifications in columns 4-8 for Sy1, Sy2 and unclassified Seyfert rows. We show the counts for VCV classification for intermediate or other Seyfert galaxies. The last row shows the total number of galaxies in each of the samples.*

Seyfert	Samples		With	X-CIGALE AGN models			
Classification	VCV	SMB	Photometry	SKIRTOR	Fritz	SKIRTOR 30/70	Fritz 30/70
Seyfert 1	13 177	13 760	8 942 / 9 421	5 913 / 6 328	6 295 / 6 683	6 064 / 6 453	6 350 / 6 723
Seyfert 2	4 567	5 040	3 284 / 3 679	1 473 / 1 626	1 535 / 1 697	1 390 / 1 544	1 361 / 1 515
Unclassified Seyfert	87	121	54/73	27 / 38	28 / 36	25 / 34	28 / 37
Intermediate Seyfert	920	···	756	507	489	492	479
Alternative Seyfert	170	···	137	72	69	60	57
Total galaxies	18 921	18 921	13 173	7 992	8 416	8 031	8 275

$\chi^2_{\rm AGN} - \chi^2_{\rm No-AGN} + \ln(N)$, as we are just fixing the $f_{\rm AGN}$ to zero in the No-AGN setup. We adopt a positive evidence criterion for No-AGN setup (Salmon et al. 2016), meaning that galaxies with a ΔBIC ≥ 2 will prefer the No-AGN setup.

We imposed some constraints in the X-CIGALE estimated values to clean the set of derived parameters for this work. First, we used galaxies with a $\log(\chi^2_{\rm red})$ between -0.5 and 0.5 (grey dashed lines in Fig. 6.3) to avoid over and underestimations, respectively. Second, we selected galaxies where the AGN setups were preferred i.e. ΔBIC < 2. Finally, we selected galaxies where their estimated 1σ error in SFR was below one dex, to obtain reliable SFR estimations. Unfortunately, this last selection causes a bias against quiescent galaxies. These constraints led us to remove between 4 757 and 5 181 galaxies (depending on the AGN setup), from which: 69–75% galaxies were over or underestimated fits, 9–15% galaxies had a better fit with the No-AGN setup, and 31-38% galaxies where SFR was not well constrained.

To summarise, we present in Table 6.3 the total number of galaxies of the original samples (SMB and VCV), samples with photometry that meet our criteria for the SED fitting procedure, and the well-constrained fits with the X-CIGALE AGN models with respect to their Seyfert classification. In Appendix 6.5, we verify the quality of the fits for the main parameters studied in this work and the parameter space used for the fitting procedure by mock analysis. The mock analysis is a standard procedure included inside the CIGALE. A detailed description of this process can be found in Boquien et al. (2019).

Verification with other estimates

We verified the estimates from our procedure by comparing with a similar study done with CIGALE by Vika et al. (2017). Vika et al. (2017) uses a sample of 1 146 galaxies selected from the CASSIS spectroscopic sample (Lebouteiller et al. 2015) with good photometric coverage from UV to mid-IR in the redshift range $0 < z < 2.5$. As all these galaxies have been observed with *Spitzer*/IRS, the sample is biased to significantly brighter mid-IR galaxies. There are two main differences between the estimated physical parameters from Vika et al. (2017) and this work. The first difference is the way the photometry was retrieved. Vika et al. (2017) used specific catalogues that contain broad-band photometry for their sample of galaxies, while in this work we use data available in CDS and NED. Thus, we include additional information as the

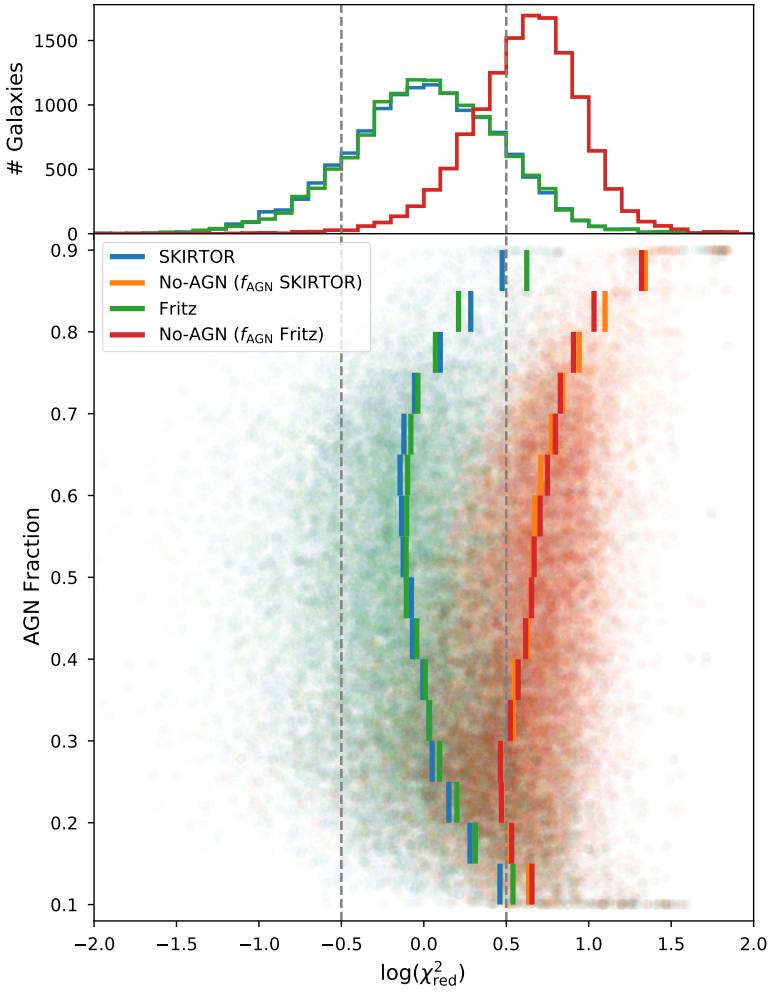

Figure 6.3 – *Reduced-chi-square (χ^2_{red}) distribution for the setups used in this work. Upper-panel: histogram of the χ^2_{red} distribution for SKIRTOR (blue), Fritz (green) and No-AGN (red) setups. Bottom-panel: χ^2_{red} against the estimated AGN fraction (f_{AGN}) for each galaxy (shaded dots) and the running median in bins of f_{AGN} values (solid lines). No-AGN setup ($f_{AGN} = 0$) is plotted assuming the f_{AGN} values of the AGN setups, thus showing slightly different median values with Fritz and SKIRTOR setups. We assumed that well fitted galaxies have a $\log(\chi^2_{red})$ between -0.5 and 0.5 (grey dashed lines).*

databases collect more broad-band photometry. The second difference is the assumed grid values for the SED modelling. Although the numerical values in most of the input parameters are not the same, here we mention the three most important differences between the grids. First, the IMF in this work comes from Chabrier (2003), while Vika et al. (2017) uses the IMF from Salpeter (1955). The use of the Chabrier IMF will lead to lower stellar mass values in this work with respect to Vika et al. (2017). Second, the parameter space for $f_{\rm AGN}$ in this work is finer sampled and more homogeneously distributed than the one from Vika et al. (2017). And third, the selected viewing angles for the AGN in Vika et al. (2017) are only $i = 0°$ and $i = 90°$ for the Fritz AGN model.

We selected the smooth torus model with two viewing angles (Fritz 30/70 setup) to compare the results from Vika et al. (2017), as it is the most similar model in this work. We cross-matched the Vika et al. (2017) catalogue with the Fritz 30/70 setup between 3″and we found 87 galaxies for the comparison in the range of $0.02 < z < 1.4$, with a median of $z = 0.13$. In Fig. 6.4, we present the comparison of six physical parameters between this work and Vika et al. (2017). In terms of SFR, AGN luminosity and dust luminosity, there is no clear difference between the estimates, the small systematic offsets are related to the different assumed values of the grid. For example, in both works the dust model from Dale et al. (2014) is used to estimate the dust luminosity, which depends mainly on the power-law parameter α of the mass distribution (Eq. 6.1). In Vika et al. (2017), they only use two values for α, while in this work we use five. On the contrary, the estimates on stellar mass are lower in this work mainly due to the adopted IMF, but the estimated median errors are similar between both works.

Age and $f_{\rm AGN}$ are the only physical parameters that are very different from Vika et al. (2017). In the case of $f_{\rm AGN}$, the estimates are constrained by the different grid selection. However, the estimated median error is lower in this work than in Vika et al. (2017), as we use more photometric bands and a finer $f_{\rm AGN}$ grid. In the case of the age, Vika et al. (2017) noted that age estimates are not well constrained in CIGALE, besides the differences in the grid values. This problem is also obvious from our estimates, which have uncertainties similar to those presented by Vika et al. (2017). In general, the physical parameters estimated in Vika et al. (2017) are similar to those presented in this work, validating our approach of obtaining data directly from astronomical databases.

6.3 Results

6.3.1 Feature selection

Recent advances in algorithms and machine learning techniques are helping to classify very complicated physical systems (Carleo et al. 2019; Virtanen et al. 2020). These classification tasks have covered the full range of galactic and extragalactic sources (e.g. Tamayo et al. 2016; Miettinen 2018; Jayasinghe et al. 2018; Bluck et al. 2020; Baqui et al. 2021). Nowadays, these methods are helping to classify astrophysical objects not only from reduced fluxes, but also from astronomical imaging surveys, where Type-1 AGN are separated from normal galaxies (Golob et al. 2021). Therefore,

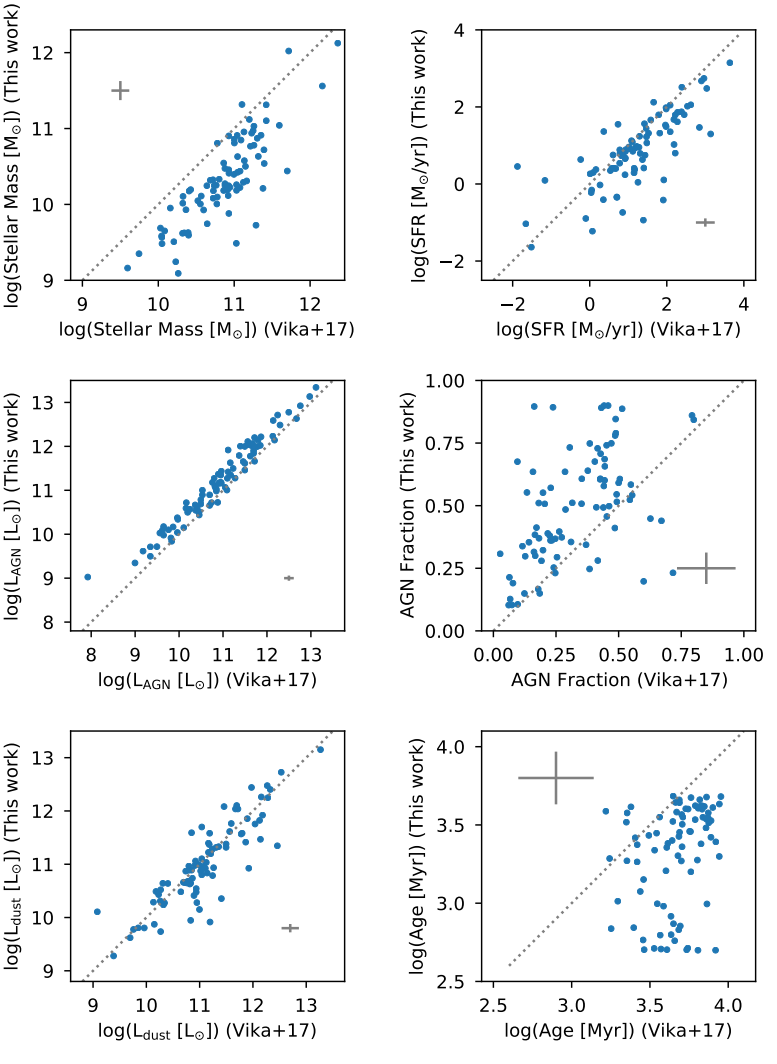

Figure 6.4 – *Comparison between the estimated physical parameters from this work and by Vika et al. (2017). The pointed line represents the 1:1 relation. The grey crosses represent the median estimated error for each of the parameters. In general, the estimates from this work agree with the results presented by Vika et al. (2017). Based on the median estimated errors, our results are better constrained than the ones from Vika et al. (2017).*

joining X-CIGALE physical parameters with these classification techniques could be useful to solve the AGN "zoo" of galaxies for different AGN scenarios.

However, X-CIGALE estimates more than 60 physical parameters depending on the number of modules included in the SED fitting. Therefore, it is necessary to select a smaller number of physical parameters which are the most informative for a classification task. We use a set of 29 737 estimates from the sum of the four X-CIGALE AGN setups (SKIRTOR, Fritz, and their respective 30/70 setups), where VCV and SMB share the same classification of Sy1 or Sy2. We split this set randomly into train and test subsets with a proportion of 80% and 20%, respectively. Then, we scale the subsets by subtracting the median and transforming according to the interquartile range. We perform this scaling to obtain classifications that are robust against outliers. We discard physical parameters directly related to inputs (e.g. redshift) and those dividing old and young stellar populations (e.g. old and young stellar masses).

We implement two machine learning ensemble techniques for the classification task: random forest and gradient boosting. Both ensemble techniques provide estimates using multiple estimators. The first technique is composed of a collection of trees randomly distributed where each one decides the most popular class (Breiman 2001). While the second technique takes into account "additive" expansions in the gradient descent estimation to improve the selection of the class (Friedman 2001). For the random forest we use the `scikit-learn` (Pedregosa et al. 2012) classifier `RandomForestClassifier`. For the gradient boosting we use the `XGBoost` (Chen & Guestrin 2016) classifier `XGBClassifier`. We tune the classifier's parameters using an estimator from `scikit-learn` that uses cross-validation in a grid-search `GridSearchCV`. The grid is defined with two parameters: the number of trees in the forest `n_estimators`, and the maximum depth of the tree `max_depth`. Values in the grid for `n_estimators` cover the range between 100 and 500 in steps of 100, while for `max_depth` the values cover the range between 10 and 40 in steps of 5. We use the F1-score to evaluate the predictions on the test set. The F1-score is the harmonic mean of precision and recall, where precision is the fraction of true positives over true and false positives and recall is the fraction of true positives over true positives and false negatives. With the `GridSearchCV` estimator, the best values for the `RandomForestClassifier` are `n_estimators` = 200 and `max_depth` = 25, while for `XGBClassifier` are `n_estimators` = 300 and `max_depth` = 25.

We apply a recursive feature elimination and cross-validation selection (RFECV) to select the ideal number of physical parameters to study. We perform 10 k-fold cross-validations with `RandomForestClassifier` and `XGBClassifier`. In Figure 6.5, we present the feature importance (the score of a feature in a predictive model) for both classifiers after the RFECV has been applied. We find that five physical parameters contribute the most in the classification task. These physical parameters are observed AGN disc luminosity, AGN viewing angle, AGN polar dust $E(B-V)$, e-folding time ($\tau_{\rm main}$), and the colour excess. Though we find that the feature importance is low for SFR and $f_{\rm AGN}$, the classification scores improve when we include these parameters, which are important when comparing to observational results. Thus, we decided to focus on these previous seven parameters to assess the impact of the viewing angles in the samples of Seyfert galaxies. In Sect. 6.4.1, we describe the role that these physical parameters play in the classification task.

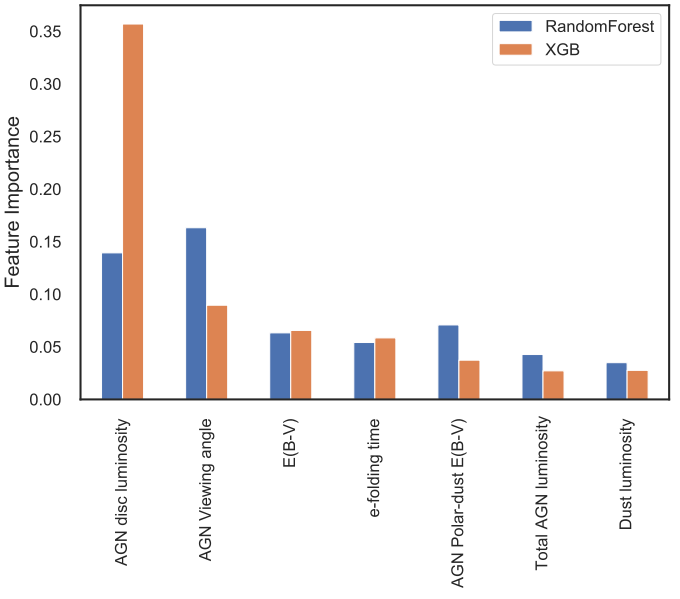

Figure 6.5 – *Feature importance scores for the seven most important physical parameters in both* `RandomForestClassifier` *and* `XGBClassifier`*. Parameters with scores below 0.03 (i.e. < 3%) are not presented. The observed AGN disc luminosity is the most important physical parameter to describe Seyfert types.*

It is necessary to clarify here that the L_{AGN}^{disc} is the observed total luminosity of the accretion disc in X-CIGALE. The AGN accretion power is the term we use to refer to the intrinsic luminosity of the accretion disc.

6.3.2 Comparison of X-CIGALE outputs from different SED fitting setups

We compare the seven selected physical parameters using the density distribution of the selected setups and Seyfert type samples. We quantify the difference between the distributions using the two-sample Kolmogorov–Smirnov (KS) test. The KS test checks the null hypothesis that two distributions are drawn from the same underlying distribution. We reject the null hypothesis if D (the distance between cumulative distributions) is higher than the critical value at a significance level of $\alpha = 0.05$ (e.g., $P < 0.05$). We visualise the distributions using a bandwidth of the density estimator following the Scott's Rule (Scott 2015). The visualisation and the KS statistics D tell us how different the samples are.

AGN setup comparison

We compare the AGN setups (SKIRTOR, Fritz, and their respective 30/70 setups) before comparing the results between different Seyfert types. We present the probability distribution functions for the parameters and their errors in Fig. 6.6.

First, we check the differences between the setups covering all viewing angles and setups with viewing angles of only 30°and 70°. We do not observe a clear difference between these two sets of setups in most of the physical parameters. Their main discrepancy in the distributions is in the viewing angle, as expected. In setups with 30°and 70°, we find galaxies with viewing angles around 50°. These values are related to the bayesian nature of the estimations, as seen in the estimated error. Something different happens in setups with the full range of viewing angles used for the SED modelling. The distribution of the viewing angle peaks at around 20°while at larger angles ($\gtrsim 45°$) the distribution is almost flat.

Second, we check the differences between SKIRTOR and Fritz setups. In this case, small differences are noticeable in some physical parameters (e.g. observed AGN disc luminosity and polar dust). However, both kinds of setups follow similar trends in all parameters. If we observe estimated errors, all of them are well constrained, partly due to our $\chi^2_{\rm red}$ selection criteria.

Finally, the null hypothesis in the KS test is almost always rejected in all the setups. The only cases where the null hypothesis cannot be rejected are at Fritz and Fritz 30/70 setups for the $\tau_{\rm main}$ and SFR parameters ($D \sim 0.02$, $P \sim 0.7$), and at SKIRTOR and SKIRTOR 30/70 setups for the SFR parameter ($D \sim 0.02$, $P \sim 0.1$). We find $D < 0.1$ in most parameters when comparing all setups except for the viewing angle due to the way we designed the setups, as expected. Variations in the setups will only give different individual results, but in general, the setups will be similar when interpreting the physical results in Seyfert galaxies.

Seyfert types 1 and 2 comparison

According to the AGN unified model, the viewing angle is the main physical parameter to classify Seyfert galaxies into Type-1 AGNs (face-on) and Type-2 AGNs (edge-on). In X-CIGALE, SED models which include AGN follow the AGN-unification scheme (Yang et al. 2020). The viewing angle estimated by the models should coincide with the classification scheme of Sy1 and Sy2 galaxies. For this analysis, we use galaxies where Seyfert classifications from VCV and SMB agree. In Fig. 6.7, we present the probability distribution functions for the seven selected parameters now separating the galaxies in Sy1 and Sy2 for the SKIRTOR and Fritz setups. The distribution of viewing angles for Sy1 coincides with the expected low values of face-on galaxies. For Sy2 galaxies, the viewing angle distribution extends over a wide range of angles, but agrees with a viewing angle close to edge-on galaxies.

KS statistic values D show: i) independent of the used AGN setup, the distributions of the parameter values are similar for the same Seyfert galaxy types, and ii) the different Seyfert types can be well separated as the distributions are different for both setups. The only exception is in terms of the $E(B-V)$, where D values for Sy1 galaxies depend on the used setup.

In terms of $f_{\rm AGN}$, Sy2 galaxies have mainly lower values than Sy1 galaxies. We observe something similar for the parameters of SFR and e-folding time ($\tau_{\rm main}$), with low values typically associated to Sy2 galaxies. Although for these two parameters, the difference with Sy1 could be due to the strong UV / optical emission that bias

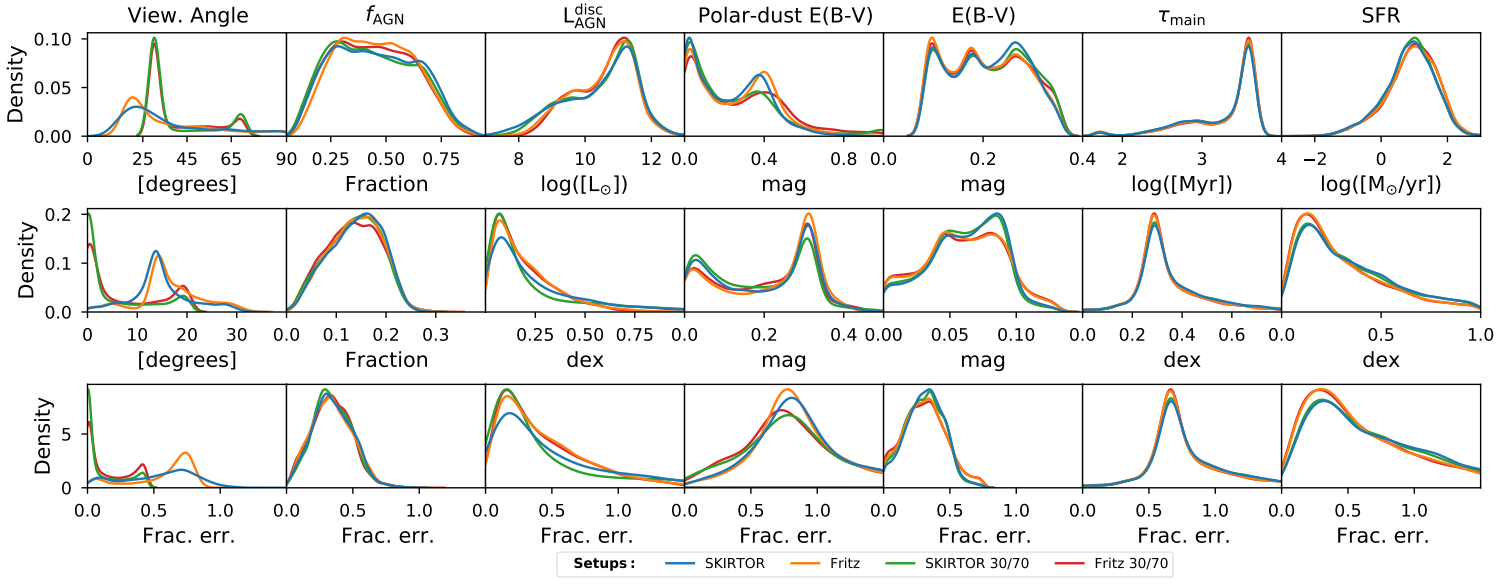

Figure 6.6 – *Probability density functions for the estimated parameters (upper panels), their respective errors (middle panels) and their relative fractional errors (lower panels). We compare the SKIRTOR (blue), SKIRTOR 30/70 (green), Fritz (orange), and Fritz 30/70 (red) setups. The difference between the setups that use the full range of viewing angles (SKIRTOR and Fritz) and those that use only two angles (30/70 setups) is significant only in the viewing angle, as expected.*

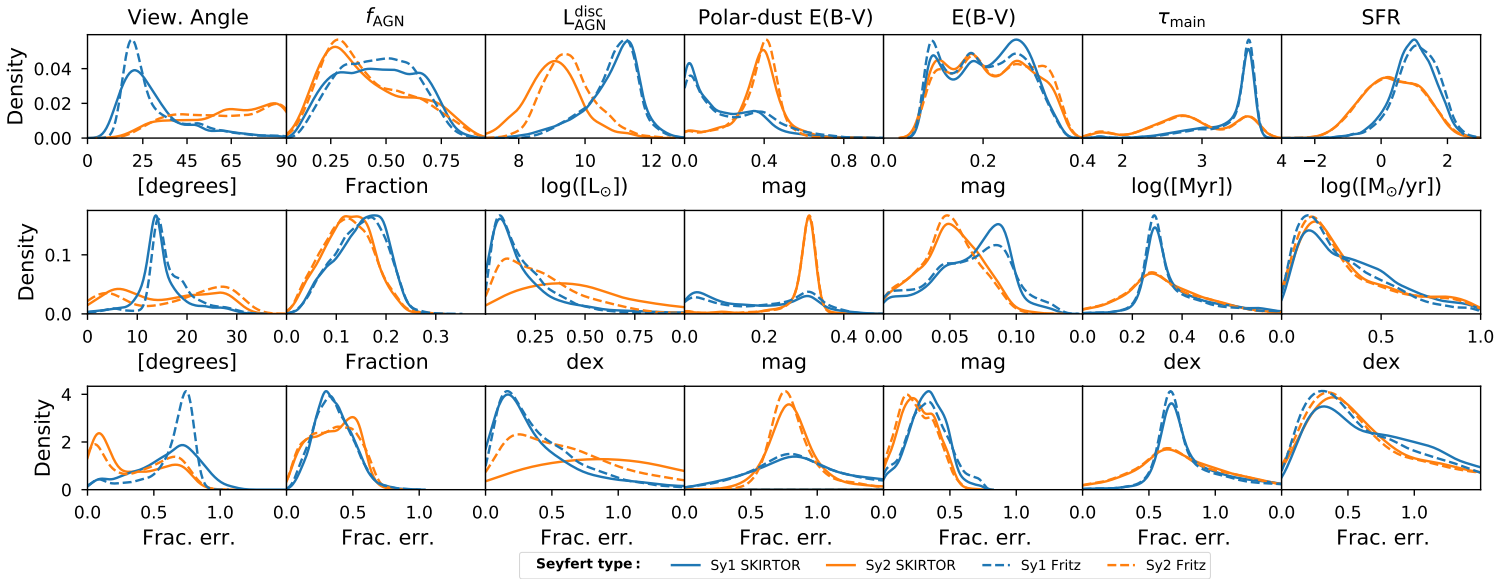

Figure 6.7 – *Probability density functions for the estimated parameters (upper panels), their respective errors (middle panels) and their relative fractional errors (lower panels). We compare the Seyfert types 1 (blue) and 2 (orange) for the SKIRTOR (solid) and Fritz (dashed) setups. A clear difference is observed in some of these parameters between Sy1 and Sy2 (e.g. viewing angle, polar dust $E(B-V)$ and observed AGN disc luminosity).*

these parameters. Values of the polar-dust show that in Sy2 galaxies the emission is obscured, in contrast to the Sy1 where $E(B-V)=0$ is more common.

The most interesting result in this comparison is the significant difference in the L_{AGN}^{disc} between the two Seyfert types. Most Sy2 have L_{AGN}^{disc} below $\sim 10^{10} L_{\odot}$. The opposite happens for Sy1 galaxies where most L_{AGN}^{disc} are above $\sim 10^{10} L_{\odot}$. This result is also verified with the KS statistic D, where higher D values are found when comparing Sy1 and Sy2 samples for i and L_{AGN}^{disc}. In a classification task, this latter physical parameter might be more informative than others, as we have seen with the feature importance score (Sect 6.3.1 and Fig. 6.5). We test this idea when predicting the classification type in unclassified and discrepant Seyfert galaxies (Sect. 6.3.5).

We verify the impact of missing bands in the SED fitting in our sample of galaxies for the estimated parameters. The differences we found between the probability density functions with and without a given band can be explained by the way the Seyfert types are distributed in the sample. For example, in the 2MASS bands, a third of the galaxies detected in these bands are classified to be Seyfert 2, while in the cases where we do not have these bands most of the galaxies (∼90%) are of Seyfert 1 type. Most of the galaxies not detected by 2MASS follow the trends of Seyfert 1 galaxies, with high AGN disc luminosities, lower viewing angles and polar dust close to zero. This shows that even if some bands are missing, the SED fitting procedure still gives similar values for the physical parameters compared to other galaxies which have the same classification. In other words, the lack of some data does not necessarily impact the results of this work significantly, but it could be important in individual cases, which is out of the scope of this work.

Intermediate Seyfert types comparison

From the results presented in Figs. 6.6 and 6.7, we observe a small difference between the AGN setups. Thus, we select the SKIRTOR setup to visualise the difference between intermediate Seyfert types. These intermediate galaxy types are classified in subgroups following Osterbrock (1977, 1981) with the quantitative approach of Winkler (1992). The subgroups are divided using the ratio between Hβ and [O III] fluxes (R) and the spectral profiles of the Balmer lines (see also Véron-Cetty & Véron (2010)). In this classification scheme Seyfert galaxies are: Sy1.0 if $R > 5$, Sy1.2 if $2.0 < R < 5.0$, Sy1.5 if $0.33 < R < 2.0$, Sy1.8 if $R < 0.33$ with a broad component in Hα and Hβ, and Sy1.9 if the broad component is visible in Hα but not in Hβ.

We show in Fig. 6.9 the probability density functions of the physical parameters for the intermediate Seyfert types. Interestingly, these intermediate types coincide with the picture observed in Sy1 and Sy2 in most of the physical parameters. In some parameters, a possible numerical sequence from Sy1 to Sy2 can be observed (Sy1.0>Sy1.2>Sy1.5>Sy1.8>Sy1.9). Three parameters show this sequence in their KS statistic D values: viewing angle, observed AGN disc luminosity and the polar dust $E(B-V)$. For the viewing angle i, Sy1.0, Sy1.2 and Sy1.5 tend to estimate values around 25°, while for Sy1.8 and Sy1.9 there are more i values above 45°. For the L_{AGN}^{disc}, the density functions for Sy1.8 and Sy1.9 peak below 10^{10} $L_{\odot}$, Sy1.5 peaks at $\sim 10^{10}$ $L_{\odot}$, and Sy1.2 and Sy1.0 peak at values above $\sim 10^{10}$ $L_{\odot}$. The polar dust in Sy1.0 and Sy1.2 shows in general no extinction, Sy1.5 shows mild values (0.1-0.3), and

Sy1.8 and Sy1.9 peak at around 0.4. These results agree with the expected behaviours of the transition between Sy1 and Sy2 type galaxies.

Other physical parameters (e.g. $f_{\rm AGN}$ or $\tau_{\rm main}$) also show similarities between close intermediate Seyfert types, but not as clear as the parameters described before. This suggests that AGN SED models, as the ones available in X-CIGALE, can estimate the possible transitional phase between different Seyfert types. In Sect. 6.4.2, we discuss a possible explanation for this transitional phase. In addition, in Sect. 6.4.4, we discuss the effect of using these AGN classifications when interpreting our results.

6.3.3 Redshift behaviour/evolution

Separating Seyfert galaxies in Type-1 or Type-2 is very important to understand the nature of these types of galaxies. In previous section (Sect. 6.3.2), we notice that some physical parameters could be used to separate the two Seyfert types. Now, we verify if the separation between Seyfert types holds and/or evolves with redshift. In Fig. 6.8, we present the evolution of the SED physical parameters as a function of redshift. For this figure, we use estimates from the SKIRTOR setup and separate the Seyfert types using the classifications from SMB.

In the upper panel of Fig. 6.8, we notice that the number of classified Sy2 galaxies are almost always below the number of classified Sy1 galaxies. Most of these classifications come from redshifts below $z \sim 0.5$, where spectroscopic information of the local Universe is more readily available, compared to high redshift galaxies. In the lower panels, the median of the physical parameters is similar between the two Seyfert types for $f_{\rm AGN}$, $E(B-V)$, $\tau_{\rm main}$ and SFR. However, the estimated values for the viewing angle, observed AGN disc luminosity and polar dust separate the two Seyfert types.

The viewing angle does not evolve with redshift, as expected. In general, the viewing angle for Sy1 galaxies is mostly located at ∼25°while for Sy2 galaxies is at ∼65°. Thus, we can use the value of 45°to separate the two types of Seyfert galaxies. For $\rm L_{AGN}^{disc}$, the difference is always above 0.8 dex (at $z \sim 0$) and can go up to ~ 2.3 dex at $z \sim 0.43$. We define a separation limit with the median values of the separation between Seyfert type as linear relation

$$\log\left(\rm L_{AGN}^{disc}\right) = (9.20 \pm 0.08) + (8.67 \pm 0.61) \times \log\left(1+z\right), \tag{6.3}$$

where z is redshift. Finally, for the polar dust, the minimum difference of ∼0.1 occurs at $z < 0.1$ and increases with redshift because most of the Sy2 galaxies above $z = 0.2$ have an estimate of 0.4 compared to values close to zero in Sy1, as expected. We do not define a separation for the polar dust due to the similarity of the estimates at $z < 0.1$.

These three parameters (viewing angle, observed AGN disc luminosity and polar dust) are tightly related in X-CIGALE (Yang et al. 2020). However, polar dust have larger uncertainties in the estimations, as show in Figs. 6.6–6.9. On the other side, $\rm L_{AGN}^{disc}$ could be a more robust parameter in classification tasks than the viewing angle. In Sy1 galaxies the median contribution from the observed disc luminosity to the total AGN luminosity is $\sim 52\%$, while for Sy2 the median contribution is around $\sim 4\%$. Furthermore, as we see in Figs. 6.7 and 6.9, the separation is more evident between

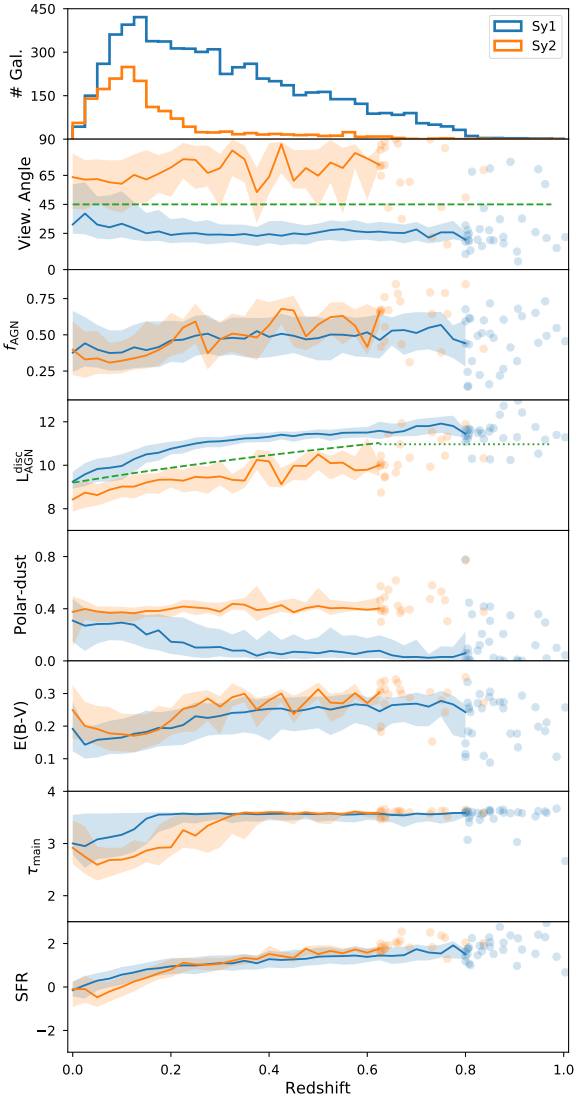

Figure 6.8 – *Redshift evolution of the physical parameters for the sample of Seyfert galaxies using the SKIRTOR setup. In the uppermost panel, we present the histogram of the Sy1 and Sy2 galaxies in terms of redshift. At redshift higher than 0.6, we find only a few Sy2 galaxies. At redshift above 0.8, the number of Sy1 galaxies reduces to tens of galaxies. In the rest of the panels, we show the running median of the physical parameters in Sy1 (blue) and 2 (orange), with the range between the 25th and 75th percentiles as shaded regions. We calculate the running median for bins with at least ten galaxies. For bins with fewer than ten galaxies, we plot the galaxies as scatter points. For the viewing angle we include the value of $i = 45°$ that separates the two Seyfert samples (green dashed line). We also include the separation limit for the two samples using L_{AGN}^{disc}, assuming the linear relation described in Eq. 6.3 (green dashed line) and a constant of $\log\left(L_{AGN}^{disc}\right) = 10.97$ at $z > 0.6$ (green pointed line). Galaxies with $z > 1$ are not shown.*

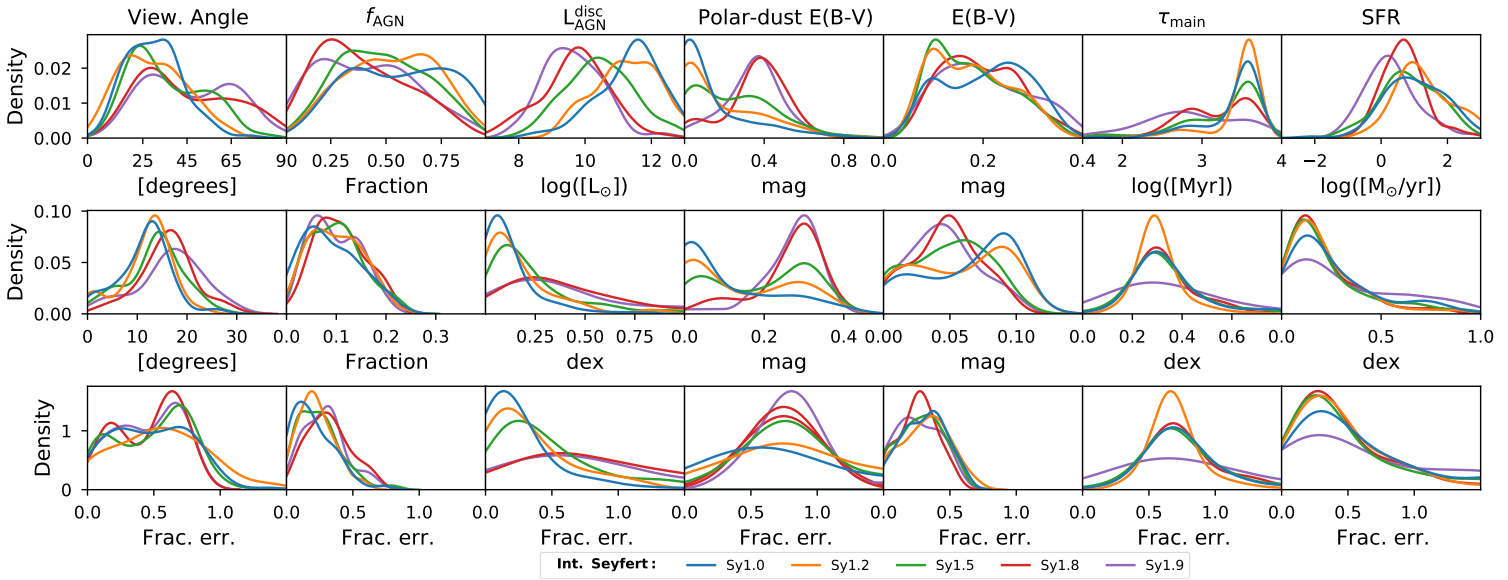

Figure 6.9 – *Probability density functions for the estimated parameters (upper panels), their respective errors (middle panels) and their relative fractional errors (lower panels). We compare the intermediate Seyfert types 1.0 (blue), 1.2 (orange), 1.5 (green), 1.8 (red), and 1.9 (purple) for the SKIRTOR setup. A transition between Sy1 and Sy2 galaxies is observed in the viewing angle, polar dust $E(B-V)$ and observed AGN disc luminosity.*

Seyfert types in observed AGN disc luminosity than in the viewing angle. The extent of the viewing angles and their uncertainties for Sy2 galaxies calls into question the use of this parameter to assess type in SED models.

6.3.4 Classifiers in Seyfert galaxies

We compare different classifiers to test if the physical parameters estimated by the SED modelling are useful to classify AGN galaxies. We define two different scenarios to do the classifications: i) using individual X-CIGALE physical parameters, ii) using ensemble methods with X-CIGALE selected physical parameters. For the first scenario, we use the two limits defined for i and L_{AGN}^{disc}, $i = 45°$and Eq. 6.3 respectively, as illustrated in Fig. 6.8. For the second scenario, we use two machine learning ensemble methods: `RandomForestClassifier` and `XGBoostClassifier`, to be consistent with the assumptions in the initial selection of physical parameters (Sect. 6.3.1). In the second scenario, we randomly split the sample of galaxies in train and test sets with a contribution of 80% and 20%, respectively. We use three different metrics to compare and test the quality of these binary classifications: Matthews correlation coefficient (MCC, Matthews 1975), F1-score and accuracy. MCC is the correlation coefficient between observed and predicted classifications, the F1-score is the harmonic mean of precision and recall (as described in Sect. 6.3.1), while accuracy is the fraction of samples correctly classified. These metrics are usually used in binary classifications, however they can be sensitive to imbalanced data (Tharwat 2020). Thus, we assess the scenarios by defining a baseline using the `DummyClassifier` from `scikit-learn`. This classifier respects the class distribution (i.e. stratified with Seyfert types) and generates random predictions, so it works as a sanity check. Therefore, we use five classifiers with three metrics to test the classification methods.

We use the subset of Seyfert galaxies where the classifications (of Type-1 and Type-2) were the same. We apply the different classifiers described previously for each AGN setup and present their metrics in Table 6.4. We notice that all classifiers outperform the baseline. In general, the best AGN setup in all the classifiers is SKIRTOR 30/70. In the scenario where we use individual X-CIGALE physical parameters, L_{AGN}^{disc} is in most cases a better discriminator than the viewing angle i. In the scenario where we use ensemble methods, both classifiers are similar but better than the L_{AGN}^{disc} and i, as expected. Part of the success of the SKIRTOR 30/70 AGN setup with individual classifiers is the assumption of using two viewing angles that follows the edge-on and face-on geometrical configurations. However, if i is the most robust parameter in the classification task, we would also expect higher numbers in the metrics for Fritz 30/70, which is not the case. For the classifier using L_{AGN}^{disc}, we find that most of the metrics between SKIRTOR and SKIRTOR 30/70 are not far from each other. These results mean that L_{AGN}^{disc} in SKIRTOR setups can be used to assess Seyfert type of a galaxy. Nevertheless, this classifier will not achieve as accurate predictions as the ensemble methods used in this work. In Sect. 6.4.4, we discuss the main drawback of using these classifiers in AGNs with SED models.

Table 6.4 – *Prediction metrics for galaxies with the same classification in SMB and VCV. The baseline is defined from a random classifier as mention in the text. In bold, we highlight the highest metric value for a given classifier. In general, the SKIRTOR 30/70 setup is the one that best separates Sy1 and Sy2 galaxies, although SKIRTOR metrics are good too.*

Classifier	Setup	Metrics		
		MCC	F1	Accu.
Viewing angle (i)	SKIRTOR	0.541	0.763	0.827
	Fritz	0.492	0.737	0.808
	SKIRTOR 30/70	**0.580**	**0.781**	**0.846**
	Fritz 30/70	0.530	0.746	0.817
AGN disk luminosity (L_{AGN}^{disc})	SKIRTOR	**0.635**	**0.808**	**0.860**
	Fritz	0.542	0.769	0.842
	SKIRTOR 30/70	0.621	0.801	**0.860**
	Fritz 30/70	0.535	0.764	0.848
Random Forest	SKIRTOR	0.717	0.856	0.907
	Fritz	0.732	0.865	0.913
	SKIRTOR 30/70	**0.737**	**0.866**	**0.914**
	Fritz 30/70	0.646	0.822	0.900
XGBoost	SKIRTOR	0.666	0.833	0.896
	Fritz	0.680	0.840	0.898
	SKIRTOR 30/70	**0.716**	**0.857**	0.910
	Fritz 30/70	0.707	0.853	**0.912**
Baseline	All	-0.003	0.499	0.691

6.3.5 Predictions on unclassified and discrepant Seyfert galaxies

We test if the estimation from the L_{AGN}^{disc} and the machine learning classifiers are robust and useful to predict the unclassified and discrepant cases in Seyfert galaxies. We compare the Seyfert type predictions of these classifiers with the literature. We gather information about other activity type classifications, outside VCV and SMB where possible, and we present this information in Table 6.7. We discuss the classifications for the 59 galaxies (14 unclassified and 45 discrepant) in the following paragraphs.

First, we focus on unclassified Seyfert galaxies by neither VCV nor SMB, as shown in Table 6.5. From the 14 unclassified Seyfert galaxies with the L_{AGN}^{disc} and ML methods: i) one galaxy (6dFGS gJ234635.0-205845) is classified for both methods as Sy2 galaxies, ii) seven are classified for both methods as Sy1 galaxies, and iii) six have mixed classifications in the two methods. From the literature classifications, two galaxies (LEDA 1485346 and MCG+00-11-002) coincides with the Seyfert classification. For MCG+03-45-003, we are not able to compare it with our classifiers because classifications change depending on the AGN setup used. In addition, we found a mixed blazar (2MASX J12140343-1921428), five quasar (CADIS 16-505716, LEDA 3095610, LEDA 3096762, QSO B1238+6232 and [HB93] 0248+011A), and a composite galaxy (2MASX J23032790+14434). We assume these seven galaxies cannot be Seyfert galaxies. However, we can compare the prediction's similarity with Type-1 AGN. For the mixed blazar and quasars (QSO), the L_{AGN}^{disc} and ML methods classify them as Sy1 types (except for QSO B1238+6232 with L_{AGN}^{disc}), which are close to Type-1 AGNs in

Table 6.5 – *Classification types in unclassified Seyfert galaxies. The second and third columns show the Seyfert type obtained using the* L_{AGN}^{disc} *and machine learning ensemble methods, respectively. The fourth column shows the assumed classification from the literature. Finally, the last column shows if the previous classifiers are similar with the AGN type in the literature. In case* L_{AGN}^{disc} *and ML classifications are different split the last column in two options.*

Object ID	Classification type			
	L_{AGN}^{disc}	ML	Literature	Sim.
2MASX J12140343-1921428	Sy1	Sy1	Mix. blazar	Y
2MASX J18121404+2153047	Sy2	Sy1	–	–
2MASX J21560047-2144325	Sy2	Sy1	–	–
2MASX J23032790+1443491	Sy1	Sy1	Composite	N
6dFGS gJ234635.0-205845	Sy2	Sy2	–	–
CADIS 16-505716	Sy1	Sy1	QSO	Y
ESO 373-13	Sy2	Sy1	–	–
LEDA 1485346	Sy1	Sy1	Sy1	Y
LEDA 3095610	Sy1	Sy1	QSO	Y
LEDA 3096762	Sy1	Sy1	QSO	Y
MCG+00-11-002	–[a]	Sy1	Sy1	–/Y
MCG+03-45-003	–[a]	–[a]	Sy2	–
QSO B1238+6232	Sy2	Sy1	QSO	N/Y
[HB93] 0248+011A	Sy1	Sy1	QSO	Y

[a] The classification changes depending on the AGN setup used.

the unified model scheme. Thus, 8/14 of the galaxies in the unclassified subset have similar classifications to the AGN type. Therefore, it may be possible to distinguish the AGN type from the classifiers presented in this work.

Second, we focus in the 45 discrepant galaxies that are unclassified in VCV but classified in SMB, and vice versa, shown in Table 6.6. From this sample, we find that 36 of them (80%) are mainly classified as Sy1 while only two are Sy2 with the L_{AGN}^{disc} and machine learning ensemble methods. For the other seven galaxies the classifications are still discrepant. We compare the similarity of the predictions of the L_{AGN}^{disc} and machine learning ensemble methods with the literature. We assume that: i) mixed blazar, BZQ, QSO, NLSy1, Sy1.5 and Sy1.2 galaxies are Type-1 AGN; ii) Sy1.8 are Type-2 AGN; and iii) LINERs cannot be classified as Type-1 nor Type-2 AGN. From galaxies with literature classifications (42 galaxies), we find that more than half of these galaxies ($\sim$ 59%) are classified as QSO, and only 11 galaxies ($\sim$ 26%) are classified as a Sy1. From the seven galaxies with still discrepant classifications five have other literature classifications. From those five, two galaxies have AGN types similar to those in the literature when L_{AGN}^{disc} is used as a classifier, and only one galaxy when we use machine learning ensemble methods. For the remaining galaxies with literature classifications (37 galaxies), we find that $\sim$ 95% of discrepant galaxies have AGN types similar to those in the literature when both classifiers coincide in the classification.

However, most of the results come from Type-1 AGNs and less than half of these

Table 6.6 – *Classification types in discrepant Seyfert galaxies. Columns are similar to Table 6.5.*

Object ID	Classification type			
	L_{AGN}^{disc}	ML	Literature	Sim.
2E 2294	Sy1	Sy1	QSO	Y
2E 2628	Sy1	Sy1	QSO	Y
2E 3786	Sy1	Sy1	QSO	Y
2MASS J00423990+3017514	Sy1	Sy1	Sy1	Y
2MASS J01341936+0146479	Sy1	Sy1	QSO	Y
2MASS J02500703+0025251	Sy1	Sy1	QSO	Y
2MASS J08171856+5201477	Sy2	Sy1	Sy1	N/Y
2MASS J09393182+5449092	Sy1	Sy1	QSO	Y
2MASS J09455439+4238399	Sy1	Sy1	NLSy1	Y
2MASS J09470326+4640425	Sy1	Sy1	QSO	Y
2MASS J09594856+5942505	Sy1	Sy1	QSO	Y
2MASS J10102753+4132389	Sy1	Sy1	QSO	Y
2MASS J10470514+5444060	Sy1	Sy1	QSO	Y
2MASS J12002696+3317286	Sy1	Sy1	QSO	Y
2MASS J15142051+4244453	Sy1	Sy1	QSO	Y
2MASSI J0930176+470720	Sy1	Sy1	QSO	Y
2MASX J02522087+0043307	Sy1	Sy1	QSO	Y
2MASX J02593816+0042167	Sy1	–[a]	QSO	Y/–
2MASX J06374318-7538458	Sy1	Sy1	Sy1	Y
2MASX J09420770+0228053	Sy1	Sy1	LINER[b]	N
2MASX J09443702-2633554	Sy1	Sy1	Sy1.5[b]	Y
2MASX J09483841+4030436	Sy2	Sy2	Sy1	N
2MASX J10155660-2002268	Sy1	–[a]	Sy1	Y/–
2MASX J10194946+3322041	Sy1	Sy1	NLSy1	Y
2MASX J15085291+6814074	Sy1	Sy1	Sy1	Y
2MASX J16383091-2055246	–[a]	Sy2	NLSy1	–/N
2MASX J21033788-0455396	Sy1	Sy1	Sy1	Y
2MASX J21512498-0757558	Sy2	Sy2	–	–
2MASX J22024516-1304538	Sy1	Sy1	Sy1	Y
2dFGRS TGN357Z241	Sy1	Sy1	QSO	Y
3C 286	Sy1	Sy1	Mix. blazar	Y
6dFGS gJ034205.4-370322	Sy1	Sy1	Mix. blazar	Y
6dFGS gJ043944.9-454043	Sy1	Sy1	QSO	Y
6dFGS gJ084628.7-121409	Sy1	Sy1	Sy1	Y
CTS 11	Sy1	Sy1	NLSy1	Y
HE 0226-4110	Sy1	Sy1	QSO	Y
ICRF J025937.6+423549	Sy1	Sy1	Mix. blazar	Y
ICRF J081100.6+571412	Sy1	Sy1	QSO	Y
ICRF J100646.4-215920	Sy1	Sy1	Mix. blazar	Y
ICRF J110153.4+624150	Sy1	Sy1	BZQ	Y
ICRF J135704.4+191907	Sy1	Sy1	BZQ	Y
IRAS 10295-1831	Sy1	Sy1	Sy1	Y
Mrk 1361	–[a]	Sy2	–	–
PB 162	–[a]	Sy1	–	–
UGC 10683	Sy2	–[a]	Sy1	N/–

Notes: [a] The classification changes depending on the AGN setup used.
[b] Decision based on VCV.

galaxies end up being Seyfert galaxies. The unclassified and discrepant Seyfert type classifications, presented in Tables 6.5 and 6.6, only show part of the predictability potential of these methods but support the usefulness of the classifiers in this work. These results, together with Table 6.4, show that we can classify AGN types correctly using the physical parameters estimated from broad-band SED fitting using X-CIGALE. In a future study, we expect to use a more complete set of intermediate numerical values for Seyfert galaxies to test their classification.

6.4 Discussions

In this section, we further examine the results of this work. First, we discuss the importance of the viewing angle in the AGN SED models (Sect. 6.4.1) and how they affect current AGN studies (Sect. 6.4.2). Then, we focus on the AGN types in terms of physical parameters as $f_{\rm AGN}$ and SFR (Sect. 6.4.3), and their classifications (Sect. 6.4.4). Finally, we examine the effect of not including X-ray data in the SED AGN models (Sect. 6.4.5).

6.4.1 The role of the viewing angle

In this work, we restricted our results to the seven most important, according to the machine learning techniques, physical parameters when classifying Sy1 and Sy2. Following the AGN unification model, the main difference between these two types of galaxies resides in the viewing angle. Therefore, by comparing these seven parameters, we could understand the role that the viewing angle plays in determining the AGN type.

In Fig. 6.6, we show how different the AGN setups are when assuming only two viewing angles instead of a full range. With the exception of the viewing angle, all the parameters show a similar behaviour between the different setups. Only small differences in the E(B-V) and observed AGN disc luminosity are observed when comparing SKIRTOR and Fritz setups. Interestingly, in setups with ten viewing angles the estimates give a frequent value of ∼25°. Therefore, it seems that the full range of viewing angles in the setup does not significantly affect the other estimated SED parameters.

By construction, the viewing angle in X-CIGALE can determine the AGN type when the angle is close to the face-on (0-30°) and edge-on (70-90°) scenarios (Yang et al. 2020). Using the Chandra COSMOS Legacy survey (Marchesi et al. 2016) and X-CIGALE, Yang et al. (2020) estimated an accuracy of ∼71% in spectroscopic Type-1 and Type-2 AGNs. In our case, the accuracy of the classifications is around 82-85% when using only the viewing angle (Table 6.4), although with a different sample size (590 AGNs in Yang et al. (2020) while ∼ 8 000 in this work). The highest value in the accuracy (and other metrics) is obtained when using a setup with only two angles. If we use a setup with the full range of viewing angles, the distributions for the viewing angle (Fig. 6.7) are similar to what Gkini et al. (2021) found, where Sy1 are located at values around 20-30°while Sy2 galaxies are more scattered in a wider range of viewing angles. However, the estimations in terms of redshift (Sect. 6.3.3 and Fig. 6.8) seems to favour values around 20-30°and 60-70°when looking at Sy1 and Sy2, respectively. Thus, AGNs classifications can be improved by forcing the viewing angle to two values

that follow the Type-1 and Type-2 classifications. This result justifies the selection of two or even three viewing angles in similar studies using CIGALE or X-CIGALE in AGN galaxies (e.g. Vika et al. 2017; Zou et al. 2019; Pouliasis et al. 2020; Wang et al. 2020; Mountrichas et al. 2021).

This dichotomy in the viewing angle of AGN SED models could indicate that using the (IR) SED is not an adequate tool to estimate the viewing angle of AGNs as a continuous distribution, compared to spectroscopic measurements of the NLR (Fischer et al. 2013; Marin 2016). In Fig. 6.5, we show that the viewing angle is not the most important physical parameter when classifying Sy1 and Sy2. The observed AGN disc luminosity has the highest importance score to classify Seyferts, although its UV-optical emission is (in theory) angle-dependent (Netzer 2015; Yang et al. 2020). The power of the observed AGN disc luminosity in separating Sy1 and Sy2 galaxies is stunning when comparing with the viewing angle and other parameters in this work (Figs. 6.7-6.8 and Table 6.4). Using supernovae host galaxies catalogues, Villarroel et al. (2017) shows that the AGN luminosity, together with the stellar age, could play an important role in the AGN unification model beyond the viewing angle estimated from the torus when counting the number of supernovae in Type-2 AGNs. This result is similar to what we find in this work, although the age in the SEDs is not well constrained (Sect. 6.2.2).

In Fig. 6.10, we verify the estimations of the AGN accretion power for the different Seyfert types and AGN models. We notice that Fritz setups estimate higher accretion powers than SKIRTOR setups. This result is due to a different anisotropy correction applied in X-CIGALE, where the AGN accretion power does not depend on the viewing angle for the Fritz models. As a result of this correction, the difference between Seyfert types (middle panel) have different accretion powers depending on the AGN model. Therefore, L_{AGN}^{disc} cannot be easily transformed into bolometric luminosities due to its complex dependencies on the models. However, when we use only one of the AGN models in the intermediate Seyfert types, we notice similar distributions as with the L_{AGN}^{disc}. Therefore, the more detailed spectroscopic classifications follow a path from Type-2 to Type-1 AGNs by increasing the observed and intrinsic AGN disc luminosity (Sect. 6.3.2 and Figs. 6.9 and 6.10). This means that the AGN disc luminosity is an important factor, if not the main, in deciphering AGN types with SED models, and that a probable evolutionary path from intermediate types can be obtained with the AGN disc luminosity (Elitzur et al. 2014).

This result could have an impact in cosmological estimations done in AGNs. For example, QSOs are used to measure the luminosity distance at high redshifts to determine cosmological parameters (e.g. Lusso et al. 2019). If there is a dependency in the luminosity distance of QSOs with the viewing angle, it would lead to incorrect cosmological estimates (Prince et al. 2021). Even if the dependency is not entirely on the viewing angle but on the brightness of the AGN, then these estimates should be reformulated.

6.4.2 The AGN dust winds

As we mentioned before, X-CIGALE follows the simple AGN-unification scheme, i.e. the viewing angle determines the AGN type (Yang et al. 2020). If this simplified

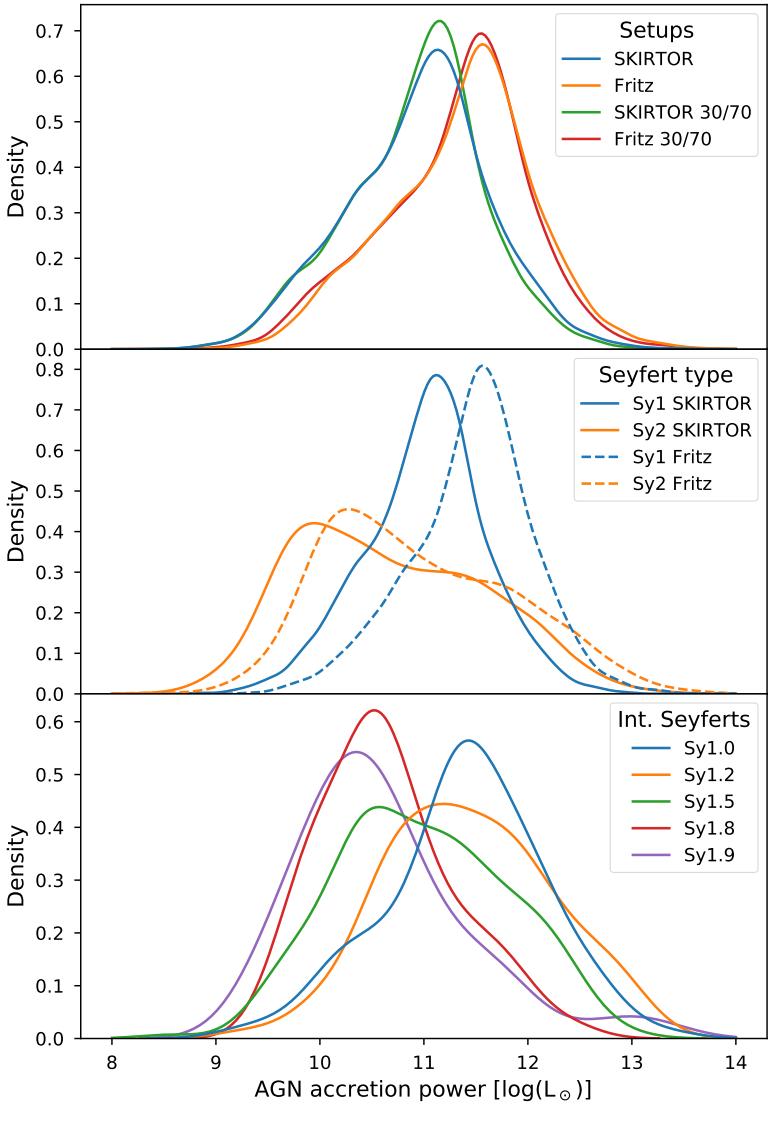

Figure 6.10 – *Probability density functions for the AGN accretion power of the four AGN setups (upper panel), Seyfert Type-1 and Type-2 for SKIRTOR and Fritz setups (middel panel), and intermediate Seyfert types for the SKIRTOR setup (lower panel). The AGN accretion power (intrinsic disc luminosity) depends on the selected AGN model, but the path from Type-2 to Type-1 AGNs by increasing the AGN accretion power.*

scheme is correct, the viewing angle should be the best classifier in our sample, but it is not. Our results for the smooth and clumpy torus models (Fritz and SKIRTOR, respectively) show that the observed AGN disc luminosity is a better classifier than the viewing angle. Perhaps the AGN "zoo" of galaxies is too complex to be explained with only the angle-dependent obscuration coming from a toroidal structure (Padovani et al. 2017).

A possible solution is the AGN disc-wind scenario (Emmering et al. 1992; Elitzur & Shlosman 2006; Netzer 2015). For example, Hönig & Kishimoto (2017) propose a model where dusty winds can explain the observational results on spatially resolved AGNs (e.g. García-Burillo et al. 2021; Alonso-Herrero et al. 2021). Later on, this model was extended by Hönig (2019), following interferometry IR and sub-mm observations (e.g. López-Gonzaga et al. 2016) where the AGN structure is composed of disc, wind and wind launching regions instead of a simple toroidal obscuration structure. In this scenario, the multiphase structure is consistent with the relation between the AGN obscured fraction (covering fraction) and AGN luminosity (Eddington ratio) in X-rays (Merloni et al. 2014; Ricci et al. 2017). The winds provide additional obscuration traced by covering factors and could separate the obscured and unobscured regions, and thus separate Type-1 and Type-2 AGNs. Similarly, Ogawa et al. (2021) describe a disc wind model analogous to Hönig (2019) using X-CLUMPY (Nenkova et al. 2008; Tanimoto et al. 2019), another SED toroidal model. In that work, X-CLUMPY is used in two different models for obscured and unobscured AGNs. With this approach, Ogawa et al. (2021) also finds a negative correlation between the Eddington ratio and the torus covering factor. Therefore, it is possible that the estimates of the AGN disc luminosity can be affected by other structures, like the ones proposed by Hönig (2019) or Ogawa et al. (2021). These structures may explain why the estimated observed AGN disc luminosity works better as a classifier in Sy1 and Sy2 with X-CIGALE.

An advantage of dusty wind structures is that they can also explain the origin of the red QSOs. These red QSOs are expected to be red due to dust in the line of sight (Webster et al. 1995), although it has recently been proposed that it is related to an evolutionary phase of QSO (Klindt et al. 2019). Calistro Rivera et al. (2021) studied the nature of red QSO using AGNfitter (Calistro Rivera et al. 2016), which resides in SED templates of smooth toroidal models (Silva et al. 2004), and found no connection between the AGN torus and the QSO reddening. However, they also found high-velocity winds in these red QSOs, suggesting that dusty wind structures may explain their nature.

Interestingly, even the intermediate Seyfert types can be explained with the disc-wind scenario. Elitzur et al. (2014) show that an evolutionary path of intermediate Seyfert types is related to the wind streamlines. When the accretion rate decreases the AGN luminosity decreases because the clouds in the wind streamlines move from high to low altitudes, generating an evolution sequence from 1.0 to 1.9 AGN types. This path is similar to the one presented in Sect. 6.3.2 and Figs. 6.9 and 6.10, where AGN disc luminosity decreases with Seyfert type. The estimations presented in this work support this idea, as the mean AGN accretion power is lower in Sy2 galaxies by ~ 0.5 dex than in Sy1 galaxies. However, we need to keep in mind that: i) the estimates of the physical parameters coming from AGN SED models depend on the chosen model

(as shown in Fig. 6.10 and by González-Martín et al. 2019a,b), and ii) the classification criteria for the intermediate types are different in Elitzur et al. (2014) and this work. Therefore, we need more information in terms of intermediate type classifications and in-depth observations of AGN regions with multi-wavelength observations to create a robust connection between theoretical models and observations.

6.4.3 AGN fraction and SFR in AGN types

To describe the difference between Type-1 and Type-2 AGN other physical parameters could be used besides AGN luminosity and viewing angle. In Sect. 6.3.2 and Fig. 6.7, we mentioned that parameters like $f_{\rm AGN}$ and SFR can also differentiate Seyfert types. However, we argue that the differences between AGN types will be smaller in these physical parameters.

In terms of the AGN contribution to the total IR luminosity ($f_{\rm AGN}$), Gruppioni et al. (2016) found that Sy2 galaxies tend to have a lower $f_{\rm AGN}$ compared with Sy1, in a sample of local galaxies. Similar results are found by Ramos Padilla et al. (2020), where six Type-1 AGN galaxies have higher $f_{\rm AGN}$ than AGN types with an estimated viewing angle of 60°, generally Type-2 AGNs. They also found that the $f_{\rm AGN}$ could increase with IR luminosity, as Alonso-Herrero et al. (2012) suggested. Therefore, Sy1 could have a higher AGN fraction than Sy2 because of their higher IR luminosity (e.g. Suh et al. 2019, and references therein). We find a similar behaviour in Fig. 6.7, although the values are widely distributed between both types. In addition, in Fig. 6.8, we find a slight increase of $f_{\rm AGN}$ with redshift as found by other works (e.g Wang et al. 2020), especially in Sy2 galaxies, but the statistic is small. The discrepancy in the results may also lie in the selection effects when comparing AGN types of galaxies, due to sample selection (e.g Calistro Rivera et al. 2016), AGN definitions (e.g Wang et al. 2020), or incorrect estimates of $f_{\rm AGN}$ because of cold-dust emission (McKinney et al. 2021). Therefore, it is not yet clear how $f_{\rm AGN}$, or its relationship to total IR luminosity, will change at higher redshifts ($z > 1$) between AGN types.

For the SFR, the relative importance is small when separating the two Seyfert types, following the feature selection (Sect. 6.3.1). This result agrees with Suh et al. (2019) and Masoura et al. (2021), who found no differences between AGN types in terms of SFR and stellar mass associated with AGN power in X-rays, although large uncertainties are present for Type-1 AGNs. On the other hand, Zou et al. (2019) show differences in stellar mass but not in SFR between AGN types, even using different SFR calculation methods (SED, H_α and IR luminosities). In this work, stellar mass is estimated with X-CIGALE but its relative importance is even lower than the importance of the SFR. In addition, we do not find a clear separation between the two Seyfert types in terms of SFR with redshift (Fig. 6.8). The SFR will increase for both types as the SFR increases with the redshift until the cosmic noon, therefore we are only seeing a selection effect.

Thus, looking at $f_{\rm AGN}$ and SFR individually may give some clues about the difference between AGN types. However, the differences are small when compared together with other physical parameters, such as the observed AGN disc luminosity. Furthermore, one of the main problems of looking at different physical parameters resides in the definition between obscured (Type-2) and unobscured (Type-1) AGNs (Hickox &

Alexander 2018), as we will see in the following subsection.

6.4.4 AGN Classifications

In this work, we treated the Seyfert types coming from SMB and VCV as true classifications. However, that does not mean that the Seyfert galaxy sample studied here is not populated with other types of AGN galaxies. For example, galaxies that we assume as unclassified Seyfert end up being mostly QSO in the literature (Sect. 6.3.5). In any case, we have presented evidence from different AGN setups that seem to agree with the general classification of Type-1 and Type-2 AGN. Then, we could use simple relations, like the one presented in Eq. 6.3, to separate AGN galaxies using the features estimated from SED modelling of the AGN component available in X-CIGALE. Machine learning ensemble techniques could also be used to classify galaxies with X-CIGALE outputs, achieving even higher accuracy than just one physical parameter (Table 6.4). Nevertheless, these methods will require correct classifications of galaxies that sometimes are not available.

As we discussed in Sect. 6.4.2, the intermediate Seyfert classifications could help solve the problems of the AGN unification model. If these galaxies are an evolutionary stage of AGNs, as proposed by Elitzur et al. (2014) with disc wind models, then classifying AGN galaxies could be crucial to join the theoretical models with observations. In this work, all the intermediate classifications come from VCV with the quantitative approach of Winkler (1992), as described in Sect. 6.3.2. However, other criteria could be applied in the classification of intermediate AGN galaxies (e.g. Stern & Laor 2012). More spectral information and consistent classifications will be needed in intermediate AGN galaxies to clarify this evolutionary stage, for example using the on-going DEVILS survey. Future observations with the *James Webb Space Telescope* (JWST) will detect obscured AGNs and calculate AGN fractional contributions through photometry and spectral line features (e.g. Satyapal et al. 2021; Yang et al. 2021a), which will elucidate the path for classifying these type of galaxies.

6.4.5 X-ray information

The efforts to create catalogues of AGN galaxies with X-ray data are crucial to classify galaxies missed in optical catalogues (e.g. Koss et al. 2017). These missed galaxies are in general obscured AGN galaxies (Type-2), which are difficult to classify, even using ensemble methods in optical wavelengths (Golob et al. 2021). However, it is good to keep in mind that Type-2 AGNs could remain undetected in IR and X-ray colour surveys (Pouliasis et al. 2020). Therefore, a multi-wavelength study, like the one presented in this work, is ideal to tackle the completeness and classification problem in AGN galaxies.

Unfortunately, a limitation of this study is the lack of X-ray data when fitting the SED, although there is a correlation between the AGN MIR luminosity and the 2–10 keV X-ray luminosity in Type-1 and Type-2 AGNs (Gandhi et al. 2009; Suh et al. 2019). We conclude that the quantity of photometric data in NED and CDS is not enough nor homogeneous to include the X-ray module in the X-CIGALE setups. None the less, X-rays can help to constrain non-physical parameters in SED tools as X-CIGALE (Yang et al. 2020). Recent studies with different SEDs models that include X-ray

data show how important the inclusion of these data could be to constrain the AGN models (e.g. Suh et al. 2019; Ogawa et al. 2021; Mountrichas et al. 2021; Masoura et al. 2021). In the future, we expect to understand the probable evolutionary paths discussed in this work for different AGN types, with the AGN luminosity, $f_{\rm AGN}$, and X-ray data. In the next decade, the combination of the *Athena space telescope* with X-CIGALE will help to unambiguously determine the presence of AGN for this type of works (Yang et al. 2020).

6.5 Conclusions

We have used a sample of 13 173 Seyfert galaxies from SMB and VCV to assess the importance of the viewing angle in AGN SED models. We have used a data-driven approach by retrieving photometric data from astronomical databases (CDS and NED) to be used in the SED analysis with X-CIGALE. Two AGN SED toroidal models (Fritz and SKIRTOR setups) were used with different viewing angle configurations to verify the effect of viewing angle selection in the estimated physical parameters. These estimates have been validated by comparing our results with those from Vika et al. (2017), showing good agreement except for the AGN fractions, which can be related to the different assumptions of the grids.

Our main conclusions are the following:

1. The estimated viewing angle from X-CIGALE seems to be the second best discriminator when assessing AGN type. This result is supported by different prediction metrics and importance scores in machine learning algorithms which favour the observed AGN disc luminosity as the most important physical parameter.

2. The initial viewing angle assumption in X-CIGALE does not significantly affect the other estimated physical parameters if at least two viewing angles that follow the AGN Type-1 and Type-2 classifications are taken into account.

3. At different redshifts ($z \lesssim 0.5$), the smooth and clumpy torus AGN models seem to favour viewing angles around 20-30°and 60-70°when looking at Sy1 (Type-1 AGN) and Sy2 (Type-2 AGN), respectively. While in terms of the observed AGN disc luminosity, we propose to use a limit (Eq. 6.3) that separates both types. These values may predict the AGN type in unclassified AGN galaxies, as shown in the case of unclassified and discrepant Seyfert galaxies.

4. Machine learning ensemble methods can be used for AGN classification tasks but require the use of several parameters from X-CIGALE. These parameters include individual physical parameters that are important for the classification (e.g. viewing angle or observed AGN disc luminosity). Nevertheless, these methods require correct classifications (training data) that often vary based on criteria.

5. The observed and intrinsic AGN disc luminosity decreases from Type-1 to Type-2 AGNs in the intermediate Seyfert types. This decrease may be explained by accretion rates within AGN disc wind models, which show an evolutionary path

among these AGN types. However, more information is needed to create a robust connection between theoretical models and observations.

In this work, we have demonstrated usefulness of the broad-band SED tool as X-CIGALE to classify AGN galaxies in Type-1 and Type-2. Thus, X-CIGALE could be a powerful tool to characterise AGNs in the upcoming years. Future space telescopes like the JWST and *Athena* will get crucial photometry and spectroscopy to constrain AGN physical parameters like luminosity and $f_{\rm AGN}$, which will improve AGN galaxy classifications. These classifications along with other physical parameters will help us understand the real scenario that describes AGN galaxies.

Data Availability

Most of the data and code underlying this article are available in Zenodo, at https://doi.org/10.5281/zenodo.5227294. The X-CIGALE estimated data from the AGN setups are available in a companion repository in Zenodo, at https://doi.org/10.5281/zenodo.5221764, which also includes a script to create similar figures like the one presented in Fig. 6.2 for all galaxies in this work.

Acknowledgements: K. Małek has been supported by the Polish National Science Centre grant (UMO-2018/30/E/ST9/00082). We would like to thank the Center for Information Technology of the University of Groningen for their support and for providing access to the Peregrine high performance computing cluster. We acknowledge the anonymous referee for a careful reading of the manuscript and very helpful suggestions and comments.

This research made use of Astropy,* a community-developed core Python package for Astronomy (Astropy Collaboration et al. 2013, 2018). This research has made use of the SIMBAD database, operated at CDS, Strasbourg, France. This research has made use of the NASA/IPAC Extragalactic Database (NED), which is funded by the National Aeronautics and Space Administration and operated by the California Institute of Technology. This research has made use of NASA's Astrophysics Data System Bibliographic Services. Part of the initial exploratory analysis was performed with TOPCAT (Taylor 2005).

* http://www.astropy.org

Appendix

Narrow-line Sy1 galaxies

The narrow-line Sy1 (NLSy1) are AGN galaxies that share similarities with Sy1 and Sy2, but cannot be identified as an intermediate Seyfert type (Osterbrock & Pogge 1985). These galaxies are classified as NLSy1 due to their i) Hβ line emission profile (broad and narrow components), ii) ratio between Hβ and [O III] fluxes below 3 ($R < 3$), iii) narrow FWHM, and iv) the presence of [Fe II] multiplets (e.g. Zhou et al. 2006; Rakshit et al. 2017; Chen et al. 2018). These characteristics indicate that NLSy1 are similar to Sy1. None the less, NLSy1 seem to show higher Eddington rates, lower black hole masses and different viewing angles (Rakshit et al. 2017).

We assumed that NLSy1 are Sy1 galaxies in the catalogue of VCV, as most of these NLSy1 were classified as Sy1 in SMB. However, the results from the AGN setups in this work show small differences in the AGN parameters between these two AGN types. In Figure 6.11, we present the density functions of the AGN physical parameters for NLSy1 and Sy1 as classified by VCV. For almost all physical parameters the median values for these AGN types are different, although it does not show any difference for the viewing angle. The null hypothesis of the KS test is always rejected when comparing NLSy1 and Sy1 galaxies. All the parameters have a higher D value than the critical value, $D_{\rm crit} = 0.04$ in the SKIRTOR setup. This means that both samples originate from different distributions. Nevertheless, in terms of the viewing angle, we notice the smallest difference between these two types ($D = 0.08$). This small difference supports our assumption that Sy1 and NLSy1 can be treated as the same type in this work, as our focus resides in the AGN viewing angle.

Conversely, there are larger differences in the total AGN luminosity, as well in terms of their components (disc and re-emitted dust) and intrinsic accretion power ($D > 0.24$). These differences may be related to the higher accretion of the AGN in this type of galaxy. Nevertheless, the difference in the AGN physical parameters for Sy1 and NLSy1 may require the X-ray bands, which were not obtained in this work (Sect. 6.4.5), as they seem to show a steeper X-ray slope in NLSy1 (Scott & Stewart 2014). Then, a more specific study of these two types of galaxies will be required to understand their nature.

Mock results

We perform a mock analysis on the main physical parameters to verify the quality of the fits inside X-CIGALE. We use the estimations from the SKIRTOR and Fritz setups. The mock analysis is performed inside X-CIGALE by creating mock values from the original photometry flux and errors and the best fit of the object. Then, the same method used in the original estimation is applied to obtain mock estimations. With this analysis, we can estimate the reliability of the obtained estimations for the physical parameters (Boquien et al. 2019; Yang et al. 2020).

Figure 6.12 shows the mock analysis for the SKIRTOR setup in the seven selected parameters used for this study (Sect. 6.3.1) with the addition of the accretion power (intrinsic disc luminosity) and stellar mass. In general, all the physical parameters

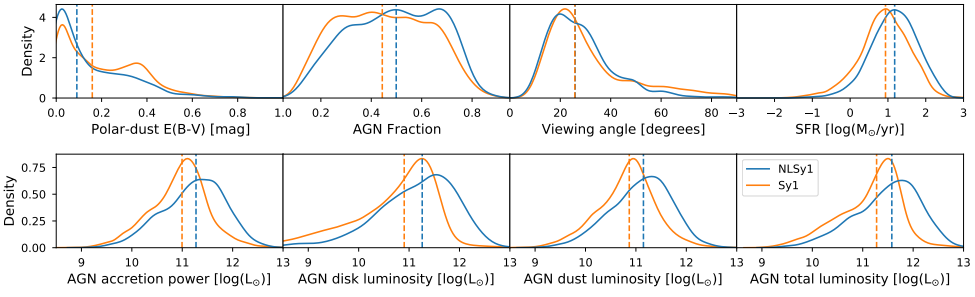

Figure 6.11 – *Probability density functions for the AGN estimated parameters and SFR comparing NLSy1 (blue) and Sy1 (orange) galaxies for the SKIRTOR setup with the respective median values (vertical dashed lines). No difference is observed in the viewing angle median estimates. However, NLSy1 show a higher luminosity in the intrinsic component (accretion power), the emitted disc, dust luminosities and the total AGN luminosity.*

are well correlated, although some parameters tend to have larger uncertainties such as polar dust and viewing angle. The e-folding time of the main stellar population ($\tau_{\rm main}$) is the most affected parameter when comparing with mock values. This result is expected as the age estimates inside CIGALE are not well constrained, as noted by Vika et al. (2017) and shown in Sect. 6.2.2. The mock analysis for the Fritz setup (found in the online repository) shows similar results as the Fritz setup. Therefore, the quality of the fits allows us to analyse the estimated physical parameters in this work.

Classifications of individual objects in the literature

For the 14 unclassified and 45 discrepant Seyfert classifications discussed in Sect. 6.3.5, we searched the literature for other classifications of activity type. In Table 6.7, we present the activity type information and the reference for these classifications. We did not find another classification for some of these galaxies, therefore these galaxies are not listed in Table 6.7.

Table 6.7 – *Activity types found in the literature for unclassified and discrepant Seyfert galaxies.*

Object ID	Activity classification types	References
	No Type in VCV & SMB	
2MASX J12140343-1921428	Mixed blazar; QSO; Sy1	2,19; 16; 18
2MASX J23032790+1443491	Composite; Sy1	20; 18
CADIS 16-505716	QSO	18
LEDA 1485346	Sy1	18
LEDA 3095610	QSO; Sy1	16; 18
LEDA 3096762	QSO	18
MCG+00-11-002	Sy1	18
MCG+03-45-003	Sy1; Sy2	18; 23
QSO B1238+6232	QSO	18
[HB93] 0248+011A	QSO	18
	Type in VCV or SMB	
2E 2294	QSO; Sy1	6,9,12,15,16,21; 18
2E 2628	QSO	6,9,13,15,16,18,21
2E 3786	QSO	6,9,13,15,16,18,21
2MASS J00423990+3017514	Sy1	18,23
2MASS J01341936+0146479	QSO; Sy1	6,16,18
2MASS J02500703+0025251	QSO; SF; Sy1	6,10,14,16,21,22; 8; 18,20
2MASS J08171856+5201477	Sy1; NLSy1	3,18,20; 7
2MASS J09393182+5449092	QSO; Sy1	16,9,4,18,21; 20
2MASS J09455439+4238399	QSO; Sy1; NLSy1	14; 20; 4,7,17,18
2MASS J09470326+4640425	QSO; NLSy1	6,9,16,18,21; 7
2MASS J09594856+5942505	QSO; Sy1	6,9,10,12,14,16,21; 18
2MASS J10102753+4132389	QSO	6,9,10,13,15,18,21
2MASS J10470514+5444060	QSO; Sy1	4,6,9,15,16,21; 18,20
2MASS J12002696+3317286	QSO; Sy1	16,9,6,14,21; 18
2MASS J15142051+4244453	QSO; Sy1	4,6,12,14,15,16,21; 18,20
2MASSI J0930176+470720	QSO; Sy1	6,9,23,16,18,21; 20
2MASX J02522087+0043307	QSO; Sy1; SF	4,14,16,22; 18,20; 8
2MASX J02593816+0042167	QSO; Sy1; SF	6,12,14,16,21,22; 18,20; 8
2MASX J06374318-7538458	Sy1	18
2MASX J09420770+0228053	Sy2; LINER	18; 20
2MASX J09443702-2633554	QSO; Sy1; Sy1.5	16; 18; 1
2MASX J09483841+4030436	QSO; Sy1	14; 18,20
2MASX J10155660-2002268	Sy1	18
2MASX J10194946+3322041	Sy1; NLSy1	18; 7
2MASX J15085291+6814074	Sy1	18
2MASX J16383091-2055246	Sy1; NLSy1	18; 1,5,11
2MASX J21033788-0455396	Sy1	18
2MASX J22024516-1304538	Sy1	18
2dFGRS TGN357Z241	QSO; Sy1	4,16; 18
3C 286	Mixed Blazar; QSO	2,19; 6,12,13,18,21
6dFGS gJ034205.4-370322	Mixed Blazar; QSO; Sy1	2,19; 16; 18
6dFGS gJ043944.9-454043	QSO	16,18
6dFGS gJ084628.7-121409	Sy1; NLSy1	18; 5
CTS 11	Sy1; NLSy1	18; 11
HE 0226-4110	QSO; NLSy1	16,18; 5
ICRF J025937.6+423549	Mixed Blazar; QSO	19; 18
ICRF J081100.6+571412	QSO	6,12,13,14,15,18,21
ICRF J100646.4-215920	Mixed Blazar; Sy1	2,19; 18
ICRF J110153.4+624150	BZQ; QSO; SF	2,18; 6,9,12,13,18,21; 8
ICRF J135704.4+191907	BZQ; QSO	2,18,23; 6,9,12,13,14,18,21
IRAS 10295-1831	Sy1	18
Mrk 1361	QSO; Sy1; Sy2	14; 20; 18
PB 162	QSO; Sy1; NLSy1; SF	14,16; 18,20; 7; 8
UGC 10683	Sy1	18

References: 1: Panessa et al. (2020); 2: D'Abrusco et al. (2019); 3: Liu et al. (2018); 4: Dong et al. (2018); 5: Chen et al. (2018); 6: Pâris et al. (2018); 7: Rakshit et al. (2017); 8: Duarte Puertas et al. (2017); 9: Gupta et al. (2016); 10: Ai et al. (2016); 11: Schmidt et al. (2016); 12: Albareti et al. (2015); 13: Gentile Fusillo et al. (2015); 14: Sun & Shen (2015); 15: Krawczyk et al. (2015); 16: Souchay et al. (2015); 17: Järvelä et al. (2015); 18: Flesch (2015), version 7.2 Flesch (2021); 19: D'Abrusco et al. (2014); 20: Toba et al. (2014); 21: Krawczyk et al. (2013); 22: Meusinger et al. (2011); 23: Cusumano et al. (2010), coordinates match within 2″, distance varies between 0.56″and 1.77″.

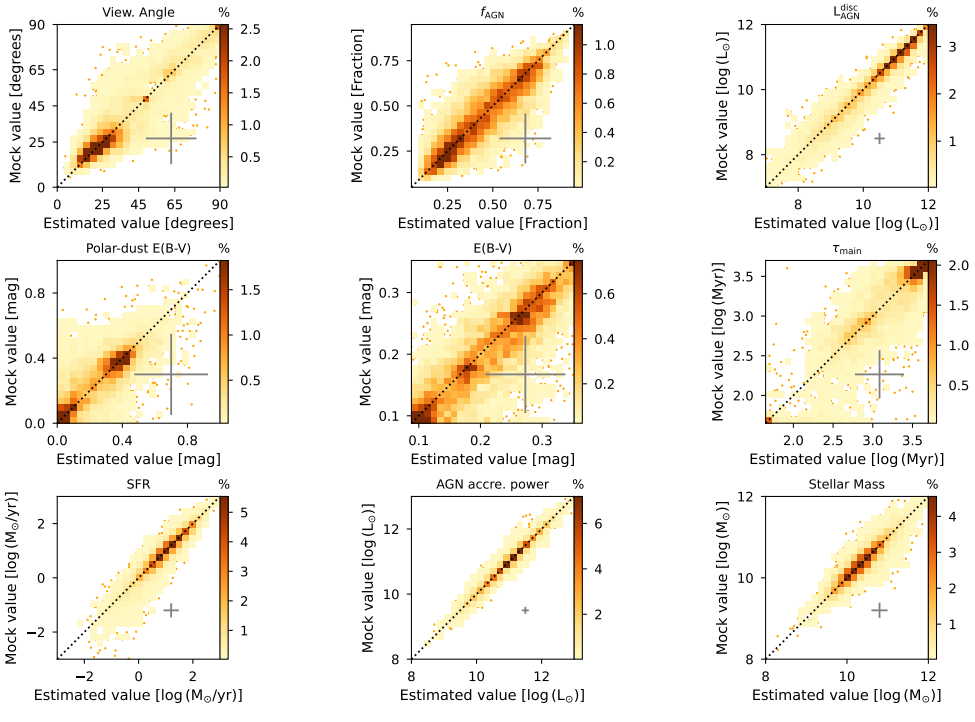

Figure 6.12 – *Mock versus estimated values from SKIRTOR setup for nine parameters studied in this work. The 2-dimensional histograms show the percentage of galaxies in the sample that fall in the parameter space sampled by 25 cells in each dimension. The pointed line represents the 1:1 relation and the grey crosses represent the median estimated error for each of the parameters. Cells with only one galaxy are not coloured, instead we draw these individual galaxies as orange dots.*

From a different perspective, an imposing wall is just a large door

Gosho Aoyama - Detective Conan

I must not make a fool of myself

Andrés looking at a door (Duwen of Trekken)

7

Conclusions

In this thesis, I have investigated the physical processes (such as star formation, AGN and mergers) and gas properties in galaxies. I have focused on diagnosing the interstellar medium (ISM) of galaxies using far-infrared (FIR) emission lines to decouple the ISM phases contributing to these lines. With the information obtained from the FIR emission lines, we have analysed their dependence on physical properties of the galaxies, such as star-formation rates (SFR) and interstellar radiation field (ISRF). We have used diagnostic diagrams to give clues about the ISM gas properties of galaxies. Furthermore, we have developed a web app to retrieve the physical parameters of galaxies. Scientists can use this tool to predict physical parameters from their observed FIR emission line luminosities.

In addition, I have examined the relation between star-formation rates and fractional AGN contributions in galaxies using SED models. We focus on interacting galaxies and typical AGNs to understand the main-sequence of star-forming galaxies (MS). We compare different stages of interaction in the MS and other physical properties of the galaxies. Finally, we turn our attention to the viewing angle parameter for classifying AGN galaxies. For this final chapter, we use publicly available data, so almost all our analysis is reproducible.

In this chapter, I briefly summarise the main results of the previous scientific chapters and provide a future perspective on the topics in this thesis.

7.1 ISM conditions at $z \lesssim 6$

The results from Chapters 2 and 3 show that by post-processing EAGLE simulations and implementing an ISM model with four phases (dense molecular gas, neutral atomic gas, diffuse ionised gas (DIG) and HII regions) it is possible to trace the properties of the ISM gas as a function of cosmic time. We are able to reproduce the observed SFR–FIR line luminosity relationship over the range $z = 0$–6 for all FIR lines. This leads us to examine the fractional contribution of the different ISM phases and how it depends on SFR and metallicity.

Briefly, the [C II] emission line (the most robust indicator of gas cooling in galaxies) changes from depending on the different ISM phases in the local Universe to depending mainly on neutral atomic gas at $z = 6$. Similarly, the [O I] lines change from being dominated by dense molecular gas and neutral atomic gas to being dominated by only neutral atomic gas at $z = 6$. Other lines like [O III] and [N III] go from being dominated by DIG and HII regions in the local Universe to being dominated solely by HII regions at $z = 6$. Finally, the [N II] lines maintain a balance between DIG and HII regions throughout cosmic time. These ISM changes have an impact on the theoretical models and typical assumptions made for these lines, where conversion factors to determine the cosmic HI mass densities (e.g. Heintz et al. 2021) can be used to better understand the origin of galaxies.

7.1.1 The deficit in FIR emission lines

One of the main problems with using FIR line luminosities to trace SFR is that when estimating their ratio to the FIR luminosity they seem to increase at a lower rate with FIR luminosity at the bright end, an effect known as the 'FIR line deficit' (e.g. Díaz-Santos et al. 2017). In Chapter 2 we show that this deficit may be related to a decrease in the size of neutral clouds due to an increase in SFR. We check whether this may also be related to an increase of intense radiation fields, such as those present in starburst galaxies. We have found that galaxies that tend to have higher specific star-formation rates (sSFR) tend to have lower luminosity–SFR ratios. In other words, galaxies with an increased value of ΔMS (the distance to the main sequence), show a larger deficit in the lines. Which again links the star-formation regulation (in starburst galaxies) as being responsible for the observed variations in some FIR line luminosities at high infrared luminosities.

These results are confirmed at redshifts up to $z = 6$ in Chapter 3, where the luminosity–SFR ratio of FIR lines almost always decreased with increasing ΔMS. Therefore, the 'line deficit' may be real for starburst galaxies, but recent studies show no clear evidence for it at high-z (e.g. Schaerer et al. 2020). More information from galaxies with "normal" brightness is required to confirm these results in other redshift ranges.

7.1.2 Diagnostic tools with FIR lines

To better understand the behaviour of the FIR lines and the contributing ISM phases, we present diagnostic diagrams in Chapter 3 comparing some physical parameters that we can obtain from the simulations with observations. We find that FIR line ratios like [C II]/[O III] and [N II]/[O I] are useful to trace parameters such as ISRF, metallicity

and sSFR in diagnostic diagrams. These diagnostic diagrams are limited in that the total number of observations we can collect for the FIR lines is not statistically significant (except for the local Universe). However, it gives some clues on the range of parameters we expect for certain line luminosity ratios.

In Chapter 4, we present a web app that uses the information of the FIR line luminosities to derive the physical parameters following the predictions of the simulations. In this way, it will be possible to diagnose the physical parameters in galaxies with different redshifts once they become available from future observations. This tool, called DIAGISM, has proven useful in deriving similar SFR in an observational sample of galaxies using only the [C II] and [O III] 88 μm emission lines. We expect that other estimates from DIAGISM can be used to predict other physical parameters such as ISRF, metallicity, external pressure (P_{ext}), total hydrogen number density in the neutral clouds ($n(H)_{cloud}$), radius of the neutral clouds (R_{cloud}), and gas mass (M_{gas}).

7.2 Star-forming and AGN galaxies

In Chapter 5, we study different samples of galaxies to understand the role that SF and AGN play in the different stages of galaxy interactions (mergers). It has been proposed that interacting galaxies tend to have high SFR close to coalescence and then the AGN emission peaks right after the peak of SFR (see Fig. 1.6). We estimate both the SFR and AGN fraction (f_{AGN}) in AGN, "classical" star-forming galaxies and interacting galaxies with the help of a SED tool (CIGALE). From those results we have found a tentative correlation between f_{AGN} and the stages of interaction. However, some of the interacting galaxies that show a low SFR and high f_{AGN} were late-stage mergers. In addition, by including an AGN component into the SED fitting we notice that estimations of SFR may change drastically when the AGN is becoming dominant ($f_{AGN}>0.2$). We confirm this finding when comparing Seyfert galaxies (a type of AGN galaxy) at different levels of f_{AGN} (Chapter 6). This result may have important implications when comparing the main-sequence of star-forming galaxies with different types of galaxies.

7.2.1 Main sequence of star-forming galaxies

With the SFR corrected by taking into account the AGN contribution, we have found that galaxies with relatively high AGN fractions (AGNs and interacting galaxies with $f_{AGN}>0.4$) still reside in the MS or are very close to it. Therefore, the AGN activity may be responsible for a significant displacement of galaxies across the MS. Results from Leslie et al. (2020) have shown that adding AGN galaxies in the MS tends to bend the relation at higher stellar masses for different redshift ranges. In order to interpret these results with spectroscopic information, we propose to use the f_{AGN}-[Ne V]/[Ne II] plane to diagnose luminous infrared galaxies in terms of their stage of interaction. However, more information is needed to clearly understand if the proposed process of evolution from star-forming to AGN galaxies comes mainly from the interaction of galaxies. For example, there is still a debate if the AGN can quench SF, even though SED tools seem to suggest there is an anti-correlation between f_{AGN} and SFR (e.g. Gao et al. 2021, and references therein).

7.2.2 AGN classifiers

To understand the nature of AGN galaxies, we delved into the classification of AGN galaxies. These AGN classifications come from the assumption that an obscuring torus explains the variety of AGN types due to their orientation with respect to the line-of-sight (e.g. Antonucci 1993; Urry & Padovani 1995). We look at the viewing angle parameter in AGN SED models in Chapter 6, where we find that the observed AGN disc luminosity is better judging the two typical AGN types (Type-1 and Type-2) than the viewing angle.

Interestingly, both the observed and intrinsic AGN disc luminosity parameters show that accretion rates within the AGN may play a role in the evolution of AGN types. Specifically, by looking at intermediate stages of the two typical AGN types (e.g. Seyfert 1.2, Seyfert 1.5, etc.) we see a trend for them to change from low accretion rates in Seyfert 1.9 to high accretion rates in Seyfert 1.2. This shows that SED models are a powerful tool to characterise AGNs and future observations could clarify whether the evolution between different AGN types comes from the line of sight or whether intrinsic properties are more important.

7.3 Outlook

7.3.1 Confirmation from observational data

ALMA

Currently, ALMA is the only facility that can provide clear insights on FIR emission lines outside the local Universe. Some ALMA receivers are ideal for observing galaxies at $z > 3$ in some FIR lines (e.g. Mordini et al. 2022, see. their Fig.1). While this information is still being collected, we expect tools like DiagISM, shown in Chapter 4, can be used to interpret the observations of the FIR lines. Wider field observations of lines such as [C II], [O III] and [O I] will test our findings in terms of the evolution of the ISM.

JWST

With upcoming JWST observations, new information will be available in the MIR wavelength range of the spectrum. This information will improve our knowledge in terms of the contribution of the AGN (e.g. $f_{\rm AGN}$) in galaxies, both using spectroscopy and broad-band photometry (e.g. Yang et al. 2021a). We expect that with more information about the [Ne II], [Ne III] and [Ne V] emission lines we will better understand the AGN accretion processes. Furthermore, observing interacting galaxies with better resolution could also help improve the treatment of radiation processes present in current SED models.

Space telescopes: SPICA-like, LETO-like or PRIMA

Although there is currently no IR space telescope in development, existing mission concepts such as a SPICA-like, LETO-like or PRIMA will allow us to confirm the possible evolution of the ISM of galaxies through cosmic times. The need for a

dedicated telescope in the FIR range is crucial to understanding the gas processes that we cannot observe at other wavelengths. Not only can the information of the ISM phases be retrieved with the FIR lines, but also some physical parameters as we saw in Chapter 4. With one of these telescopes, we will hopefully obtain crucial FIR observations on statistically large samples of galaxies from the local Universe to the cosmic dawn that will help us discover how galaxies form and evolve.

7.3.2 A step forward in ISM and AGN modelling

As we mentioned before, our understanding of the physics involved in the formation and evolution of galaxies will improve with the arrival of new facilities such as JWST and new observations from ALMA, and hopefully a dedicated FIR space telescope. With that knowledge, it will be possible to extend and improve the model presented in Chapter 3. An improvement that we would like to add in the future is the addition of AGN in terms of X-ray dominated regions (e.g. Wolfire et al. 2022). With this, it will be possible to check the effects on the emission of FIR emission lines. At the same time, we would also like to verify how FIR lines change in terms of morphological and kinematic properties of galaxies.

In addition to that, we expect that cosmological hydrodynamical simulations will also improve significantly. For example, some of the most important parameters for describing the Universe with simulations are mass resolution, volume and the number of galaxies that can be retrieved. Improving these parameters is and will be crucial for the next generation of simulations to succeed (Nelson et al. 2019).

7.3.3 Open science as a paradigm shift

I cannot find a better way to finish this thesis than by talking about the future of science. Halfway through the process to obtain the scientific chapters presented before, I was delighted to learn about the new paradigm of open science and how it can help to improve several aspects of scientific discoveries. Part of the information presented in this thesis is publicly available, not only in terms of scientific outputs but also in terms of the code and information involved in the analysis of all the data. Two of our chapters (Chapters 4 and 6) are fully reproduced in a web browser, and in one of them, I provide a user-friendly environment to "play" with the data (Chapter 4). We follow the FAIR principles on data (Wilkinson et al. 2016) and we expect that it will become more common in the near future, not only in astronomy research but in all sciences.

The COVID-19 pandemic not only affected the way this PhD thesis was developed but also helped to show that scientists need to communicate with different communities of the society to generate trust in scientific discoveries (Miedema 2022). I believe open science can be the idea that can help us to achieve that connection with society. Publishing and sharing knowledge without economical barriers, using data as the rule instead of articles and prioritising the reproducibility of results will help us to change the scientific paradigm for the best of all humans on this small planet in the Universe.

7

Bibliography

Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, ApJS, 182, 543
Abdullah, A., Brandl, B. R., Groves, B., et al. 2017, ApJ, 842, 4
Abdullah, A. & Tielens, A. G. G. M. 2020, A&A, 639, A110
Abel, N. P. & Satyapal, S. 2008, ApJ, 678, 686
Accurso, G., Saintonge, A., Catinella, B., et al. 2017, MNRAS, 470, 4750
Ai, Y. L., Wu, X.-B., Yang, J., et al. 2016, AJ, 151, 24
Alatalo, K., Bitsakis, T., Lanz, L., et al. 2017, ApJ, 843, 9
Albareti, F. D., Comparat, J., Gutiérrez, C. M., et al. 2015, MNRAS, 452, 4153
Alonso-Herrero, A., García-Burillo, S., Hönig, S. F., et al. 2021, A&A, 652, A99
Alonso-Herrero, A., Pereira-Santaella, M., Rieke, G. H., & Rigopoulou, D. 2012, ApJ, 744, 2
Amblard, A., Riguccini, L., Temi, P., et al. 2014, ApJ, 783, 135
André, P., Di Francesco, J., Ward-Thompson, D., et al. 2014, in Protostars and Planets VI, ed. H. Beuther, R. S. Klessen, C. P. Dullemond, & T. Henning, 27
Antonucci, R. 1993, ARA&A, 31, 473
Antonucci, R. R. J. & Miller, J. S. 1985, ApJ, 297, 621
Arata, S., Yajima, H., Nagamine, K., Abe, M., & Khochfar, S. 2020, MNRAS, 498, 5541
Aravena, M., Decarli, R., Walter, F., et al. 2016, ApJ, 833, 71
Armus, L., Mazzarella, J. M., Evans, A. S., et al. 2009, PASP, 121, 559
Ashby, M. L. N., Mahajan, S., Smith, H. A., et al. 2011, PASP, 123, 1011
Ashby, M. L. N., Stern, D., Brodwin, M., et al. 2009, ApJ, 701, 428
Asmus, D., Hönig, S. F., Gandhi, P., Smette, A., & Duschl, W. J. 2014, MNRAS, 439, 1648
Asplund, M., Grevesse, N., Sauval, A. J., & Scott, P. 2009, ARA&A, 47, 481
Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, AJ, 156, 123
Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33
Bacchini, C., Fraternali, F., Iorio, G., & Pezzulli, G. 2019, A&A, 622, A64
Baes, M., Davies, J. I., Dejonghe, H., et al. 2003, MNRAS, 343, 1081
Bakx, T. J. L. C., Tamura, Y., Hashimoto, T., et al. 2020, MNRAS, 493, 4294
Baldwin, J. A., Phillips, M. M., & Terlevich, R. 1981, PASP, 93, 5
Baqui, P. O., Marra, V., Casarini, L., et al. 2021, A&A, 645, A87
Barber, C., Crain, R. A., & Schaye, J. 2018, MNRAS, 479, 5448
Bellstedt, S., Robotham, A. S. G., Driver, S. P., et al. 2021, MNRAS, 503, 3309
Benitez-Llambay, A. 2015, Py-Sphviewer: Py-Sphviewer V1.0.0, Zenodo, Zenodo, 10.5281/zenodo.21703
Bernard-Salas, J., Spoon, H. W. W., Charmandaris, V., et al. 2009, ApJS, 184, 230
Bernhard, E., Tadhunter, C., Mullaney, J. R., et al. 2021, MNRAS, 503, 2598
Bertin, E. & Arnouts, S. 1996, A&AS, 117, 393
Béthermin, M., De Breuck, C., Gullberg, B., et al. 2016, A&A, 586, L7
Béthermin, M., Fudamoto, Y., Ginolfi, M., et al. 2020, A&A, 643, A2
Bethermin, M., Gkogkou, A., Van Cuyck, M., et al. 2022, arXiv e-prints, arXiv:2204.12827
Béthermin, M., Wu, H.-Y., Lagache, G., et al. 2017, A&A, 607, A89
Bialy, S. & Sternberg, A. 2019, ApJ, 881, 160
Bishop, C. M. et al. 1995, Neural networks for pattern recognition (Oxford university press)
Bitsakis, T., Charmandaris, V., da Cunha, E., et al. 2011, A&A, 533, A142
Black, J. H. 1987, in Interstellar Processes, ed. D. J. Hollenbach & J. Thronson, Harley A., Vol. 134, 731
Blecha, L., Snyder, G. F., Satyapal, S., & Ellison, S. L. 2018, MNRAS, 478, 3056
Blitz, L., Fukui, Y., Kawamura, A., et al. 2007, in Protostars and Planets V, ed. B. Reipurth,

D. Jewitt, & K. Keil, 81
Bluck, A. F. L., Maiolino, R., Sánchez, S. F., et al. 2020, MNRAS, 492, 96
Boissay, R., Ricci, C., & Paltani, S. 2016, A&A, 588, A70
Bolatto, A. D., Leroy, A. K., Rosolowsky, E., Walter, F., & Blitz, L. 2008, ApJ, 686, 948
Bonato, M., Negrello, M., Cai, Z. Y., et al. 2015, MNRAS, 452, 356
Bonatto, C. & Bica, E. 2011, MNRAS, 415, 2827
Boquien, M., Burgarella, D., Roehlly, Y., et al. 2019, A&A, 622, A103
Bordalo, V., Plana, H., & Telles, E. 2009, ApJ, 696, 1668
Boselli, A., Voyer, E., Boissier, S., et al. 2014, A&A, 570, A69
Bouwens, R., González-López, J., Aravena, M., et al. 2020, ApJ, 902, 112
Bouwens, R. J., Smit, R., Schouws, S., et al. 2022, ApJ, 931, 160
Braatz, J. A., Wilson, A. S., Gezari, D. Y., Varosi, F., & Beichman, C. A. 1993, ApJ, 409, L5
Bradley, L., Sipocz, B., Robitaille, T., et al. 2018, Astropy/Photutils: V0.5, Zenodo, Zenodo, 10.5281/zenodo.1340699
Brandl, B. R., Bernard-Salas, J., Spoon, H. W. W., et al. 2006, ApJ, 653, 1129
Brandl, B. R., Martín-Hernández, N. L., Schaerer, D., Rosenberg, M., & van der Werf, P. P. 2012, A&A, 543, A61
Brassington, N. J., Zezas, A., Ashby, M. L. N., et al. 2015, ApJS, 218, 6
Brauher, J. R., Dale, D. A., & Helou, G. 2008, ApJS, 178, 280
Breiman, L. 2001, Machine Learning, 45, 5
Brinch, C. & Hogerheijde, M. R. 2010, A&A, 523, A25
Brinchmann, J., Charlot, S., White, S. D. M., et al. 2004, MNRAS, 351, 1151
Brisbin, D., Ferkinhoff, C., Nikola, T., et al. 2015, ApJ, 799, 13
Brunet, C., Allen, M., Brouty, M., et al. 2018, in European Physical Journal Web of Conferences, Vol. 186, European Physical Journal Web of Conferences, 02004
Bruzual, G. & Charlot, S. 2003, MNRAS, 344, 1000
Buat, V., Ciesla, L., Boquien, M., Małek, K., & Burgarella, D. 2019, A&A, 632, A79
Buat, V., Mountrichas, G., Yang, G., et al. 2021, A&A, 654, A93
Burgarella, D., Buat, V., & Iglesias-Páramo, J. 2005, MNRAS, 360, 1413
Bussmann, R. S., Pérez-Fournon, I., Amber, S., et al. 2013, ApJ, 779, 25
Buta, R. J., Sheth, K., Athanassoula, E., et al. 2015, ApJS, 217, 32
Cairós, L. M., Caon, N., Zurita, C., et al. 2010, A&A, 520, A90
Calistro Rivera, G., Alexander, D. M., Rosario, D. J., et al. 2021, A&A, 649, A102
Calistro Rivera, G., Lusso, E., Hennawi, J. F., & Hogg, D. W. 2016, ApJ, 833, 98
Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, ApJ, 533, 682
Cameron, M., Storey, J. W. V., Rotaciuc, V., et al. 1993, ApJ, 419, 136
Camps, P. & Baes, M. 2015, Astronomy and Computing, 9, 20
Camps, P. & Baes, M. 2020, Astronomy and Computing, 31, 100381
Capak, P. L., Carilli, C., Jones, G., et al. 2015, Nature, 522, 455
Caplar, N. & Tacchella, S. 2019, MNRAS, 487, 3845
Cappellari, M., McDermid, R. M., Alatalo, K., et al. 2012, Nature, 484, 485
Cardamone, C., Schawinski, K., Sarzi, M., et al. 2009, MNRAS, 399, 1191
Carilli, C. L. & Walter, F. 2013, ARA&A, 51, 105
Carleo, G., Cirac, I., Cranmer, K., et al. 2019, Reviews of Modern Physics, 91, 045002
Carniani, S., Ferrara, A., Maiolino, R., et al. 2020, MNRAS, 499, 5136
Carniani, S., Maiolino, R., Pallottini, A., et al. 2017, A&A, 605, A42
Casey, C. M. 2012, MNRAS, 425, 3094
Chabrier, G. 2003, PASP, 115, 763
Chae, K.-H., Lelli, F., Desmond, H., et al. 2020, ApJ, 904, 51
Chang, Y.-Y., van der Wel, A., da Cunha, E., & Rix, H.-W. 2015, ApJS, 219, 8
Chen, S., Berton, M., La Mura, G., et al. 2018, A&A, 615, A167
Chen, T. & Guestrin, C. 2016, arXiv e-prints, arXiv:1603.02754
Cheng, C., Cao, X., Lu, N., et al. 2020, ApJ, 898, 33
Ching, T. C., Li, D., Heiles, C., et al. 2022, Nature, 601, 49

Cibinel, A., Daddi, E., Sargent, M. T., et al. 2019, MNRAS, 485, 5631
Cid Fernandes, R., Stasińska, G., Mateus, A., & Vale Asari, N. 2011, MNRAS, 413, 1687
Ciesla, L., Boquien, M., Boselli, A., et al. 2014, A&A, 565, A128
Ciesla, L., Charmandaris, V., Georgakakis, A., et al. 2015, A&A, 576, A10
Ciesla, L., Elbaz, D., Schreiber, C., Daddi, E., & Wang, T. 2018, A&A, 615, A61
Cigan, P., Young, L., Cormier, D., et al. 2016, AJ, 151, 14
Clark, C. J. R., Verstocken, S., Bianchi, S., et al. 2018, A&A, 609, A37
Cohen, M., Megeath, S. T., Hammersley, P. L., Martín-Luis, F., & Stauffer, J. 2003a, AJ, 125, 2645
Cohen, M., Wheaton, W. A., & Megeath, S. T. 2003b, AJ, 126, 1090
Cole, S., Lacey, C. G., Baugh, C. M., & Frenk, C. S. 2000, MNRAS, 319, 168
Collister, A. A. & Lahav, O. 2004, PASP, 116, 345
Cooray, A., Calanog, J., Wardlow, J. L., et al. 2014, ApJ, 790, 40
Coppin, K. E. K., Danielson, A. L. R., Geach, J. E., et al. 2012, MNRAS, 427, 520
Cormier, D., Abel, N. P., Hony, S., et al. 2019, A&A, 626, A23
Cormier, D., Lebouteiller, V., Madden, S. C., et al. 2012, A&A, 548, A20
Cormier, D., Madden, S. C., Lebouteiller, V., et al. 2015, A&A, 578, A53
Cortés, J. R., Kenney, J. D. P., & Hardy, E. 2008, ApJ, 683, 78
Cowie, L. L., Songaila, A., Hu, E. M., & Cohen, J. G. 1996, AJ, 112, 839
Cox, D. P. & Smith, B. W. 1974, ApJ, 189, L105
Crain, R. A., Schaye, J., Bower, R. G., et al. 2015, MNRAS, 450, 1937
Croton, D. J., Springel, V., White, S. D. M., et al. 2006, MNRAS, 365, 11
Croxall, K. V., Smith, J. D., Pellegrini, E., et al. 2017, ApJ, 845, 96
Cunningham, D. J. M., Chapman, S. C., Aravena, M., et al. 2020, MNRAS, 494, 4090
Cusumano, G., La Parola, V., Segreto, A., et al. 2010, A&A, 524, A64
da Cunha, E., Charlot, S., & Elbaz, D. 2008, MNRAS, 388, 1595
D'Abrusco, R., Álvarez Crespo, N., Massaro, F., et al. 2019, ApJS, 242, 4
D'Abrusco, R., Massaro, F., Paggi, A., et al. 2014, ApJS, 215, 14
Daddi, E., Dickinson, M., Morrison, G., et al. 2007, ApJ, 670, 156
Dai, Y. S., Wilkes, B. J., Bergeron, J., et al. 2018, MNRAS, 478, 4238
Dale, D. A., Bendo, G. J., Engelbracht, C. W., et al. 2005, ApJ, 633, 857
Dale, D. A., Helou, G., Magdis, G. E., et al. 2014, ApJ, 784, 83
Dalla Vecchia, C. & Schaye, J. 2012, MNRAS, 426, 140
Dasyra, K. M., Ho, L. C., Netzer, H., et al. 2011, ApJ, 740, 94
Davé, R., Anglés-Alcázar, D., Narayanan, D., et al. 2019, MNRAS, 486, 2827
Davé, R., Thompson, R., & Hopkins, P. F. 2016, MNRAS, 462, 3265
Dayal, P. & Ferrara, A. 2018, Phys. Rep., 780, 1
De Breuck, C., Weiß, A., Béthermin, M., et al. 2019, A&A, 631, A167
De Looze, I., Baes, M., Bendo, G. J., Cortese, L., & Fritz, J. 2011, MNRAS, 416, 2712
De Looze, I., Cormier, D., Lebouteiller, V., et al. 2014, A&A, 568, A62
De Rossi, M. E., Bower, R. G., Font, A. S., Schaye, J., & Theuns, T. 2017, MNRAS, 472, 3354
de Vaucouleurs, G. 1959, Handbuch der Physik, 53, 275
de Vaucouleurs, G. 1961, ApJS, 5, 233
de Vaucouleurs, G. 1977, in Evolution of Galaxies and Stellar Populations, ed. B. M. Tinsley & D. C. Larson, Richard B. Gehret, 43
de Vaucouleurs, G., de Vaucouleurs, A., Corwin, Herold G., J., et al. 1991, Third Reference Catalogue of Bright Galaxies
Decarli, R., Walter, F., Carilli, C., et al. 2014, ApJ, 782, L17
Decarli, R., Walter, F., Neri, R., et al. 2012, ApJ, 752, 2
Decarli, R., Walter, F., Venemans, B. P., et al. 2017, Nature, 545, 457
Decataldo, D., Lupi, A., Ferrara, A., Pallottini, A., & Fumagalli, M. 2020, MNRAS, 497, 4718
Deharveng, L., Schuller, F., Anderson, L. D., et al. 2010, A&A, 523, A6
Dessauges-Zavadsky, M., Richard, J., Combes, F., et al. 2019, Nature Astronomy, 3, 1115

Di Matteo, T., Springel, V., & Hernquist, L. 2005, Nature, 433, 604
Díaz-Santos, T., Armus, L., Charmandaris, V., et al. 2017, ApJ, 846, 32
Díaz-Santos, T., Armus, L., Charmandaris, V., et al. 2013, ApJ, 774, 68
Dietrich, J., Weiner, A. S., Ashby, M. L. N., et al. 2018, MNRAS, 480, 3562
Dixon, T. G. & Joseph, R. D. 2011, ApJ, 740, 99
Doi, M., Tanaka, M., Fukugita, M., et al. 2010, AJ, 139, 1628
Dolag, K., Borgani, S., Murante, G., & Springel, V. 2009, MNRAS, 399, 497
Domínguez Sánchez, H., Huertas-Company, M., Bernardi, M., Tuccillo, D., & Fischer, J. L. 2018, MNRAS, 476, 3661
Dong, X. Y., Wu, X.-B., Ai, Y. L., et al. 2018, AJ, 155, 189
Donley, J. L., Koekemoer, A. M., Brusa, M., et al. 2012, ApJ, 748, 142
Dopita, M. A., Pereira, M., Kewley, L. J., & Capaccioli, M. 2002, ApJS, 143, 47
Draine, B. T. 2011, Physics of the Interstellar and Intergalactic Medium
Duarte Puertas, S., Vilchez, J. M., Iglesias-Páramo, J., et al. 2017, A&A, 599, A71
Dullemond, C. P., Juhasz, A., Pohl, A., et al. 2012, RADMC-3D: A multi-purpose radiative transfer tool, Astrophysics Source Code Library, record ascl:1202.015
Dullemond, C. P. & van Bemmel, I. M. 2005, A&A, 436, 47
Eales, S., de Vis, P., Smith, M. W. L., et al. 2017, MNRAS, 465, 3125
Efstathiou, A. 2006, MNRAS, 371, L70
Efstathiou, A., Hough, J. H., & Young, S. 1995, MNRAS, 277, 1134
Efstathiou, A. & Rowan-Robinson, M. 1995, MNRAS, 273, 649
Elbaz, D., Daddi, E., Le Borgne, D., et al. 2007, A&A, 468, 33
Elbaz, D., Dickinson, M., Hwang, H. S., et al. 2011, A&A, 533, A119
Elbaz, D., Leiton, R., Nagar, N., et al. 2018, A&A, 616, A110
Elitzur, M., Ho, L. C., & Trump, J. R. 2014, MNRAS, 438, 3340
Elitzur, M. & Shlosman, I. 2006, ApJ, 648, L101
Ellison, S. 2019, in Linking Galaxies from the Epoch of Initial Star Formation to Today, 37
Ellison, S. L., Teimoorinia, H., Rosario, D. J., & Mendel, J. T. 2016, MNRAS, 455, 370
Ellison, S. L., Viswanathan, A., Patton, D. R., et al. 2019, MNRAS, 487, 2491
Elmegreen, B. G. 1989, ApJ, 344, 306
Emmering, R. T., Blandford, R. D., & Shlosman, I. 1992, ApJ, 385, 460
Engelbracht, C. W., Blaylock, M., Su, K. Y. L., et al. 2007, PASP, 119, 994
Erroz-Ferrer, S., Carollo, C. M., den Brok, M., et al. 2019, MNRAS, 484, 5009
Evans, A. S., Solomon, P. M., Tacconi, L. J., Vavilkin, T., & Downes, D. 2006, AJ, 132, 2398
Fabian, A. C., Vasudevan, R. V., & Gandhi, P. 2008, MNRAS, 385, L43
Faesi, C. M., Lada, C. J., & Forbrich, J. 2018, ApJ, 857, 19
Fahrion, K., Cormier, D., Bigiel, F., et al. 2017, A&A, 599, A9
Faisst, A. L., Capak, P. L., Yan, L., et al. 2017, ApJ, 847, 21
Farrah, D., Lebouteiller, V., Spoon, H. W. W., et al. 2013, ApJ, 776, 38
Faucher-Giguère, C.-A. 2020, MNRAS, 493, 1614
Fazio, G. G., Hora, J. L., Allen, L. E., et al. 2004, ApJS, 154, 10
Ferkinhoff, C., Brisbin, D., Nikola, T., et al. 2011, ApJ, 740, L29
Ferkinhoff, C., Brisbin, D., Nikola, T., et al. 2015, ApJ, 806, 260
Ferkinhoff, C., Hailey-Dunsheath, S., Nikola, T., et al. 2010, ApJ, 714, L147
Ferland, G. J. 1996, Hazy, A Brief Introduction to Cloudy 90
Ferland, G. J., Chatzikos, M., Guzmán, F., et al. 2017, Rev. Mexicana Astron. Astrofis., 53, 385
Ferland, G. J., Korista, K. T., Verner, D. A., et al. 1998, PASP, 110, 761
Ferland, G. J., Porter, R. L., van Hoof, P. A. M., et al. 2013, Rev. Mexicana Astron. Astrofis., 49, 137
Fernández-Ontiveros, J. A., Pérez-Montero, E., Vílchez, J. M., Amorín, R., & Spinoglio, L. 2021, A&A, 652, A23
Fernández-Ontiveros, J. A., Spinoglio, L., Pereira-Santaella, M., et al. 2016, ApJS, 226, 19
Ferrara, A., Vallini, L., Pallottini, A., et al. 2019, MNRAS, 489, 1
Field, G. B., Blackman, E. G., & Keto, E. R. 2011, MNRAS, 416, 710

Field, G. B., Goldsmith, D. W., & Habing, H. J. 1969, ApJ, 155, L149
Fischer, T. C., Crenshaw, D. M., Kraemer, S. B., & Schmitt, H. R. 2013, ApJS, 209, 1
Flesch, E. W. 2015, PASA, 32, e010
Flesch, E. W. 2021, VizieR Online Data Catalog, VII/290
Florez, J., Jogee, S., Sherman, S., et al. 2020, MNRAS, 497, 3273
Förster Schreiber, N. M. & Wuyts, S. 2020, ARA&A, 58, 661
Fraser-McKelvie, A., Brown, M. J. I., & Pimbblet, K. A. 2014, MNRAS, 444, L63
Fraternali, F., Karim, A., Magnelli, B., et al. 2021, A&A, 647, A194
Frenk, C. S. & White, S. D. M. 2012, Annalen der Physik, 524, 507
Friedman, J. H. 2001, Annals of statistics, 1189
Fritz, J., Franceschini, A., & Hatziminaoglou, E. 2006, MNRAS, 366, 767
Fujimoto, S., Oguri, M., Brammer, G., et al. 2021, ApJ, 911, 99
Furlong, M., Bower, R. G., Theuns, T., et al. 2015, MNRAS, 450, 4486
Gaetz, T. J. & Salpeter, E. E. 1983, ApJS, 52, 155
Gámez Rosas, V., Isbell, J. W., Jaffe, W., et al. 2022, Nature, 602, 403
Gandhi, P., Horst, H., Smette, A., et al. 2009, A&A, 502, 457
Gao, F., Wang, L., Efstathiou, A., et al. 2021, A&A, 654, A117
García-Burillo, S., Alonso-Herrero, A., Ramos Almeida, C., et al. 2021, A&A, 652, A98
Gardner, M. W. & Dorling, S. 1998, Atmospheric environment, 32, 2627
Gavazzi, G., Boselli, A., Scodeggio, M., Pierini, D., & Belsole, E. 1999, MNRAS, 304, 595
Gavazzi, R., Cooray, A., Conley, A., et al. 2011, ApJ, 738, 125
Gentile Fusillo, N. P., Gänsicke, B. T., & Greiss, S. 2015, MNRAS, 448, 2260
Genzel, R., Lutz, D., Sturm, E., et al. 1998, ApJ, 498, 579
Gkini, A., Plionis, M., Chira, M., & Koulouridis, E. 2021, A&A, 650, A75
Glenn, J., Bradford, C. M., Rosolowsky, E., et al. 2021, Journal of Astronomical Telescopes, Instruments, and Systems, 7, 034004
Gnedin, N. Y. & Kravtsov, A. V. 2011, ApJ, 728, 88
Goicoechea, J. R. & Le Bourlot, J. 2007, A&A, 467, 1
Goldsmith, P. F. 2019, ApJ, 887, 54
Goldsmith, P. F., Langer, W. D., Pineda, J. L., & Velusamy, T. 2012, ApJS, 203, 13
Goldsmith, P. F., Yıldız, U. A., Langer, W. D., & Pineda, J. L. 2015, ApJ, 814, 133
Golob, A., Sawicki, M., Goulding, A. D., & Coupon, J. 2021, MNRAS, 503, 4136
Gong, Y., Cooray, A., Silva, M., et al. 2012, ApJ, 745, 49
Gonzalez Garcia, M., Le Bourlot, J., Le Petit, F., & Roueff, E. 2008, A&A, 485, 127
González-Martín, O., Masegosa, J., García-Bernete, I., et al. 2019a, ApJ, 884, 10
González-Martín, O., Masegosa, J., García-Bernete, I., et al. 2019b, ApJ, 884, 11
Goulding, A. D. & Alexander, D. M. 2009, MNRAS, 398, 1165
Granato, G. L. & Danese, L. 1994, MNRAS, 268, 235
Gray, W. J. & Scannapieco, E. 2016, ApJ, 818, 198
Gray, W. J. & Scannapieco, E. 2017, ApJ, 849, 132
Gray, W. J., Scannapieco, E., & Kasen, D. 2015, ApJ, 801, 107
Greve, T. R., Vieira, J. D., Weiß, A., et al. 2012, ApJ, 756, 101
Griffin, M. J., Abergel, A., Abreu, A., et al. 2010, A&A, 518, L3
Groves, B., Dopita, M. A., Sutherland, R. S., et al. 2008, ApJS, 176, 438
Gruppioni, C., Berta, S., Spinoglio, L., et al. 2016, MNRAS, 458, 4297
Gruppioni, C., Béthermin, M., Loiacono, F., et al. 2020, A&A, 643, A8
Guillard, P., Ogle, P. M., Emonts, B. H. C., et al. 2012, ApJ, 747, 95
Gullberg, B., De Breuck, C., Vieira, J. D., et al. 2015, MNRAS, 449, 2883
Gullberg, B., Swinbank, A. M., Smail, I., et al. 2018, ApJ, 859, 12
Gunn, J. E., Carr, M., Rockosi, C., et al. 1998, AJ, 116, 3040
Gunn, J. E., Siegmund, W. A., Mannery, E. J., et al. 2006, AJ, 131, 2332
Gupta, M., Sikora, M., & Nalewajko, K. 2016, MNRAS, 461, 2346
Habing, H. J. 1968, Bull. Astron. Inst. Netherlands, 19, 421
Haffner, L. M., Dettmar, R. J., Beckman, J. E., et al. 2009, Reviews of Modern Physics, 81, 969

Hahn, C., Starkenburg, T. K., Choi, E., et al. 2019, ApJ, 872, 160
Haines, C. P., Iovino, A., Krywult, J., et al. 2017, A&A, 605, A4
Harikane, Y., Ouchi, M., Inoue, A. K., et al. 2020, ApJ, 896, 93
Harrington, K. C., Vishwas, A., Weiß, A., et al. 2019, MNRAS, 488, 1489
Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, Nature, 585, 357
Harrison, C. M., Alexander, D. M., Rosario, D. J., Scholtz, J., & Stanley, F. 2021, in Nuclear Activity in Galaxies Across Cosmic Time, ed. M. Pović, P. Marziani, J. Masegosa, H. Netzer, S. H. Negu, & S. B. Tessema, Vol. 356, 199–203
Hashimoto, T., Inoue, A. K., Tamura, Y., et al. 2019, PASJ, 71, 109
Hayward, C. C., Lanz, L., Ashby, M. L. N., et al. 2014, MNRAS, 445, 1598
Heckman, T. M. 1980, A&A, 87, 152
Heintz, K. E., Watson, D., Oesch, P. A., Narayanan, D., & Madden, S. C. 2021, ApJ, 922, 147
Hernán-Caballero, A., Alonso-Herrero, A., Hatziminaoglou, E., et al. 2015, ApJ, 803, 109
Herrera-Camus, R., Bolatto, A., Smith, J. D., et al. 2016, ApJ, 826, 175
Herrera-Camus, R., Bolatto, A. D., Wolfire, M. G., et al. 2015, ApJ, 800, 1
Herrera-Camus, R., Sturm, E., Graciá-Carpio, J., et al. 2018a, ApJ, 861, 94
Herrera-Camus, R., Sturm, E., Graciá-Carpio, J., et al. 2018b, ApJ, 861, 95
Hickox, R. C. & Alexander, D. M. 2018, ARA&A, 56, 625
Higuera-G., M. A. & Ramos P., A. F. 2013, Rev. Mexicana Astron. Astrofis., 49, 301
Higuera-G, M. A., Rodríguez-Ardila, A., & Tejeiro, J. M. 2009, in Astronomical Society of the Pacific Conference Series, Vol. 408, The Starburst-AGN Connection, ed. W. Wang, Z. Yang, Z. Luo, & Z. Chen, 72
Ho, L. C. & Keto, E. 2007, ApJ, 658, 314
Hodge, J. A. & da Cunha, E. 2020, Royal Society Open Science, 7, 200556
Hogerheijde, M. R. & van der Tak, F. F. S. 2000, A&A, 362, 697
Holincheck, A. J., Wallin, J. F., Borne, K., et al. 2016, MNRAS, 459, 720
Hollenbach, D. J. & Tielens, A. G. G. M. 1999, Reviews of Modern Physics, 71, 173
Hönig, S. F. 2019, ApJ, 884, 171
Hönig, S. F. & Kishimoto, M. 2017, ApJ, 838, L20
Hopkins, P. F. 2013, MNRAS, 428, 2840
Hopkins, P. F., Hernquist, L., Cox, T. J., et al. 2006, ApJS, 163, 1
Hopkins, P. F., Kereš, D., Oñorbe, J., et al. 2014, MNRAS, 445, 581
Hopkins, P. F., Younger, J. D., Hayward, C. C., Narayanan, D., & Hernquist, L. 2010, MNRAS, 402, 1693
Houck, J. R., Roellig, T. L., van Cleve, J., et al. 2004, ApJS, 154, 18
Hubble, E. P. 1926, ApJ, 64, 321
Hubble, E. P. 1936, Realm of the Nebulae
Hughes, A., Meidt, S. E., Colombo, D., et al. 2013, ApJ, 779, 46
Hughes, T. M., Ibar, E., Villanueva, V., et al. 2017, A&A, 602, A49
Hui, L. & Gnedin, N. Y. 1997, MNRAS, 292, 27
Hunter, J. D. 2007, Computing in Science & Engineering, 9, 90
Indriolo, N., Geballe, T. R., Oka, T., & McCall, B. J. 2007, ApJ, 671, 1736
Indriolo, N., Neufeld, D. A., Gerin, M., et al. 2015, ApJ, 800, 40
Inoue, A. K., Shimizu, I., Tamura, Y., et al. 2014, ApJ, 780, L18
Inoue, A. K., Tamura, Y., Matsuo, H., et al. 2016, Science, 352, 1559
Inoue, S., Matsuo, H., Yoshida, N., Yajima, H., & Moriwaki, K. 2021, arXiv e-prints, arXiv:2102.10752
Iono, D., Yun, M. S., Elvis, M., et al. 2006, ApJ, 645, L97
Ishibashi, W. & Fabian, A. C. 2016, MNRAS, 463, 1291
James, G., Witten, D., Hastie, T., & Tibshirani, R. 2013, An introduction to statistical learning, Vol. 112 (Springer)
Järvelä, E., Lähteenmäki, A., & León-Tavares, J. 2015, A&A, 573, A76
Jayasinghe, T., Kochanek, C. S., Stanek, K. Z., et al. 2018, MNRAS, 477, 3145
Jenkins, E. B. 2009, ApJ, 700, 1299

Ji, X. & Yan, R. 2022, A&A, 659, A112
Johnson, B. D., Leja, J., Conroy, C., & Speagle, J. S. 2021, ApJS, 254, 22
Johnson, K. E., Hibbard, J. E., Gallagher, S. C., et al. 2007, AJ, 134, 1522
Juvela, M. 2020, A&A, 644, A151
Kamenetzky, J., Rangwala, N., Glenn, J., Maloney, P. R., & Conley, A. 2016, ApJ, 829, 93
Kannan, R., Smith, A., Garaldi, E., et al. 2022, MNRAS, 514, 3857
Karachentsev, I. D. & Makarov, D. A. 1996, AJ, 111, 794
Karachentsev, I. D., Tully, R. B., Wu, P.-F., Shaya, E. J., & Dolphin, A. E. 2014, ApJ, 782, 4
Kass, R. E. & Raftery, A. E. 1995, Journal of the american statistical association, 90, 773
Katsianis, A., Blanc, G., Lagos, C. P., et al. 2017, MNRAS, 472, 919
Katz, H., Galligan, T. P., Kimm, T., et al. 2019, MNRAS, 487, 5902
Katz, H., Kimm, T., Sijacki, D., & Haehnelt, M. G. 2017, MNRAS, 468, 4831
Katz, H., Rosdahl, J., Kimm, T., et al. 2022, MNRAS, 510, 5603
Kauffmann, G., Heckman, T. M., Tremonti, C., et al. 2003a, MNRAS, 346, 1055
Kauffmann, G., Heckman, T. M., White, S. D. M., et al. 2003b, MNRAS, 341, 54
Kaufman, M. J., Wolfire, M. G., & Hollenbach, D. J. 2006, ApJ, 644, 283
Kaufman, M. J., Wolfire, M. G., Hollenbach, D. J., & Luhman, M. L. 1999, ApJ, 527, 795
Kawamuro, T., Ueda, Y., Tazaki, F., & Terashima, Y. 2013, ApJ, 770, 157
Keel, W. C., Kennicutt, R. C., J., Hummel, E., & van der Hulst, J. M. 1985, AJ, 90, 708
Kennicutt, Robert C., J. 1998a, ARA&A, 36, 189
Kennicutt, Robert C., J. 1998b, ApJ, 498, 541
Kennicutt, R. C. & Evans, N. J. 2012, ARA&A, 50, 531
Keremedjiev, M., Hao, L., & Charmandaris, V. 2009, ApJ, 690, 1105
Kessler, M. F., Steinz, J. A., Anderegg, M. E., et al. 1996, A&A, 315, L27
Keto, E. & Rybicki, G. 2010, ApJ, 716, 1315
Keto, E., Rybicki, G. B., Bergin, E. A., & Plume, R. 2004, ApJ, 613, 355
Keto, E. R. 1990, ApJ, 355, 190
Kewley, L. J. & Ellison, S. L. 2008, ApJ, 681, 1183
Kewley, L. J., Groves, B., Kauffmann, G., & Heckman, T. 2006, MNRAS, 372, 961
Kewley, L. J., Nicholls, D. C., & Sutherland, R. S. 2019, ARA&A, 57, 511
Khan-Ali, A., Carrera, F. J., Page, M. J., et al. 2015, MNRAS, 448, 75
Klindt, L., Alexander, D. M., Rosario, D. J., Lusso, E., & Fotopoulou, S. 2019, MNRAS, 488, 3109
Knudsen, K. K., Richard, J., Kneib, J.-P., et al. 2016, MNRAS, 462, L6
Knuth, K. H. 2006, arXiv e-prints, physics/0605197
Komatsu, E., Dunkley, J., Nolta, M. R., et al. 2009, ApJS, 180, 330
Kormendy, J. & Bender, R. 2012, ApJS, 198, 2
Koss, M., Trakhtenbrot, B., Ricci, C., et al. 2017, ApJ, 850, 74
Kramida, A., Yu. Ralchenko, Reader, J., & and NIST ASD Team. 2020, NIST Atomic Spectra Database (ver. 5.8), [Online]. Available: `https://physics.nist.gov/asd` [2021, July 23]. National Institute of Standards and Technology, Gaithersburg, MD.
Krawczyk, C. M., Richards, G. T., Gallagher, S. C., et al. 2015, AJ, 149, 203
Krawczyk, C. M., Richards, G. T., Mehta, S. S., et al. 2013, ApJS, 206, 4
Krolik, J. H. & Begelman, M. C. 1988, ApJ, 329, 702
Kroupa, P. 2001, MNRAS, 322, 231
Kroupa, P. 2012, PASA, 29, 395
Krumholz, M. R. 2014, MNRAS, 437, 1662
Kurtz, S. 2005, in Massive Star Birth: A Crossroads of Astrophysics, ed. R. Cesaroni, M. Felli, E. Churchwell, & M. Walmsley, Vol. 227, 111–119
Lacy, M., Storrie-Lombardi, L. J., Sajina, A., et al. 2004, ApJS, 154, 166
Lagache, G. 2018, in Atacama Large-Aperture Submm/mm Telescope (AtLAST), 22
Lagache, G., Cousin, M., & Chatzikos, M. 2018, A&A, 609, A130
Lagos, C. d. P., Crain, R. A., Schaye, J., et al. 2015, MNRAS, 452, 3815
Lamarche, C., Verma, A., Vishwas, A., et al. 2018, ApJ, 867, 140

LaMassa, S. M., Cales, S., Moran, E. C., et al. 2015, ApJ, 800, 144
Langer, W. D., Pineda, J. L., Goldsmith, P. F., et al. 2021, A&A, 651, A59
Lanz, L., Hayward, C. C., Zezas, A., et al. 2014, ApJ, 785, 39
Lanz, L., Zezas, A., Brassington, N., et al. 2013, ApJ, 768, 90
Lanzuisi, G., Delvecchio, I., Berta, S., et al. 2017, A&A, 602, A123
Larson, R. B. 1981, MNRAS, 194, 809
Le Fèvre, O., Béthermin, M., Faisst, A., et al. 2020, A&A, 643, A1
Le Petit, F., Nehmé, C., Le Bourlot, J., & Roueff, E. 2006, ApJS, 164, 506
Lebouteiller, V., Barry, D. J., Goes, C., et al. 2015, ApJS, 218, 21
Lee, M. M., Nagao, T., De Breuck, C., et al. 2019, ApJ, 883, L29
Lee, M. M., Nagao, T., De Breuck, C., et al. 2021, ApJ, 913, 41
Leisawitz, D., Amatucci, E., Carter, R., et al. 2018, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 10698, Space Telescopes and Instrumentation 2018: Optical, Infrared, and Millimeter Wave, ed. M. Lystrup, H. A. MacEwen, G. G. Fazio, N. Batalha, N. Siegler, & E. C. Tong, 1069815
Leitherer, C., Ekström, S., Meynet, G., et al. 2014, ApJS, 212, 14
Leitherer, C., Li, I. H., Calzetti, D., & Heckman, T. M. 2002, ApJS, 140, 303
Leitherer, C., Vacca, W. D., Conti, P. S., et al. 1996, ApJ, 465, 717
Leja, J., Johnson, B. D., Conroy, C., & van Dokkum, P. 2018, ApJ, 854, 62
Leja, J., Johnson, B. D., Conroy, C., van Dokkum, P. G., & Byler, N. 2017, ApJ, 837, 170
Leslie, S. K., Kewley, L. J., Sanders, D. B., & Lee, N. 2016, MNRAS, 455, L82
Leslie, S. K., Schinnerer, E., Liu, D., et al. 2020, ApJ, 899, 58
Lestrade, J.-F., Carilli, C. L., Thanjavur, K., et al. 2011, ApJ, 739, L30
Leung, T. K. D., Olsen, K. P., Somerville, R. S., et al. 2020, ApJ, 905, 102
Li, J., Wang, R., Cox, P., et al. 2020a, ApJ, 900, 131
Li, Y., Gu, M. F., Yajima, H., Zhu, Q., & Maji, M. 2020b, MNRAS, 494, 1919
Li, Y., Hopkins, P. F., Hernquist, L., et al. 2008, ApJ, 678, 41
Lintott, C., Schawinski, K., Bamford, S., et al. 2011, MNRAS, 410, 166
Lintott, C. J., Schawinski, K., Slosar, A., et al. 2008, MNRAS, 389, 1179
Liu, H.-Y., Yuan, W., Dong, X.-B., Zhou, H., & Liu, W.-J. 2018, ApJS, 235, 40
López-Gonzaga, N., Burtscher, L., Tristram, K. R. W., Meisenheimer, K., & Schartmann, M. 2016, A&A, 591, A47
Luo, W., Yang, X., & Zhang, Y. 2014, ApJ, 789, L16
Lupi, A. & Bovino, S. 2020, MNRAS, 492, 2818
Lupi, A., Pallottini, A., Ferrara, A., et al. 2020, MNRAS, 496, 5160
Lusso, E., Piedipalumbo, E., Risaliti, G., et al. 2019, A&A, 628, L4
Lyu, J. & Rieke, G. H. 2018, ApJ, 866, 92
Madau, P. & Dickinson, M. 2014, ARA&A, 52, 415
Magdis, G. E., Rigopoulou, D., Hopwood, R., et al. 2014, ApJ, 796, 63
Maiolino, R., Carniani, S., Fontana, A., et al. 2015, MNRAS, 452, 54
Maiolino, R., Cox, P., Caselli, P., et al. 2005, A&A, 440, L51
Małek, K., Buat, V., Roehlly, Y., et al. 2018, A&A, 620, A50
Malhotra, S., Kaufman, M. J., Hollenbach, D., et al. 2001, ApJ, 561, 766
Mannucci, F., Belfiore, F., Curti, M., et al. 2021, MNRAS, 508, 1582
Marchesi, S., Civano, F., Elvis, M., et al. 2016, ApJ, 817, 34
Marin, F. 2016, MNRAS, 460, 3679
Martin, D. C., Fanson, J., Schiminovich, D., et al. 2005, ApJ, 619, L1
Martínez-Galarza, J. R., Smith, H. A., Lanz, L., et al. 2016, ApJ, 817, 76
Masoura, V. A., Mountrichas, G., Georgantopoulos, I., & Plionis, M. 2021, A&A, 646, A167
Matthee, J. & Schaye, J. 2019, MNRAS, 484, 915
Matthee, J., Sobral, D., Boogaard, L. A., et al. 2019, ApJ, 881, 124
Matthews, B. W. 1975, Biochimica et Biophysica Acta (BBA)-Protein Structure, 405, 442
McAlpine, S., Helly, J. C., Schaller, M., et al. 2016, Astronomy and Computing, 15, 72
McKee, C. F. 1990, in Astronomical Society of the Pacific Conference Series, Vol. 12, The Evolution of the Interstellar Medium, ed. L. Blitz, 3–29

McKee, C. F. & Ostriker, E. C. 2007, ARA&A, 45, 565
McKee, C. F. & Ostriker, J. P. 1977, ApJ, 218, 148
McKinney, J., Hayward, C. C., Rosenthal, L. J., et al. 2021, ApJ, 921, 55
Merloni, A., Bongiorno, A., Brusa, M., et al. 2014, MNRAS, 437, 3550
Meusinger, H., Hinze, A., & de Hoon, A. 2011, A&A, 525, A37
Mezger, P. G., Mathis, J. S., & Panagia, N. 1982, A&A, 105, 372
Miedema, F. 2022, Open Science: the Very Idea (Springer Nature)
Miettinen, O. 2018, Ap&SS, 363, 197
Milgrom, M. & Sanders, R. H. 2003, ApJ, 599, L25
Miller, T. B., Chapman, S. C., Aravena, M., et al. 2018, Nature, 556, 469
Mitsuhashi, I., Matsuda, Y., Smail, I., et al. 2021, ApJ, 907, 122
Miville-Deschênes, M.-A., Murray, N., & Lee, E. J. 2017, ApJ, 834, 57
Mordini, S., Spinoglio, L., & Fernández-Ontiveros, J. A. 2021, A&A, 653, A36
Mordini, S., Spinoglio, L., & Fernández-Ontiveros, J. A. 2022, PASA, 39, e012
Morgan, W. W. & Mayall, N. U. 1957, PASP, 69, 291
Morisset, C., Delgado-Inglada, G., & Flores-Fajardo, N. 2015, Rev. Mexicana Astron. Astrofis., 51, 103
Moriwaki, K., Yoshida, N., Shimizu, I., et al. 2018, MNRAS, 481, L84
Morrissey, P., Conrow, T., Barlow, T. A., et al. 2007, ApJS, 173, 682
Mountrichas, G., Buat, V., Yang, G., et al. 2021, A&A, 646, A29
Muñoz, J. A. & Oh, S. P. 2016, MNRAS, 463, 2085
Mullaney, J. R., Alexander, D. M., Goulding, A. D., & Hickox, R. C. 2011, MNRAS, 414, 1082
Müller-Sánchez, F., González-Martín, O., Fernández-Ontiveros, J. A., Acosta-Pulido, J. A., & Prieto, M. A. 2010, ApJ, 716, 1166
Murray, N. 2011, ApJ, 729, 133
Naab, T. & Ostriker, J. P. 2017, ARA&A, 55, 59
Nagamine, K., Wolfe, A. M., & Hernquist, L. 2006, ApJ, 647, 60
Nagao, T., Maiolino, R., Marconi, A., & Matsuhara, H. 2011, A&A, 526, A149
Narayanan, D., Cox, T. J., Kelly, B., et al. 2008a, ApJS, 176, 331
Narayanan, D., Cox, T. J., Shirley, Y., et al. 2008b, ApJ, 684, 996
Narayanan, D., Dey, A., Hayward, C. C., et al. 2010, MNRAS, 407, 1701
Narayanan, D. & Krumholz, M. R. 2017, MNRAS, 467, 50
Neeleman, M., Bañados, E., Walter, F., et al. 2019, ApJ, 882, 10
Neeleman, M., Prochaska, J. X., Kanekar, N., & Rafelski, M. 2020, Nature, 581, 269
Nelson, D., Pillepich, A., Springel, V., et al. 2019, MNRAS, 490, 3234
Nenkova, M., Ivezić, Ž., & Elitzur, M. 2002, ApJ, 570, L9
Nenkova, M., Sirocky, M. M., Nikutta, R., Ivezić, Ž., & Elitzur, M. 2008, ApJ, 685, 160
Netzer, H. 2015, ARA&A, 53, 365
Neugebauer, G., Habing, H. J., van Duinen, R., et al. 1984, ApJ, 278, L1
Noeske, K. G., Weiner, B. J., Faber, S. M., et al. 2007, ApJ, 660, L43
Noll, S., Burgarella, D., Giovannoli, E., et al. 2009, A&A, 507, 1793
Novak, M., Smolčić, V., Delhaize, J., et al. 2017, A&A, 602, A5
Oberto, A., Loup, C., Allen, M., et al. 2020, in Astronomical Society of the Pacific Conference Series, Vol. 522, Astronomical Data Analysis Software and Systems XXVII, ed. P. Ballester, J. Ibsen, M. Solar, & K. Shortridge, 105
Ogawa, S., Ueda, Y., Tanimoto, A., & Yamada, S. 2021, ApJ, 906, 84
Oh, K., Yi, S. K., Schawinski, K., et al. 2015, ApJS, 219, 1
Olivares E., F., Hamuy, M., Pignata, G., et al. 2010, ApJ, 715, 833
Olsen, K., Greve, T. R., Narayanan, D., et al. 2017, ApJ, 846, 105
Olsen, K., Greve, T. R., Narayanan, D., et al. 2018a, ApJ, 857, 148
Olsen, K., Pallottini, A., Wofford, A., et al. 2018b, Galaxies, 6, 100
Olsen, K. P., Burkhart, B., Mac Low, M.-M., et al. 2021, ApJ, 922, 88
Olsen, K. P., Greve, T. R., Narayanan, D., et al. 2015, ApJ, 814, 76
Osterbrock, D. E. 1977, ApJ, 215, 733

Osterbrock, D. E. 1981, ApJ, 249, 462
Osterbrock, D. E. & Pogge, R. W. 1985, ApJ, 297, 166
Oteo, I., Ivison, R. J., Dunne, L., et al. 2016, ApJ, 827, 34
Oteo, I., Zwaan, M. A., Ivison, R. J., Smail, I., & Biggs, A. D. 2017, ApJ, 837, 182
Padovani, P., Alexander, D. M., Assef, R. J., et al. 2017, A&A Rev., 25, 2
Pallottini, A., Ferrara, A., Bovino, S., et al. 2017a, MNRAS, 471, 4128
Pallottini, A., Ferrara, A., Decataldo, D., et al. 2019, MNRAS, 487, 1689
Pallottini, A., Ferrara, A., Gallerani, S., et al. 2022, MNRAS, 513, 5621
Pallottini, A., Ferrara, A., Gallerani, S., et al. 2017b, MNRAS, 465, 2540
Panessa, F., Castangia, P., Malizia, A., et al. 2020, A&A, 641, A162
Pâris, I., Petitjean, P., Aubourg, É., et al. 2018, A&A, 613, A51
Paturel, G., Fouqué, P., Buta, R., & Garcia, A. M. 1991, A&A, 243, 319
Pavesi, R., Riechers, D. A., Capak, P. L., et al. 2016, ApJ, 832, 151
Peñaloza, C. H., Clark, P. C., Glover, S. C. O., & Klessen, R. S. 2018, MNRAS, 475, 1508
Pearson, W. J., Wang, L., Alpaslan, M., et al. 2019a, A&A, 631, A51
Pearson, W. J., Wang, L., Hurley, P. D., et al. 2018, A&A, 615, A146
Pearson, W. J., Wang, L., Trayford, J. W., Petrillo, C. E., & van der Tak, F. F. S. 2019b, A&A, 626, A49
Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2012, arXiv e-prints, arXiv:1201.0490
Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, Journal of Machine Learning Research, 12, 2825
Pelupessy, F. I. & Papadopoulos, P. P. 2009, ApJ, 707, 954
Pentericci, L., Carniani, S., Castellano, M., et al. 2016, ApJ, 829, L11
Pereira-Santaella, M., Diamond-Stanic, A. M., Alonso-Herrero, A., & Rieke, G. H. 2010, ApJ, 725, 2270
Pérez-González, P. G., Rieke, G. H., Villar, V., et al. 2008, ApJ, 675, 234
Pérez-Torres, M., Mattila, S., Alonso-Herrero, A., Aalto, S., & Efstathiou, A. 2021, A&A Rev., 29, 2
Peterson, B. M. 1997, An Introduction to Active Galactic Nuclei
Petric, A. O., Armus, L., Howell, J., et al. 2011, ApJ, 730, 28
Pettini, M. & Pagel, B. E. J. 2004, MNRAS, 348, L59
Pier, E. A. & Krolik, J. H. 1992, ApJ, 401, 99
Pilbratt, G. L., Riedinger, J. R., Passvogel, T., et al. 2010, A&A, 518, L1
Pillepich, A., Nelson, D., Springel, V., et al. 2019, MNRAS, 490, 3196
Pillepich, A., Springel, V., Nelson, D., et al. 2018, MNRAS, 473, 4077
Pineda, J. L., Fischer, C., Kapala, M., et al. 2018, ApJ, 869, L30
Pineda, J. L., Langer, W. D., & Goldsmith, P. F. 2014, A&A, 570, A121
Pineda, J. L., Langer, W. D., Velusamy, T., & Goldsmith, P. F. 2013, A&A, 554, A103
Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2014a, A&A, 571, A1
Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2014b, A&A, 571, A16
Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, A&A, 594, A13
Ploeckinger, S. & Schaye, J. 2020, MNRAS, 497, 4857
Podigachoski, P., Barthel, P. D., Haas, M., et al. 2015, A&A, 575, A80
Poetrodjojo, H., D'Agostino, J. J., Groves, B., et al. 2019, MNRAS, 487, 79
Poglitsch, A., Waelkens, C., Geis, N., et al. 2010, A&A, 518, L2
Pokhrel, N. R., Simpson, C. E., & Bagetakos, I. 2020, AJ, 160, 66
Popping, G., Narayanan, D., Somerville, R. S., Faisst, A. L., & Krumholz, M. R. 2019, MNRAS, 482, 4906
Pouliasis, E., Mountrichas, G., Georgantopoulos, I., et al. 2020, MNRAS, 495, 1853
Pound, M. W. & Wolfire, M. G. 2008, in Astronomical Society of the Pacific Conference Series, Vol. 394, Astronomical Data Analysis Software and Systems XVII, ed. R. W. Argyle, P. S. Bunclark, & J. R. Lewis, 654
Pound, M. W. & Wolfire, M. G. 2011, PDRT: Photo Dissociation Region Toolbox, Astrophysics Source Code Library, record ascl:1102.022
Prevot, M. L., Lequeux, J., Maurice, E., Prevot, L., & Rocca-Volmerange, B. 1984, A&A,

132, 389
Prince, R., Czerny, B., & Pollo, A. 2021, ApJ, 909, 58
Prugniel, P. & Heraudeau, P. 1998, A&AS, 128, 299
Rahmati, A., Pawlik, A. H., Raičević, M., & Schaye, J. 2013, MNRAS, 430, 2427
Rakshit, S., Stalin, C. S., Chand, H., & Zhang, X.-G. 2017, ApJS, 229, 39
Ramos Almeida, C. & Ricci, C. 2017, Nature Astronomy, 1, 679
Ramos Padilla, A. F., Ashby, M. L. N., Smith, H. A., et al. 2020, MNRAS, 499, 4325
Ramos Padilla, A. F., Wang, L., Malek, K., Efstathiou, A., & Yang, G. 2021a, Reproduction repository for "The viewing angle in AGN SED models, a data-driven analysis", Zenodo, 10.5281/zenodo.5227294
Ramos Padilla, A. F., Wang, L., Małek, K., Efstathiou, A., & Yang, G. 2022, MNRAS, 510, 687
Ramos Padilla, A. F., Wang, L., Małek, K., Efstathiou, A., & Yang, G. 2021b, Complementary dataset of the paper "The viewing angle in AGN SED models, a data-driven analysis", Zenodo, 10.5281/zenodo.5221764
Ramos Padilla, A. F., Wang, L., Ploeckinger, S., van der Tak, F. F. S., & Trager, S. C. 2021, A&A, 645, A133
Ramos Padilla, A. F., Wang, L., van der Tak, F., & Trager, S. 2022, Supplementary material "Diagnosing the interstellar medium of galaxies with far-infrared emission lines II. [C II], [O I], [O III], [N II] and [N III] up to z=6", Zenodo, 10.5281/zenodo.6134281
Ramos Padilla, A. F., Wang, L., van der Tak, F. F. S., & Trager, S. 2022, arXiv e-prints, arXiv:2205.11955
Reissl, S., Brauer, R., Klessen, R. S., & Pellegrini, E. W. 2019, ApJ, 885, 15
Reissl, S., Wolf, S., & Brauer, R. 2016, A&A, 593, A87
Rémy-Ruyer, A., Madden, S. C., Galliano, F., et al. 2013, A&A, 557, A95
Renzini, A. & Peng, Y.-j. 2015, ApJ, 801, L29
Reynolds, J. H. 1920, MNRAS, 80, 746
Ricci, C., Trakhtenbrot, B., Koss, M. J., et al. 2017, Nature, 549, 488
Rieke, G. H., Young, E. T., Engelbracht, C. W., et al. 2004, ApJS, 154, 25
Rigopoulou, D., Pearson, C., Ellison, B., et al. 2021, Experimental Astronomy, 51, 699
Rigopoulou, D., Pereira-Santaella, M., Magdis, G. E., et al. 2018, MNRAS, 473, 20
Rizzo, F., Vegetti, S., Fraternali, F., Stacey, H. R., & Powell, D. 2021, MNRAS, 507, 3952
Rizzo, F., Vegetti, S., Powell, D., et al. 2020, Nature, 584, 201
Robinson, D., Avestruz, C., & Gnedin, N. Y. 2022, ApJ, 936, 50
Robitaille, T. P. & Whitney, B. A. 2010, ApJ, 710, L11
Robotham, A. S. G., Bellstedt, S., Lagos, C. d. P., et al. 2020, MNRAS, 495, 905
Rodríguez, Ó., Clocchiatti, A., & Hamuy, M. 2014, AJ, 148, 107
Rodríguez-Puebla, A., Behroozi, P., Primack, J., et al. 2016, MNRAS, 462, 893
Roelfsema, P. R., Shibai, H., Armus, L., et al. 2018, PASA, 35, e030
Röllig, M., Abel, N. P., Bell, T., et al. 2007, A&A, 467, 187
Röllig, M., Ossenkopf, V., Jeyakumar, S., Stutzki, J., & Sternberg, A. 2006, A&A, 451, 917
Röllig, M. & Ossenkopf-Okada, V. 2022, A&A, 664, A67
Rosas-Guevara, Y. M., Bower, R. G., Schaye, J., et al. 2015, MNRAS, 454, 1038
Rosenberg, M. J. F., van der Werf, P. P., Aalto, S., et al. 2015, ApJ, 801, 72
Rowan-Robinson, M., Oliver, S., Wang, L., et al. 2016, MNRAS, 461, 1100
Rowan-Robinson, M., Wang, L., Farrah, D., et al. 2018, A&A, 619, A169
Ruiz, A., Risaliti, G., Nardini, E., Panessa, F., & Carrera, F. J. 2013, A&A, 549, A125
Rybak, M., da Cunha, E., Groves, B., et al. 2021, ApJ, 909, 130
Rybak, M., Zavala, J. A., Hodge, J. A., Casey, C. M., & Werf, P. v. d. 2020, ApJ, 889, L11
Ryden, B. & Pogge, R. W. 2021, Interstellar and Intergalactic Medium (Cambridge University Press)
Saintonge, A., Catinella, B., Cortese, L., et al. 2016, MNRAS, 462, 1749
Salmon, B., Papovich, C., Long, J., et al. 2016, ApJ, 827, 20
Salpeter, E. E. 1955, ApJ, 121, 161
Sandage, A. 2005, ARA&A, 43, 581

Sanders, D. B. & Mirabel, I. F. 1996, ARA&A, 34, 749
Sanders, R. L., Shapley, A. E., Zhang, K., & Yan, R. 2017, ApJ, 850, 136
Saturni, F. G., Trevese, D., Vagnetti, F., Perna, M., & Dadina, M. 2016, A&A, 587, A43
Satyapal, S., Kamal, L., Cann, J. M., Secrest, N. J., & Abel, N. P. 2021, ApJ, 906, 35
Satyapal, S., Secrest, N. J., Ricci, C., et al. 2017, ApJ, 848, 126
Schaerer, D., Ginolfi, M., Béthermin, M., et al. 2020, A&A, 643, A3
Schaller, G., Schaerer, D., Meynet, G., & Maeder, A. 1992, A&AS, 96, 269
Schaye, J. 2001, ApJ, 559, 507
Schaye, J., Crain, R. A., Bower, R. G., et al. 2015, MNRAS, 446, 521
Schaye, J. & Dalla Vecchia, C. 2008, MNRAS, 383, 1210
Schmidt, E. O., Ferreiro, D., Vega Neme, L., & Oio, G. A. 2016, A&A, 596, A95
Schmidt, M. 1959, ApJ, 129, 243
Scott, A. E. & Stewart, G. C. 2014, MNRAS, 438, 2253
Scott, D. W. 2015, Multivariate Density Estimation: Theory, Practice, and Visualization
Seon, K.-I., Edelstein, J., Korpela, E., et al. 2011, ApJS, 196, 15
Serra, P., Amblard, A., Temi, P., et al. 2011, ApJ, 740, 22
Shaw, G., Ferland, G. J., & Ploeckinger, S. 2020, Research Notes of the American Astronomical Society, 4, 78
Shi, Y., Rieke, G. H., Ogle, P. M., Su, K. Y. L., & Balog, Z. 2014, ApJS, 214, 23
Shields, G. A. 1990, ARA&A, 28, 525
Siebenmorgen, R., Heymann, F., & Efstathiou, A. 2015, A&A, 583, A120
Silva, A., Marchesini, D., Silverman, J. D., et al. 2018, ApJ, 868, 46
Silva, L., Maiolino, R., & Granato, G. L. 2004, MNRAS, 355, 973
Singh, R., van de Ven, G., Jahnke, K., et al. 2013, A&A, 558, A43
Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, AJ, 131, 1163
Smith, B. J., Wagstaff, P., Struck, C., et al. 2019, AJ, 158, 169
Smith, J. D. T., Armus, L., Dale, D. A., et al. 2007a, PASP, 119, 1133
Smith, J. D. T., Croxall, K., Draine, B., et al. 2017, ApJ, 834, 5
Smith, J. D. T., Draine, B. T., Dale, D. A., et al. 2007b, ApJ, 656, 770
Smith, R. J. 2020, ARA&A, 58, 577
Somerville, R. S. & Davé, R. 2015, ARA&A, 53, 51
Somerville, R. S. & Primack, J. R. 1999, MNRAS, 310, 1087
Souchay, J., Andrei, A. H., Barache, C., et al. 2015, A&A, 583, A75
Speagle, J. S., Steinhardt, C. L., Capak, P. L., & Silverman, J. D. 2014, ApJS, 214, 15
Spergel, D. N., Verde, L., Peiris, H. V., et al. 2003, ApJS, 148, 175
Spilker, J. S., Marrone, D. P., Aravena, M., et al. 2016, ApJ, 826, 112
Spilker, J. S., Phadke, K. A., Aravena, M., et al. 2020, ApJ, 905, 85
Spinoglio, L., Alonso-Herrero, A., Armus, L., et al. 2017, PASA, 34, e057
Spinoglio, L., Mordini, S., Fernández-Ontiveros, J. A., et al. 2021, PASA, 38, e021
Spinoglio, L., Pereira-Santaella, M., Dasyra, K. M., et al. 2015, ApJ, 799, 21
Springel, V. 2005, MNRAS, 364, 1105
Springel, V., Di Matteo, T., & Hernquist, L. 2005, MNRAS, 361, 776
Springel, V., White, S. D. M., Tormen, G., & Kauffmann, G. 2001, MNRAS, 328, 726
Stacey, G. J., Geis, N., Genzel, R., et al. 1991, ApJ, 373, 423
Stacey, G. J., Hailey-Dunsheath, S., Ferkinhoff, C., et al. 2010, ApJ, 724, 957
Stacey, H. R., McKean, J. P., Powell, D. M., et al. 2021, MNRAS, 500, 3667
Stalevski, M., Fritz, J., Baes, M., Nakos, T., & Popović, L. Č. 2012, MNRAS, 420, 2756
Stalevski, M., Ricci, C., Ueda, Y., et al. 2016, MNRAS, 458, 2288
Stern, D., Assef, R. J., Benford, D. J., et al. 2012, ApJ, 753, 30
Stern, D., Eisenhardt, P., Gorjian, V., et al. 2005, ApJ, 631, 163
Stern, J. & Laor, A. 2012, MNRAS, 426, 2703
Stierwalt, S., Armus, L., Surace, J. A., et al. 2013, ApJS, 206, 1
Sturm, E., Lutz, D., Verma, A., et al. 2002, A&A, 393, 821
Su, S., Kong, X., Li, J., & Fang, G. 2013, ApJ, 778, 10

Suh, H., Civano, F., Hasinger, G., et al. 2019, ApJ, 872, 168
Sun, G., Hensley, B. S., Chang, T.-C., Doré, O., & Serra, P. 2019, ApJ, 887, 142
Sun, J. & Shen, Y. 2015, ApJ, 804, L15
Sutter, J., Dale, D. A., Croxall, K. V., et al. 2019, ApJ, 886, 60
Swinbank, A. M., Karim, A., Smail, I., et al. 2012, MNRAS, 427, 1066
Swinyard, B. M., Ade, P., Baluteau, J. P., et al. 2010, A&A, 518, L4
Tacchella, S., Dekel, A., Carollo, C. M., et al. 2016, MNRAS, 457, 2790
Tacconi, L. J., Genzel, R., & Sternberg, A. 2020, ARA&A, 58, 157
Tadaki, K.-i., Iono, D., Hatsukade, B., et al. 2019, ApJ, 876, 1
Tamayo, D., Silburt, A., Valencia, D., et al. 2016, ApJ, 832, L22
Tanimoto, A., Ueda, Y., Odaka, H., et al. 2019, ApJ, 877, 95
Tarantino, E., Bolatto, A. D., Herrera-Camus, R., et al. 2021, ApJ, 915, 92
Tateuchi, K., Konishi, M., Motohara, K., et al. 2015, ApJS, 217, 1
Tayal, S. S. 2011, ApJS, 195, 12
Taylor, M. B. 2005, in Astronomical Society of the Pacific Conference Series, Vol. 347, Astronomical Data Analysis Software and Systems XIV, ed. P. Shopbell, M. Britton, & R. Ebert, 29
Telles, E. & Terlevich, R. 1995, MNRAS, 275, 1
Tharwat, A. 2020, Applied Computing and Informatics
The EAGLE team. 2017, arXiv e-prints, arXiv:1706.09899
Theuns, T., Leonard, A., Efstathiou, G., Pearce, F. R., & Thomas, P. A. 1998, MNRAS, 301, 478
Thob, A. C. R., Crain, R. A., McCarthy, I. G., et al. 2019, MNRAS, 485, 972
Thorne, J., Robotham, A., Davies, L., & Bellstedt, S. 2022, AGN Unification Diagram, Zenodo, 10.5281/zenodo.6381013
Thorne, J. E., Robotham, A. S. G., Davies, L. J. M., et al. 2021, MNRAS, 505, 540
Tielens, A. G. G. M. 2010, The Physics and Chemistry of the Interstellar Medium
Tielens, A. G. G. M. & Hollenbach, D. 1985, ApJ, 291, 722
Tinsley, B. M. 1968, ApJ, 151, 547
Tinsley, B. M. 1972, A&A, 20, 383
Toba, Y., Oyabu, S., Matsuhara, H., et al. 2014, ApJ, 788, 45
Toba, Y., Ueda, Y., Gandhi, P., et al. 2021, ApJ, 912, 91
Tomczak, A. R., Quadri, R. F., Tran, K.-V. H., et al. 2016, ApJ, 817, 118
Tommasin, S., Spinoglio, L., Malkan, M. A., & Fazio, G. 2010, ApJ, 709, 1257
Tommasin, S., Spinoglio, L., Malkan, M. A., et al. 2008, ApJ, 676, 836
Trayford, J. W., Theuns, T., Bower, R. G., et al. 2015, MNRAS, 452, 2879
Tress, R. G., Smith, R. J., Sormani, M. C., et al. 2020a, MNRAS, 492, 2973
Tress, R. G., Sormani, M. C., Glover, S. C. O., et al. 2020b, MNRAS, 499, 4455
Tully, R. B., Courtois, H. M., Dolphin, A. E., et al. 2013, AJ, 146, 86
Urry, C. M. & Padovani, P. 1995, PASP, 107, 803
Uzgil, B. D., Bradford, C. M., Hailey-Dunsheath, S., Maloney, P. R., & Aguirre, J. E. 2016, ApJ, 832, 209
Vaddi, S., O'Dea, C. P., Baum, S. A., et al. 2016, ApJ, 818, 182
Vallini, L., Ferrara, A., Pallottini, A., Carniani, S., & Gallerani, S. 2020, MNRAS, 495, L22
Vallini, L., Ferrara, A., Pallottini, A., Carniani, S., & Gallerani, S. 2021, MNRAS, 505, 5543
Vallini, L., Ferrara, A., Pallottini, A., & Gallerani, S. 2017, MNRAS, 467, 1300
Vallini, L., Gallerani, S., Ferrara, A., & Baek, S. 2013, MNRAS, 433, 1567
Vallini, L., Gallerani, S., Ferrara, A., Pallottini, A., & Yue, B. 2015, ApJ, 813, 36
Valtchanov, I., Virdee, J., Ivison, R. J., et al. 2011, MNRAS, 415, 3473
van der Tak, F. F. S., Black, J. H., Schöier, F. L., Jansen, D. J., & van Dishoeck, E. F. 2007, A&A, 468, 627
van der Tak, F. F. S., Madden, S. C., Roelfsema, P., et al. 2018, PASA, 35, e002
Vandenbroucke, B. & Wood, K. 2019, MNRAS, 488, 1977
Vanzella, E., Cristiani, S., Fontana, A., et al. 2004, A&A, 423, 761
Vayner, A., Wright, S. A., Murray, N., et al. 2021, ApJ, 910, 44

Veilleux, S., Rupke, D. S. N., Kim, D. C., et al. 2009, ApJS, 182, 628
Venemans, B. P., Walter, F., Neeleman, M., et al. 2020, ApJ, 904, 130
Véron-Cetty, M. P. & Véron, P. 2010, A&A, 518, A10
Vika, M., Ciesla, L., Charmandaris, V., Xilouris, E. M., & Lebouteiller, V. 2017, A&A, 597, A51
Villanueva, V., Ibar, E., Hughes, T. M., et al. 2017, MNRAS, 470, 3775
Villarroel, B., Nyholm, A., Karlsson, T., et al. 2017, ApJ, 837, 110
Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, Nature Methods, 17, 261
Vishwas, A., Ferkinhoff, C., Nikola, T., et al. 2018, ApJ, 856, 174
Vogelsberger, M., Genel, S., Springel, V., et al. 2014, MNRAS, 444, 1518
Wada, K. 2015, ApJ, 812, 82
Walter, F., Riechers, D., Novak, M., et al. 2018, ApJ, 869, L22
Wang, L., Gao, F., Duncan, K. J., et al. 2019a, A&A, 631, A109
Wang, L., Norberg, P., Gunawardhana, M. L. P., et al. 2016, MNRAS, 461, 1898
Wang, L., Pearson, W. J., Cowley, W., et al. 2019b, A&A, 624, A98
Wang, L., Rowan-Robinson, M., Norberg, P., Heinis, S., & Han, J. 2014, MNRAS, 442, 2739
Wang, T.-W., Goto, T., Kim, S. J., et al. 2020, MNRAS, 499, 4068
Wardlow, J. L., Cooray, A., Osage, W., et al. 2017, ApJ, 837, 12
Weaver, K. A., Meléndez, M., Mushotzky, R. F., et al. 2010, ApJ, 716, 1151
Webster, R. L., Francis, P. J., Petersont, B. A., Drinkwater, M. J., & Masci, F. J. 1995, Nature, 375, 469
Wenger, M., Ochsenbein, F., Egret, D., et al. 2000, A&AS, 143, 9
Wes McKinney. 2010, in Proceedings of the 9th Python in Science Conference, ed. Stéfan van der Walt & Jarrod Millman, 56 – 61
Westhues, C., Haas, M., Barthel, P., et al. 2016, AJ, 151, 120
Whitaker, K. E., Franx, M., Leja, J., et al. 2014, ApJ, 795, 104
Whitaker, K. E., van Dokkum, P. G., Brammer, G., & Franx, M. 2012, ApJ, 754, L29
Wiersma, R. P. C., Schaye, J., & Smith, B. D. 2009a, MNRAS, 393, 99
Wiersma, R. P. C., Schaye, J., Theuns, T., Dalla Vecchia, C., & Tornatore, L. 2009b, MNRAS, 399, 574
Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., et al. 2016, Scientific data, 3, 1
Wilson, C. D., Elmegreen, B. G., Bemis, A., & Brunetti, N. 2019, ApJ, 882, 5
Winkler, H. 1992, MNRAS, 257, 677
Wolfire, M. G., Hollenbach, D., McKee, C. F., Tielens, A. G. G. M., & Bakes, E. L. O. 1995, ApJ, 443, 152
Wolfire, M. G., McKee, C. F., Hollenbach, D., & Tielens, A. G. G. M. 2003, ApJ, 587, 278
Wolfire, M. G., Vallini, L., & Chevance, M. 2022, ARA&A, 60, 247
Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, AJ, 140, 1868
Wu, Y.-Z., Zhao, Y.-H., & Meng, X.-M. 2011, ApJS, 195, 17
Yajima, H., Li, Y., Zhu, Q., & Abel, T. 2012, MNRAS, 424, 884
Yan, R. & Blanton, M. R. 2012, ApJ, 747, 61
Yang, G., Boquien, M., Brandt, W. N., et al. 2022a, ApJ, 927, 192
Yang, G., Boquien, M., Buat, V., et al. 2020, MNRAS, 491, 740
Yang, G., Papovich, C., Bagley, M. B., et al. 2021a, ApJ, 908, 144
Yang, S. & Lidz, A. 2020, MNRAS, 499, 3417
Yang, S., Lidz, A., & Popping, G. 2021b, MNRAS, 504, 723
Yang, S., Popping, G., Somerville, R. S., et al. 2022b, ApJ, 929, 140
Yang, S., Somerville, R. S., Pullen, A. R., et al. 2021c, ApJ, 911, 132
York, D. G., Adelman, J., Anderson, John E., J., et al. 2000, AJ, 120, 1579
Yuan, T. T., Kewley, L. J., & Sanders, D. B. 2010, ApJ, 709, 884
Yun, M. S., Reddy, N. A., & Condon, J. J. 2001, ApJ, 554, 803
Zanella, A., Daddi, E., Magdis, G., et al. 2018, MNRAS, 481, 1976
Zavala, J. A., Casey, C. M., Scoville, N., et al. 2019, ApJ, 887, 183
Zenocratti, L. J., De Rossi, M. E., Lara-López, M. A., & Theuns, T. 2020, MNRAS, 496, L33

Zhang, Z.-Y., Ivison, R. J., George, R. D., et al. 2018, MNRAS, 481, 59
Zhao, Y., Lu, N., Xu, C. K., et al. 2016a, ApJ, 819, 69
Zhao, Y., Yan, L., & Tsai, C.-W. 2016b, ApJ, 824, 146
Zhou, H., Wang, T., Yuan, W., et al. 2006, ApJS, 166, 128
Zou, F., Yang, G., Brandt, W. N., & Xue, Y. 2019, ApJ, 878, 11
Zucker, C., Walker, L. M., Johnson, K., et al. 2016, ApJ, 821, 113

Samenvatting

De ruimte tussen de sterren in het heelal is geen volledig vacuüm, er is ook materie in deze ruimte. Deze materie is slechts een zeer kleine fractie van wat we op aarde gewend zijn. Bijvoorbeeld, de hoeveelheid materie per kubieke meter (ook wel dichtheid genoemd) lucht in onze atmosfeer vergeleken met de ruimte is als het vergelijken van een menselijk lichaam met een enkel atoom. Daarom zijn deze hoeveelheden materie extreem verschillend, en hun omstandigheden zijn ook verschillend.

Bij die zeer lage dichtheden vind je gas en stof. Al dat gas en stof tussen de sterren noemen we het interstellaire medium (of kortweg ISM). De temperaturen in het ISM zijn nauw gerelateerd aan hun dichtheden. In sommige gevallen, wanneer de dichtheden vertienvoudigen, kan de temperatuur bijna duizendvoudig dalen. Deze temperatuurveranderingen zijn erg belangrijk om de voorwaarden van het ISM te bepalen, door middel van verwarmings- en koelprocessen. Die processen zijn over het algemeen in nauw evenwicht, zodat verschillende ISM-regimes naast elkaar kunnen bestaan. Daarvoor moet echter wat energie worden "betaald" om vergelijkbare temperaturen te behouden. Dit lijkt sterk op processen in het menselijk lichaam, waar koeling en verwarming worden uitgedrukt in termen van zweten (afkoelen) of bibberen (verwarmen) om een gewenste lichaamstemperatuur te behouden.

In het ISM zijn deze verhittingsprocessen vaak gerelateerd aan energetische gebeurtenissen. Denk bijvoorbeeld aan de straling die sterren kunnen uitzenden door energetische fotonen, of nog heftiger, kosmische explosies als sterren sterven. Die processen zullen sommige van de atomen of moleculen exciteren (verwarmen). Wanneer deze atomen uiteindelijk de-exciteren, zenden ze fotonen uit, waardoor energie verder weg wordt getransporteerd en daardoor het ISM afkoelt. Het meest voorkomende koelproces is wat we fijnstructuur koeling noemen. Deze fijne structuur hangt samen met kleine energie verschillen binnen de energieniveaus van atomen die ervoor zorgen dat fotonen met een bepaalde energie ontsnappen. In het elektromagnetische spectrum, waar de energie van de fotonen wordt vergeleken met een bepaalde golflengte of frequentie, komen deze fijne structuren in zeer specifieke gebieden voor door emissie- of absorptie kenmerken. Het feit dat deze kenmerken straling uitzenden of absorberen in een zeer specifiek golflengtebereik, heeft hen ook de naam van spectraallijnen gegeven.

Hoewel deze spectraallijnen in verschillende golflengtebereiken van het elektromagnetische spectrum kunnen voorkomen, van gammastraling tot radio, bevat het infrarode deel van het elektromagnetische spectrum de belangrijkste fijnstructuur lijnen voor koeling bij emissie. De belangrijkste emissielijn voor koeling wordt geproduceerd door het koolstofatoom. Wanneer dit atoom wordt geëxciteerd (geïoniseerd) en vervolgens gedeëxciteerd, fungeert het als een thermostaat om de temperatuur op een stabiel niveau te houden. Deze emissielijn is belangrijk tussen warme ($\sim$1000 K) en koude ($\sim$10 K) temperaturen in het ISM. Een andere belangrijke emissie van infrarood lijnen komt van het zuurstofatoom; deze emissielijn bestrijkt echter een ander temperatuur-

bereik, van ∼10000 K tot ∼1000 K. Daarom zullen de fysieke omstandigheden van het ISM in de verschillende emissielijnen liggen en het vergelijken ervan is cruciaal om die omstandigheden te herstellen van ruimte waarnemingen.

Dan, we zijn in staat om de processen die op grotere schalen zoals sterrenstelsels plaatsvinden te begrijpen met de informatie die we over het ISM uit de emissielijnen kunnen halen. In sterrenstelsels, zoals de Melkweg, is het mogelijk om deze lijnen te observeren en hun globale fysieke eigenschappen te herstellen. Een interessant gebied binnen sterrenstelsels is hun kern, waar gewelddadige processen en intense stralingen het ISM verstoren. Deze straling is afkomstig van het superzware zwarte gat in het centrum van sterrenstelsels en de omringende materie die erin valt en zich ophoopt in de vorm van een schijf. Nabijgelegen gaswolken worden beïnvloed door de rotatie van deze schijf en veranderen de signalen die we krijgen van de emissielijnen. Deze lijnen kunnen breed of smal zijn, afhankelijk van het gebied van de kernen waar we naar kijken. Ten slotte omringt een grotere stofstructuur de vorige in een “ donut-achtige ” vorm, waardoor de straling die uit deze gebieden komt, wordt verdoezeld vanwege de zichtlijn.

Waarnemingen in het infrarode elektromagnetische spectrum kunnen helpen bij het karakteriseren van die structuren en de gasprocessen van het ISM in sterrenstelsels. De hoeveelheden data van die waarnemingen zijn momenteel echter beperkt en gemakkelijk beschikbaar. Bovendien observeren we quasi-statische beelden van het heelal, die onze analyse van de evolutie van dergelijke processen beperken. Daarom is het noodzakelijk om rekenhulpmiddelen te gebruiken die zullen fungeren als onze laboratoria om onze experimenten uit te voeren. Deze hulpmiddelen moeten rekening houden met alle fysieke kennis die we momenteel weten over de processen van het heelal. Als ze dat eenmaal hebben, kunnen ze worden gebruikt om het waargenomen spectrum van sterrenstelsels te reproduceren en te voorspellen wat er in de toekomst zou worden verwacht.

Dit proefschrift

Het belangrijkste doel van dit proefschrift is om de gasprocessen van sterrenstelsels door de kosmische tijd heen te begrijpen, met de nadruk op fysieke parameters die stervorming beschrijven. Daarom, in Hoofdstuk 1, introduceer ik de achtergrondinformatie om de inhoud van dit proefschrift te begrijpen, waar ik de fysieke processen binnen het ISM, enkele belangrijke kenmerken van sterrenstelsels en de hulpmiddelen die ik gebruik voor de analyse diepgaand toelicht.

In Hoofdstukken 2–4 concentreer ik me op het ISM van sterrenstelsels. In Hoofdstuk 2 beschrijf ik hoe ik simulaties kan gebruiken om de emissielijn van het koolstofatoom in het infrarood te berekenen. Om dit te doen, implementeer ik een fysiek gemotiveerd ISM-model dat wordt toegepast in gesimuleerde sterrenstelsels die in kosmologische tijd vergelijkbaar zijn met de Melkweg. Vervolgens breidde ik het ISM-model uit naar andere infrarood lijnen en oudere kosmologische tijden in Hoofdstuk 3. Die kosmologische tijden brengen ons terug in de tijd, dicht bij de vorming van de eerste sterrenstelsels. Met die informatie presenteer ik een web-app die kan worden gebruikt om schattingen te geven van de fysieke eigenschappen van sterrenstelsels op basis

van waargenomen lijn emissies (Hoofdstuk 4). Van al deze Hoofdstukken vergelijk ik onze resultaten met informatie die is verzameld uit waarnemingen in vergelijkbare golflengtebereiken.

In Hoofdstukken 5 en 6 concentreer ik me op verschillende soorten sterrenstelsels en hun elektromagnetische spectrum vorm. In Hoofdstuk 5 gebruik ik een tool om de fysieke eigenschappen van de sterrenstelsels te schatten op basis van de vorm van het elektromagnetische spectrum dat de golflengtebereiken tussen ultraviolet en ver-infrarood bestrijkt. Deze fysische eigenschappen worden geanalyseerd rekening houdend met hoe intens de stralingsprocessen in het centrum van de sterrenstelsels zijn. Ik bestudeer sterrenstelsels die op elkaar inwerken en in de toekomst zullen samensmelten tot één. Ten slotte analyseer ik hoe tools die rekening houden met de vorm van het elektromagnetische spectrum een specifiek type sterrenstelsel kunnen classificeren om te begrijpen waar ze vandaan komen met behulp van openbaar beschikbare data (Hoofdstuk 6).

Met dit proefschrift over "Diagnose van het ISM van sterrenstelsels en energetische processen in een kosmologische context" verwacht ik de lezer een idee te geven van hoe interessant sommige fysieke processen in sterrenstelsels zijn en hoe ze in de toekomst kunnen worden gebruikt om te begrijpen hoe de Universum werk.

Summary for non-experts

The space between the stars in the Universe is not a complete vacuum, there is also matter in this space. This matter is only a very small fraction of what we are used to on Earth. For example, the amount of matter per cubic metre (also known as density) of air in our atmosphere compared to outer space is like comparing a human body with a single atom. Therefore, these amounts of matter are extremely different, and their conditions are as well different.

Gas and dust can be found at those very low densities. All that gas and dust between the stars is what we called the interstellar medium (or ISM for short). The temperatures in the ISM are tightly related to their densities. In some cases, when densities increase tenfold, temperatures can drop by almost a thousandfold. These temperature changes are very important to define the conditions of the ISM, through heating and cooling processes. Those processes are generally in close equilibrium so that different ISM regimes can coexist. However, for that, some energy needs to be "paid" to maintain similar temperatures. This is very similar to processes found in the human body, where cooling and heating are expressed in terms of sweating (cooling) or shivering (heating) to maintain a desirable body temperature.

In the ISM, these heating processes are often related to energetic events. Think for example of the radiation that stars can emit through energetic photons, or even more violent, cosmic explosions when stars die. Those processes will excite (heat) some of the atoms or molecules. When these atoms eventually de-excite, they emit photons, transporting energy further away and therefore cooling the ISM. The most common cooling process is what we call fine-structure cooling. This fine structure is related to small energy differences within the energy levels of atoms that cause photons to escape with a certain energy. In the electromagnetic spectrum, where the energy of the photons is compared with a certain wavelength or frequency, these fine structures occur in very specific regions through emission or absorption features. The fact that these features emit or absorb radiation in a very specific range of wavelengths has also given them the name of spectral lines.

Although these spectral lines can occur in different wavelength ranges of the electromagnetic spectrum, from gamma rays to radio, the infrared part of the electromagnetic spectrum contains the most important fine-structure lines for cooling in emission. The most important emission line for cooling is produced by the carbon atom. When this atom is excited (ionised) and then de-excited, it acts as a thermostat to keep temperatures at a stable level. This emission line is important between warm ($\sim$1000 K) and cold ($\sim$10 K) temperatures in the ISM. Another important infrared line emission comes from the oxygen atom; however, this emission line covers a different range of temperatures, from $\sim$10000 K to $\sim$1000 K. Therefore, the physical conditions of the ISM will reside in the different emission lines and comparing them is crucial to recover those conditions from space observations.

Thanks to this, we are able to understand the processes that occur on larger scales like galaxies with the information we can get about the ISM from the emission lines. In galaxies, like the Milky Way, it is possible to observe these lines and recover their global physical properties. One region of interest inside galaxies is their nucleus, where violent processes and intense radiations perturb the ISM. This radiation comes from the supermassive black hole at the centre of galaxies and the surrounding matter that falls into it and accumulates in the form of a disc. Nearby gas clouds are affected by the rotation of this disc and change the signals we get from the emission lines. These lines can be broad or narrow depending on the region of the nuclei we are looking at. Finally, a larger dust structure surrounds the previous ones in a "doughnut-like" shape, obscuring the radiation coming from these regions depending on the line-of-sight.

Observations in the infrared electromagnetic spectrum can help characterise those structures and the gas processes of the ISM in galaxies. However, the amounts of data from those observations are currently limited and are not readily available. Furthermore, we are observing quasi-static images of the Universe, which limit our analysis of the evolution of such processes. Therefore, it is necessary to use computational tools that will function as our laboratories to run our experiments. These tools must consider all the physical knowledge that we currently know about the processes of the Universe. Once they have that, they can be used to reproduce the observed spectrum of galaxies and predict what would be expected to observe in the future.

This thesis

The main goal of this PhD thesis is to understand the gas processes of galaxies through cosmic time, focusing on physical parameters that describe star formation. Therefore, in Chapter 1, I introduce the background information to understand the contents of this thesis, where I explain in depth the physical processes within the ISM, some important characteristics of galaxies and the tools I use for the analysis.

In chapters 2–4 I focus on the ISM of galaxies. In Chapter 2, I describe how to use simulations to calculate the emission line of the carbon atom in the infrared. To do this, I implement a physically motivated ISM model that is applied in simulated galaxies similar in cosmological time to the Milky Way. I then extend the ISM model to other infrared lines and older cosmological times in Chapter 3. Those cosmological times will take us back in time, close to the formation of the first galaxies. With that information, I present a web app that can be used to give estimates of the physical properties of galaxies based on observed line emissions (Chapter 4). Among these chapters, I compare our results with information gathered from observations in similar wavelength regimes.

In chapters 5 and 6 I focus on different types of galaxies and their electromagnetic spectrum shape. In Chapter 5, I use a tool to estimate the physical properties of the galaxies from the shape of the electromagnetic spectrum that covers the wavelength ranges between ultraviolet and far-infrared. These physical properties are analysed taking into account how intense are the radiative processes in the centre of the galaxies. I study galaxies that are interacting and in the future will merge into one. Finally, I analyse how tools that take the electromagnetic spectrum shape into account can

classify a specific type of galaxy to understand where they come from using publicly available data (Chapter 6).

With this thesis on "Diagnosing the ISM of galaxies and energetic processes in a cosmological context" I expect to give the reader an idea of how interesting some physical processes in galaxies are and how they can be used in the future to understand how the Universe works.

Resumen

El espacio entre las estrellas en el Universo no es solamente vacío, hay materia en él. Sin embargo, dicha materia es una fracción infinitesimal de la cantidad a la que estamos acostumbrados en la Tierra. Por ejemplo, la cantidad de materia por metro cúbico (también conocida como densidad) de aire en nuestra atmósfera en contraste con el espacio exterior es como comparar el cuerpo humano con un solo átomo. Es decir, las cantidades de materia son extremadamente diferentes y, por lo tanto, sus condiciones también lo son.

En las bajas densidades del espacio vacío se puede encontrar gas y polvo, una combinación que al hallarse entre las estrellas forma lo que denominamos el medio interestelar (o ISM por sus siglas en Inglés). Las temperaturas en este ISM están estrechamente relacionadas con sus densidades. En algunos casos, cuando las densidades aumentan diez veces, las temperaturas pueden descender casi mil veces. Estos cambios de temperatura son muy importantes para definir las condiciones del ISM, a través de procesos de calentamiento y enfriamiento. Dichos procesos están generalmente en equilibrio para que puedan coexistir diferentes regímenes del ISM. Sin embargo, se necesita "pagar" algo de energía para lograr mantener temperaturas similares. Esto es muy parecido a los procesos que ocurren en el cuerpo humano, donde el enfriamiento y el calentamiento se expresan en términos de sudoración (enfriamiento) o escalofríos (calentamiento) con el fin de mantener una temperatura corporal deseable.

En el ISM, estos procesos de calentamiento a menudo están relacionados con eventos energéticos. Pensemos, por ejemplo, en la radiación que las estrellas pueden emitir a través de fotones energéticos, o incluso en explosiones cósmicas más violentas cuando mueren las estrellas. Esos procesos excitarán (calentarán) algunos de los átomos o moléculas y, cuando estos átomos finalmente se desexcitan, emitirán fotones que transportarán energía más lejos y, por lo tanto, enfriarán el ISM. El proceso de enfriamiento más común es lo que llamamos enfriamiento de estructura fina, el cual está relacionado con cambios en los estados de energía dentro de los átomos que hacen que los fotones escapen. En el espectro electromagnético, donde la energía de los fotones se compara con una cierta longitud de onda o frecuencia, estas estructuras finas ocurren en regiones muy específicas a través de características de emisión o absorción, hecho que les ha dado el nombre de líneas espectrales.

Sin embargo, aunque estas líneas espectrales pueden ocurrir en diferentes rangos de longitud de onda del espectro electromagnético, desde rayos gamma hasta radio, la parte infrarroja del espectro electromagnético contiene las líneas de estructura fina más importantes para el enfriamiento por emisión. Allí, la línea de emisión más importante para el enfriamiento es producida por el átomo de carbono. Cuando este átomo se excita (ioniza) y luego se desexcita, actúa como un termostato para mantener las temperaturas en un nivel estable. Esta línea de emisión es importante entre las temperaturas cálidas ($\sim$1000 K) y frías ($\sim$10 K) en el ISM. Otra emisión

importante de líneas infrarrojas proviene del átomo de oxígeno; sin embargo, esta línea de emisión cubre un rango diferente de temperaturas, desde ∼10000 K hasta ∼1000 K. Condiciones físicas del ISM, como las descritas anteriormente, residirán en las diferentes líneas de emisión y, por lo tanto, compararlas es crucial para recuperar esas condiciones de las observaciones espaciales.

Gracias a la información que podemos obtener de las líneas de emisión sobre el ISM somos capaces de comprender los procesos que ocurren en escalas más grandes como en las galaxias. Por ejemplo, es posible observar estas líneas en la Vía Láctea y con ello recuperar sus propiedades físicas globales. Ahora bien, una región de interés dentro de las galaxias es su núcleo, donde los procesos violentos y las radiaciones intensas perturban el ISM. Dichas radiaciónes provienen del agujero negro supermasivo en el centro de las galaxias y la materia circundante que cae en él y se acumula en forma de disco. Las nubes de gas cercanas se ven afectadas por la rotación de este disco y hacen que cambien las señales que recibimos de las líneas de emisión. Estas líneas pueden ser anchas o estrechas dependiendo de la región de los núcleos que estemos mirando. Finalmente, una estructura de polvo más grande rodea a todas las descritas anteriormente en forma de "rosquilla", oscureciendo la radiación proveniente de estas regiones dependiendo del ángulo de visión.

Las observaciones en el espectro electromagnético infrarrojo pueden ayudar a caracterizar esas estructuras y los procesos de gas del ISM en las galaxias. Sin embargo, las cantidades de datos de esas observaciones son actualmente limitadas y no están fácilmente disponibles. Además, estamos observando imágenes cuasiestáticas del Universo que limitan nuestro análisis de la evolución de tales procesos. Por lo tanto, es necesario utilizar herramientas computacionales que funcionen como laboratorios para ejecutar nuestros experimentos y que consideren todo el conocimiento físico que conocemos actualmente sobre los procesos del Universo. Una vez que tengan eso, pueden usarse para reproducir el espectro observado de galaxias y predecir lo que se esperaría observar en el futuro.

Esta tesis

El objetivo principal de esta tesis doctoral es comprender los procesos del gas en las galaxias a lo largo del tiempo cósmico, centrándose en los parámetros físicos que describen la formación estelar. Por lo tanto, en el Capitulo 1 presento los antecedentes para comprender el contenido de esta tesis. Allí explico en profundidad los procesos físicos dentro del ISM, algunas características importantes de las galaxias y las herramientas que utilizo para el análisis.

En los capítulos 2–4 me centro en el ISM de las galaxias. En el Capítulo 2 describo cómo usar simulaciones para calcular la línea de emisión del átomo de carbono en el infrarrojo. Para ello, implemento un modelo ISM motivado físicamente que se aplica en galaxias simuladas similares en tiempo cosmológico a la Vía Láctea. Luego extiendo el modelo ISM a otras líneas infrarrojas y tiempos cosmológicos más antiguos en el Capítulo 3. Esos tiempos cosmológicos nos llevarán atrás en el tiempo, cercano a la formación de las primeras galaxias. Con esa información presento una aplicación web que se puede usar para obtener estimaciones de las propiedades físicas de las galaxias

basadas en las emisiones lineales observadas (Capítulo 4). Durante estos capítulos comparo los resultados con información recopilada de observaciones en regímenes de longitud de onda similares.

En los capítulos 5 y 6 me centro en diferentes tipos de galaxias y la forma de su espectro electromagnético. En el Capítulo 5, utilizo una herramienta para estimar las propiedades físicas de las galaxias a partir de la forma del espectro electromagnético que cubre los rangos de longitud de onda entre el ultravioleta y el infrarrojo lejano. Estas propiedades físicas se analizan teniendo en cuenta la intensidad de los procesos radiativos en el centro de las galaxias. En este trabajo, estudio galaxias que están interactuando y en el futuro se fusionarán en una sola. Finalmente, analizo cómo las herramientas que tienen en cuenta la forma del espectro electromagnético pueden clasificar un tipo específico de galaxia para comprender de dónde provienen utilizando datos disponibles públicamente (Capítulo 6).

Con esta tesis sobre "Diagnóstico del ISM de galaxias y procesos energéticos en un contexto cosmológico" espero darle al lector una idea de lo interesantes que son algunos procesos físicos en las galaxias y cómo pueden ser usados en el futuro para entender cómo funciona el Universo.

Acknowledgments

The idea of becoming a PhD in astronomy, and being able to write this thesis, began 12 years ago. Thanks to the support of many people, I was able to present this thesis. When I initially approached scientific research, I was very happy to be able to say that astronomy looks like a good subject to work on, and now I know it's true. Since then, I have learned a lot from professors, colleagues, friends and family. It is to my family that I have dedicated this thesis, they have been my support at all times. Without them, I would not have finished my undergraduate studies and started my master's studies. Thanks to them I also had the opportunity to travel to have my first international experiences, which led me to live in the Netherlands. At the end of my master's degree, fortunately, I found a very good opportunity to develop a doctoral project in Groningen. After these four years in Groningen, I can only express my gratitude to all those who in one way or another help me achieve this personal goal.

My biggest support during this journey was, without a doubt, Jennifer. Thanks to you I had the tranquillity and calm that I needed to not despair. I had the most joyful moments with you whenever I needed them. Thank you also for having discussions about how the world works, how to learn a new language, experiment with food and so many things that make me happy. Hopefully, we can have more of those moments and we can take advantage of everything that life gives us.

I am also very grateful to Lingyu for being such an amazing supervisor. I am so glad I have had such a patient guide so close to me. I always liked that you tried to educate me on the important details of science and language. Half of this doctoral thesis is thanks to everything I was able to learn from you, you are the best supervisor I could have. Scott and Floris, thank you for always being so supportive during this process. I am very grateful for the good treatment you had with me and how you taught me to always be a better scientist.

To Rafael and Armando, thank you very much for opening the doors for me to investigate astronomy. Without you, none of the steps I would have taken would have paid off. To Howard and Matt for always keeping me in mind and being able to talk about research topics. To Bernhard and Jarle for allowing me to develop my master's projects with you and for teaching me so many things that were useful for my PhD.

Finally, I also want to thank all the other researchers who shared the data with us and made this thesis possible. To the secretaries of SRON and Kapteyn, firstly sorry for so many questions and secondly thank you very much for your kindness whenever I had to ask for help. I also want to thank all the support staff within the institutes and the university with whom I came in contact. To my colleagues in the office, thank you for helping me during the first months of the doctorate and allowing me to focus so much on the pleasant atmosphere that we always had. To all the members of the interstellar medium group, thank you for having learned so many things from

each other, about the ISM and science in general. Thanks to all my PhD mates with whom I enjoy the time inside and outside of the institute. Thank you very much to everyone.

On a more personal note, I think that giving a long list of names misses the point of personally thanking the people indirectly involved in the development of this thesis. In addition, it is possible to hurt sensitivities with some comments. I decided to personally thank them for particular details and situations. To my closest friends for strengthening my spirit and supporting me. Thank you very much to all ...