UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE FÍSICA

Who is who? FADO and galaxy classification

Duarte Muñoz Santos

Mestrado em Física
Especialização em Astrofísica e Cosmologia

Dissertação orientada por:
Dr. Cirino Pappalardo

2022

Agradecimentos

Muitas histórias de pessoas que entram em Astrofísica começam com o fascínio pelas estrelas, que quando olharam pela primeira vez para o céu e viram esses inúmeros pontos luminosos, ficaram imediatamente deslumbrados. Não posso afirmar que a minha história começa de forma diferente, mas agora que terminei o Mestrado, planeio escrevê-la da minha maneira.

Aos meus pais, Pedro e Carmen, agradeço profundamente por todas as oportunidades que me deram. Desde os inúmeros livros e Atlas do Céu que me compraram enquanto criança, à enorme paciência de me ouvirem falar sobre assuntos que nem sabem onde começar por compreender, um enorme obrigado. À minha irmã, Diana, agradeço por ser a pessoa a quem eu me dirigia para me pôr a cabeça no tronco e nos membros. Obrigado pelos teus inúmeros conselhos sobre a faculdade e a vida. Espero que aguentes com mais algumas perguntas da minha parte.

Sem ter entrado nesta Faculdade, decerto que não seria a pessoa que sou hoje. Muitas pessoas deram-me a capacidade e os connhecimentos para progredir, e gostaria de agradecer em especial a algumas delas (mas reforço que gosto muito de todos vós).

Ao Fábio, obrigado por me ouvires queixar sobre o que quer que fosse, sobre as discussões interessantes sobre uma determinada série de videojogos e, nomeadamente, por estares sempre presente. Não estaria aqui sem a tua inspiração e persistência (a bem e a mal). À Joana, agradeço vivamente não só a tua paciência, mas também a tua amizade. Sem a tua presença creio que nunca saíria da cepa torta e, por isso, devo-te um obrigado de dimensões cósmicas. Não sei quem seria sem ti aqui. Ao Miguel P., obrigado por seres o meu irmão de Mestrado (e adiante), pelas conversas noturnas que se tornaram madrugais e, acima de tudo, pelo apoio perene e incessante. Os semestres sem a tua presença sem dúvida que seriam mais sombrios. À Inês, desde o Renascimento que não podia ter escolhido pessoa melhor para falar durante horas a fio. Obrigado pelo teu pragmatismo e capacidade de abater os meus grandes stresses, pelo dinamismo constante e pelas grandes ideias que surgem. Ao Miguel C., agradeço acima de tudo o exemplo de resiliência que se deve ter. Além de dares uma desculpa para ires a Cascais, a tua presença é, para mim, símbolo de evolução e crescimento pessoal. Não me aventurava tanto sem ti. Ao Afonso, agradeço os nossos planos falhados, as nossas comunicações mensais e as aventuras que me proporcionaste ao longo destes cinco anos. De mesas a Alfarim, espero que, mesmo que escassamente, nunca paremos de falar e ter coisas para dizer. Agradeço também à Ganna, à Catarina, à Vanessa, ao Bruno, ao João Rodrigo, ao João Augusto, ao Filipe e ao Rafael, por terem sido meus companheiros nesta viagem.

Não posso terminar esta secção sem deixar de agradecer àqueles que me inspiraram a entrar no mundo da investigação científica.

Aos meus caros mosqueteiros: Maria, Lara, Constança e Duarte, por me motivarem ao longo destes dois anos do Mestrado a nunca desistir e a seguir o que realmente gosto. Sem vocês de certeza que qualquer viagem feita aos Açores seria exponencialmente mais entediante. Ao grupo de Galáxias do OAL: José Afonso, Israel Matute, Rodrigo Carvajal e Davi Barbosa, pelas nossas interessantes (e longas)

reuniões semanais onde se fala sobre tudo e mais alguma coisa. Ao Henrique, obrigado pelas discussões e apoio em todas as questões que tivesse sobre galáxias e outras temáticas.

Finalmente, mas não com significado menor, devo também um agradecimento especial ao meu orientador, Cirino Pappalardo. Por me ter aceite como orientado, por ter tido a paciência de ter reuniões no bar do C8, pelas conversas estimulantes na Tapada da Ajuda e pelo apoio constante ao longo deste ano, esta tese não teria o formato nem o nexo que tem sem a sua orientação.

Resumo

A introdução dos diagramas BPT, assim chamados devido aos seus criadores (Baldwin, Phillips, & Terlevich, 1981), mudou a maneira como classificamos galáxias. Em vez de se basearem em componentes específicas dos espetros das galáxias, os autores decidiram basear-se nos mecanismos de excitação do gás presentes nestas, através do quociente do fluxo de linhas de emissão específicas. Após investigação destes diagramas, descobriu-se que os melhores a separar as galáxias em dois tipos (galáxias que formam estrelas, SF, e galáxias com um núcleo galático ativo, AGN) eram: $[\mathrm{NII}]\lambda 6584/\mathrm{H}\alpha$ vs. $[\mathrm{OIII}]\lambda 5007/\mathrm{H}\beta$, $[\mathrm{SII}]\lambda\lambda 6717,6731/\mathrm{H}\alpha$ vs. $[\mathrm{OIII}]\lambda 5007/\mathrm{H}\beta$, e $[\mathrm{OI}]\lambda 6300/\mathrm{H}\alpha$ vs. $[\mathrm{OIII}]\lambda 5007/\mathrm{H}\beta$, que denominamos por diagrams BPT, SII e OI, respetivamente (Veilleux & Osterbrock, 1987). Vinte anos depois, Kauffmann et al. (2003), Kewley et al. (2001) e Kewley et al. (2006) descobriram que o diagrama BPT tinha uma sensibilidade única à metalicidade das galáxias, o que permitiu que fosse criado uma terceira categoria: as *Composite*, que são galáxias que formam estrelas e têm um AGN presente no seu centro.

Para classificar estas galáxias, necessitamos de extrair os fluxos das linhas de emissão. Assim, precisamos de modular os espetros das galáxias, que pode ser concretizado através de modelos de síntese de população estelar. Estes consistem em dois passos: a criação de populações estelares simples, através de isócronas, a função de massa inicial e bibliotecas de espetros de estrelas; e a criação da população estelar composta, que junta as populações estelares simples com a evolução da taxa de formação de estrelas e da metalicidade, tal com a atenuação e emissão da poeira (Conroy, 2013). Estes modelos, geralmente, não incluem a contribuição nebular de uma forma auto-consistente, optando por inlcuí-la à parte. Isto pode provocar problemas quando ajustarmos os espetros observados em modelos puramente estelares, nomeadamente em galáxias que estão a sofrer episódios violentos de formação estelar, onde a emissão nebular contribui tanto ou mais que a emissão estelar (Amorín et al., 2012; Fernández et al., 2022).

Para resolver este problema, FADO (*Fitting Analysis using Differential evolution Optimization*, Gomes and Papaderos 2017) foi criado no Instituto de Astrofísica e Ciências do Espaço com o intuito de ser o primeiro código a ajustar de uma forma auto-consistente a contribuição nebular nos espetros das galáxias. Assim, o objetivo desta dissertação é comparar o FADO com um modelo puramente estelar, para perceber se esta mudança nos modelos de síntese de população estelar altera de alguma forma a classificação de galáxias através dos diagramas BPT.

Para tal, decidimos comparar a análise do FADO (Cardoso et al., 2022) com a análise do MPA-JHU das galáxias do SDSS DR-7 (Abazajian et al., 2009). Verificámos que a mediana do *redshift* desta amostra (916 118 galáxias) é aproximadamente 0.116. Após a extração dos fluxos necessários, vimos que as diferenças são maiores nas linhas de emissão mais azuis (55% e 58% para $\mathrm{H}\beta$ e $[\mathrm{OIII}]\lambda 5007$, respetivamente) do que para linhas de emissão mais vermelhas (19% e 22% para $\mathrm{H}\alpha$ e $[\mathrm{NII}]\lambda 6584$ respetivamente). Isto acontece pois o FADO e o MPA-JHU usam differentes leis de extinção quando ajustam os espetros das galáxias (Calzetti et al. 2000 e Charlot and Fall 2000, respetivamente). As larguras equivalentes mudam estas diferenças, aumentando-as para as linhas de emissão de Balmer e

diminuindo-as para as linhas de transições proíbidas (29%, 75%, 38% e 9% para Hα, Hβ, [OIII]λ5007 e [NII]λ6584, respetivamente). Isto deve-se ao facto do FADO medir as linhas de emissão através do traço de absorção, algo que as transições proibidas não possuem e que MPA-JHU não faz, aumentado as diferenças entre *datasets*.

Após selecionarmos as galáxias com S/N$>$ 3 e fluxos entre $10^{-18} < F_\lambda\,[\mathrm{erg\ s^{-1}cm^{-2}}] < 10^{-12}$, no FADO e MPA-JHU, para as linhas de emissão do diagrama BPT, corrigimos o fluxo extraído devido à extinção na linha de visão que existe entre nós e as galáxias. Com isto, ficámos com uma amostra de 156 748 galáxias em comum nos dois *datasets*, onde 2% não tem a mesma classificação entre FADO e MPA-JHU. Vimos que estas galáxias estão na fronteira entre classificações, então decidimos descartá-las da nossa análise, visto que as incertezas compensam pela sua posição. Assim, comparámos as restantes galáxias, e vimos que as diferenças têm um máximo de 10% entre *datasets*, onde as incertezas (que variam entre 10% e 45%) compensam por este valor. Neste caso, vimos que a adição da emissão nebular auto-consistente não afeta significativamente a classificação de galáxias.

Para perceber se a tendência das galáxias serem classificadas igualmente entre o FADO e MPA-JHU é única ao diagrama BPT, comparámos também o diagrama SII e OI nestes dois *datasets*. Após uma seleção idêntica à anterior no S/N e no fluxo nas linhas de emissão destes três diagrams (exceto no [OI]λ6300, devido à sua fraqueza intrínseca), obtivemos uma amostra de 143 177 galáxias. Desta amostra, vimos que os resultados do FADO tinham 15 840 galáxias Ambíguas. Nestas galáxias, vimos que as Ambíguas que são consideradas SF no diagrama BPT são 223% menos massivas e possuem 95% menos metalicidade que a amostra completa de galáxias SF. Ademais, vimos que da amostra total de galáxias SF, as galáxias Ambíguas representam 26% da amostra com metalicidade inferior à do Sol e massas inferiores a $10^9\,M_\odot$, enquanto que apenas 8% do oposto. Para perceber a origem destas galáxias, comparámos com dados do AllWISE, através dum diagrama de classificação que compara as bandas de observação (Bond et al., 2012; Cutri et al., 2021). Os dados do WISE e do diagrama BPT concordam um com o outro, o que significa que esta tendência de classificar galáxias com metalicidades inferiores à do Sol e massas estelares baixas é algo que se propaga nestas medições. Isto deve-se, muito provavelmente, ao facto de estas galáxias terem um AGN fraco no seu núcleo. No diagrama SII e especialmente no diagrama OI (devido à sua baixa sensibilidade em relação à metalicidade), estas galáxias já são classificadas como AGNs, dando-nos uma melhor compreensão da sua natureza. Contudo, isto exige mais investigação para perceber se realmente é um AGN que modifica a posição destas galáxias nestes diagramas, ou se é algo diferente presente nelas. Também verificámos que a existência de galáxias Ambíguas *Composites* e AGNs são devido à sensibilidade da metalicidade de cada diagrama.

Além destas galáxias, também comparámos as galáxias não-Ambíguas entre o FADO e MPA-JHU, e vimos que as diferenças nos diagramas BPT e SII são baixas (máximo de 4%) e maiores no diagrama OI (máximo de 22%), devido ao facto de não termos limitado o S/N da linha de emissão de [OI]λ6300. Contudo, as incertezas em cada diagrama compensam por estas diferenças, o que nos leva a concluir o mesmo que anteriormente: a adição auto-consistente da contribuição nebular não afeta significativamente a classificação de galáxias.

Na nossa investigação, reparámos que se comparármos a largura equivalente de Hα e Hβ com o quociente das larguras equivalentes de Hβ e [OIII]λ5007, as galáxias dividem-se naturalmente em dois tipos, de acordo com o diagrama BPT: SF e *Composites* numa nuvem, e AGNs noutra. Tentámos comparar apenas os fluxos destas linhas de emissão, mas vimos que esta propriedade só se manifesta nas larguras equivalentes, devido à sua independência do contínuo dos espetros das galáxias. Assim, criámos três funções de separação para três diagrams: Hα vs [OIII]λ5007/Hβ, Hβ vs [OIII]λ5007/Hβ e (Hα + [NII]λ6584) vs [OIII]λ5007/Hβ, ou diagramas EWA, EWB e EWAII, respetivamente. Estas

funções minimizam o número de galáxias contaminadas. Vimos que o diagrama EWB tem o menor número de galáxias contaminadas, então decidimos comparar este diagrama com os diagramas BPT, SII e OI. Comparando galáxias contaminadas e Ambíguas, vimos que há uma correspondência de 84% para galáxias SF, 92% para galáxias *Composite* e 43% para galáxias AGNs. O número baixo para AGNs é devido à forma como definimos as galáxias Ambíguas AGNs. Estes diagramas são úteis no caso de não termos todas as linhas de emissão para gerar o diagrama BPT; para ter noção da natureza das galáxias Ambíguas, devido ao alto nível de correspondência entre contaminadas e Ambíguas; e pode ser útil no caso de termos espetros com baixa resolução.

Finalmente, decidimos comparar a sequência principal de galáxias entre o FADO e MPA-JHU, que é a comparação entre a taxa de formação de estrelas (SFR) e a massa estelar das galáxias. Utilizámos o fluxo da linha de emissão do Hα para calcular a SFR das galáxias SF, mas para as galáxias *Composite* e AGN decidimos utilizar a SFR do MPA-JHU devido às suas considerações para a idade, que são importantes. Para a massa estelar, usámos a de cada *dataset*. Vimos que, apesar das diferenças serem grandes (nomeadamente na massa estelar, com diferenças máximas de 42%), as incertezas compensam estes valores (máximo de 58%). Novamente, temos que a adição auto-consistente da contribuição nebular não afeta, de forma significativa, a posição das galáxias na sequência principal. Também comparámos a posição das galáxias Ambíguas na sequência principal e vimos que estão aproximadamente nos locais onde "devem" estar, o que significa que a natureza destas galáxias não consegue ser detetada com este esquema.

Palavras chave: códigos de síntese espetral, classificação de galáxias, evolução de galáxias, AGNs

Abstract

From the spectra of galaxies, we are able to extract many physical properties: star-formation rate, stellar mass and metallicity, for example. To calculate them, we must first classify galaxies according to their main mechanism of emission: star-formation (SF) or Active Galactive Nuclei (AGNs). The BPT diagrams, named after their creators, are a classification method that does exactly this, through emission line flux ratios.

To extract the emission line flux of these galaxies, one uses models such as stellar population synthesis. Most of these models do not add the nebular contribution, only considering the stellar component. FADO was developed to tackle this issue in SED modelling.

The purpose of this work is to compare the analysis of the Sloan Digital Sky Survey Data Release 7 galaxies (with a median $z \approx 0.116$) of MPA-JHU and FADO, to understand if the nebular contribution affects galaxy classification through the standardised BPT diagrams.

We have found that the self-consistent addition of nebular emission does not significantly affect galaxy classification through emission line ratio diagrams. The same is true of the galaxies in the Main Sequence, where the galaxies in FADO and MPA-JHU fall in the same regions of the parameter space.

Furthermore, we found that the BPT diagram may possess a bias towards low stellar and sub-solar metallicity galaxies, as they are more likely to be misclassified by other emission line ratio diagrams.

Finally, we have also found that comparing the equivalent width of Balmer lines (Hα or Hβ) with the ratio of the equivalent widths of [OIII]λ5007 and Hβ gives us a new classification diagram that separates galaxy in two types: SF and AGN. This simple classification could be used in future large spectroscopic surveys, such as MOONRISE, where at high redshifts some important emission lines for the BPT diagram are unable to be detected.

Keywords: spectral synthesis codes, galaxy classification, galaxy evolution, AGNs

Table of Contents

List of Figures

List of Tables

List of Abbreviations and Symbols

Abbreviations:
AGN: Active galactive Nucleus
BPT diagram: Diagram comparing the ratio of $[\mathrm{NII}]\lambda 6584/\mathrm{H}\alpha$ with the ratio of $[\mathrm{OIII}]\lambda 5007/\mathrm{H}\beta$
CSP: Composite Stellar Population
EW: Equivalent Width
FADO: Fitting Analysis using Differential evolution Optimization
IMF: Initial Mass Function
LINER: Low Ionisation Narrow Emission-line Region
MPA-JHU: Max-Planck institute for Astrophysics - John Hopkins University
OI diagram: Diagram comparing the ratio of $[\mathrm{OI}]\lambda 6300/\mathrm{H}\alpha$ with the ratio of $[\mathrm{OIII}]\lambda 5007/\mathrm{H}\beta$
S/N: Signal-to-Noise ratio
SDSS: Sloan Digital Sky Survey
SDSS-DR7: Seventh Data Release of the Sloan Digital Sky Survey
SED: Spectral Energy Distribution
SF: Star-forming/Star formation
SFH: Star-Formation History
SFR: Star-Formation Rate
sSFR: Specific Star-Formation Rate
SPS: Stellar Population Synthesis
SSP: Simple Stellar Population
SII diagram: Diagram comparing the ratio of $[\mathrm{SII}]\lambda\lambda 6717,6731/\mathrm{H}\alpha$ with the ratio of $[\mathrm{OIII}]\lambda 5007/\mathrm{H}\beta$
WISE: Wide-field Infrared Survey Explorer
Symbols:
z: Redshift
$M_{\odot}$: Solar Mass
Z: Metallicity
σ_{EW}: Uncertaintiy associated with the Equivalent Width

Chapter 1

Introduction

Galaxies are complex and dynamic systems, made up of many moving parts, such as stars, gas and dust. These components emit in a wide range of wavelengths in the electromagnetic spectrum, with different strengths. If we decompose the light of a galaxy, its spectrum, or its spectral energy distribution (SED), would be a net sum of the contribution of each of these elements, be it emission or absorption. From these SEDs, we can extract relevant physical properties such as the star-formation rate (SFR), the stellar mass (usually measured in solar masses, $M_{\odot}$) and the metallicity (Z, defined in Astronomy as the abundance of elements present in these systems other than hydrogen and helium).

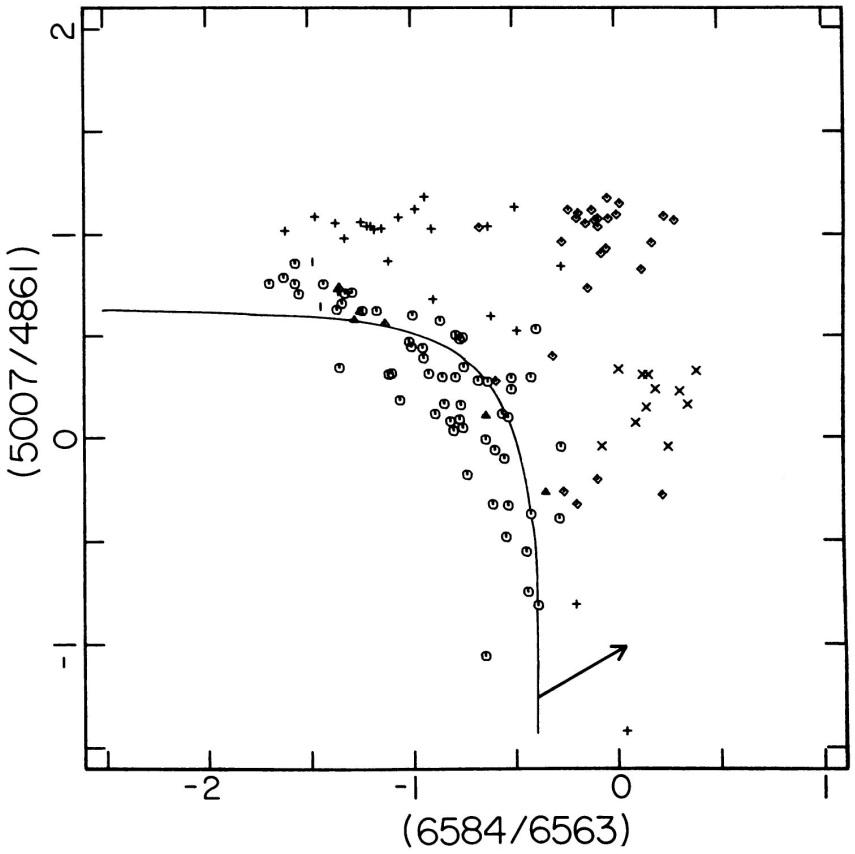

Figure 1.1: Comparison between the emission line flux ratios of [NII]λ6584/Hα and [OIII]λ5007/Hβ. The circles represent normal HII regions, the triangles detached HII regions, the plus signs planetary nebulae, the diamonds objects photoionised by a power law, and the crosses shock-heated galaxies. The line represented is a first attempt at separating galaxies from their different excitation mechanisms. The arrow represents how the galaxies move from one region of the diagram into another. Taken from Figure 5 of Baldwin, Phillips, and Terlevich (1981).

1.1 Galaxy Classification

Before we can extract the physical properties of a galaxy, we need to know its general internal behaviour: is it a galaxy that is currently forming stars (SF)? Is it a galaxy with a peculiar activity in its core? Is it something else entirely? Answering these questions is fundamental as, through them, we can model galaxies according to their nature. Trying to do so with wrong assumptions can lead to incorrect estimations.

Before the 1980's, galaxies were classified according to specific characteristics in their spectra, leading to a wide range of classifications to objects, such as quasi-stellar objects, Seyferts, active galaxies, and many others (e.g. Chiu 1964; Marshall et al. 1978; Seyfert 1943). This was confusing to scientists at the time, since classifying galaxies proved to be a challenge. As such, Baldwin, Phillips, and Terlevich (1981), realising this was a problem, decided to classify galaxies not through these specific characteristics, but rather through their excitation mechanisms. They divided these into four categories:

1. Photoionisiation by O and B-type stars;
2. Photoionisation by a power-law continuum-source;
3. Shock wave heating;
4. Planetary nebulae.

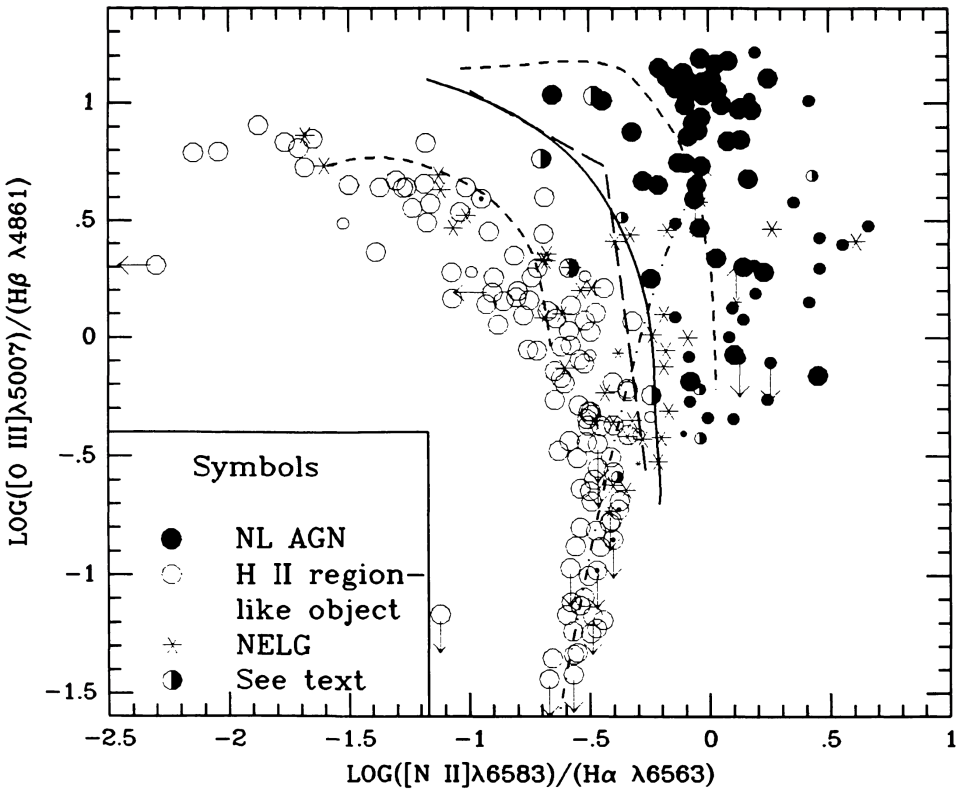

Figure 1.2: Comparison between the emission line flux ratios of [NII]λ6584/Hα and [OIII]λ5007/Hβ. Besides what is shown in the figure, the most relevant information is that the half-filled circles are peculiar objects and the solid line is the empirical separation between SF and AGN galaxies. Figure taken from Figure 4 of Veilleux and Osterbrock (1987).

They did this because they claimed that the excitation mechanisms were what truly separated one type of galaxy from another. The authors proved this claim, finding that they could compare these different

excitation mechanisms by using specific emission lines present in the SEDs of galaxies. These emission lines were considered as proxies for these physical processes: for example, Hα is a good indicator of the photoionisation by O and B-type stars. Emission line flux ratios such as [OIII]λ5007/Hβ are also a useful example to distinguish the ionised regions in galaxies, as the range of values of these ratios gives a spread of the heavy element abundances in different objects, i.e., it gives us a distribution of the density and metallicity of a galaxy. However, they were unable to find that just one emission line flux ratio could distinguish galaxies clearly, so they opted to compare at least two emission line flux ratios to better distinguish the object types.

They found that many emission line ratios could be used to separate galaxies into four distinct regions, one for each excitation mechanism, and noted that one of the best diagrams at distinguishing galaxies was the comparison between the emission line flux ratios of [NII]λ6584/Hα and [OIII]λ5007/Hβ (Figure 1.1). This was because the combination of these emission line flux ratios was particularly insensitive to reddening - the absorption and scattering of photons with shorter wavelengths by dust grains present in the interstellar medium, which makes objects appear redder, and therefore cooler, than they actually are. Other examples of emission line flux ratio diagrams that they considered to be good at distinguishing the principal excitation mechanisms include [OII]λ3727/[OIII]λ5007 vs [OIII]λ5007/Hβ and [OII]λ3727/[OIII]λ5007 vs. [OI]λ6300/Hα. These classfication schemes were then named the BPT diagrams, in honour of their creators.

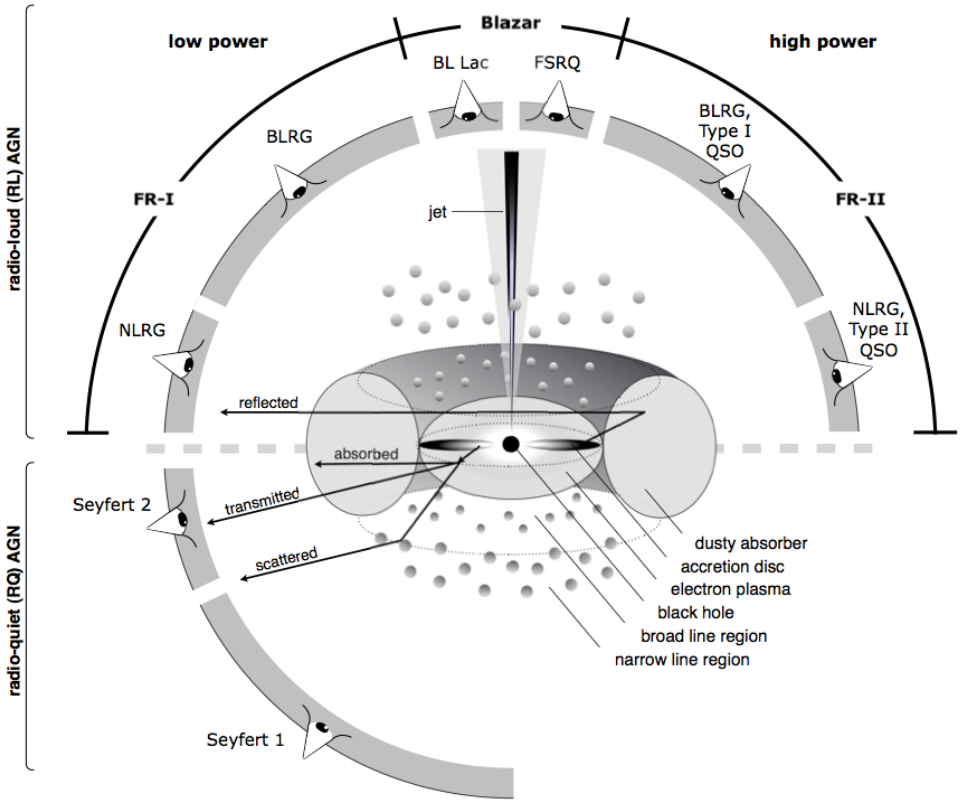

Figure 1.3: Schematics of the unified AGN model. Taken from Figure 4.16 of Beckmann and Shrader (2012).

Due to the fact that [OII]λ3727 is a weak emission line compared to others such as Hα or Hβ, latter research by Veilleux and Osterbrock (1987) opted to classify galaxies only through strong emission lines, so more data could be compared. They found that the [NII]λ6584/Hα vs. [OIII]λ5007/Hβ, [SII]$\lambda\lambda$6717,6731/Hα vs. [OIII]λ5007/Hβ, and [OI]λ6300/Hα vs. [OIII]λ5007/Hβ diagrams,

henceforth to be known as the BPT, SII and OI diagrams, were the best between distinguishing SF galaxies and galaxies with a unique element present in their cores: an Active Galactive Nucleus (AGN). The authors also clarified that the best diagram to separate galaxy types was the OI diagram, but, due to the weakness of the [OI]λ6300 emission line, very few objects are able to be put in the diagram, unlike the BPT and SII diagrams. They also attempted to separate SF and AGN galaxies using empirical calculations (i.e. Figure 1.2).

AGNs, in their unified theory (Beckmann & Shrader, 2012), are objects at the centre of galaxies, near an accreting supermassive black hole. They are composed of three main components: a torus of dust, an accretion disk and radio jets (Figure 1.3). The way these objects are tilted according to our line-of-sight is what gives them different properties in their SED.

Nearly twenty years later, technology greatly evolved, and many more extragalactic objects were able to be detected by efforts such as the Sloan Digital Sky Survey, SDSS (most recently, Abdurro'uf et al. 2022; Blanton et al. 2017), which observed (and is still observing) more than a million objects present at many epochs, but mostly in the Local Universe (with redshifts in the order of 0.1). With all these galaxies observed, many more emission line fluxes were able to be detected and, as such, the emission line flux ratio diagrams were further studied.

After selecting the BPT, SII and OI diagrams to analyse (as Veilleux and Osterbrock (1987) showed were the best diagrams at separating galaxies), Kewley et al. (2001) calculated a "maximum starburst" line that separates galaxies where emission is dominated by star-forming regions or by an AGN (Figure 1.4). They achieved this by using stellar population synthesis (SPS) models, namely PEGASEv2.0 (Fioc & Rocca-Volmerange, 1997) and STARBUST99 (Leitherer et al., 1999), to replicate the SED of a galaxy, as well a photoionisation model, MAPPINGS III (Groves, Dopita, & Sutherland, 2004), to replicate the nebular component of a galaxy. This separation line essentially tells the maximum position where a SF galaxy can be placed, and above this line there are only AGNs, due to the different ionisation sources.

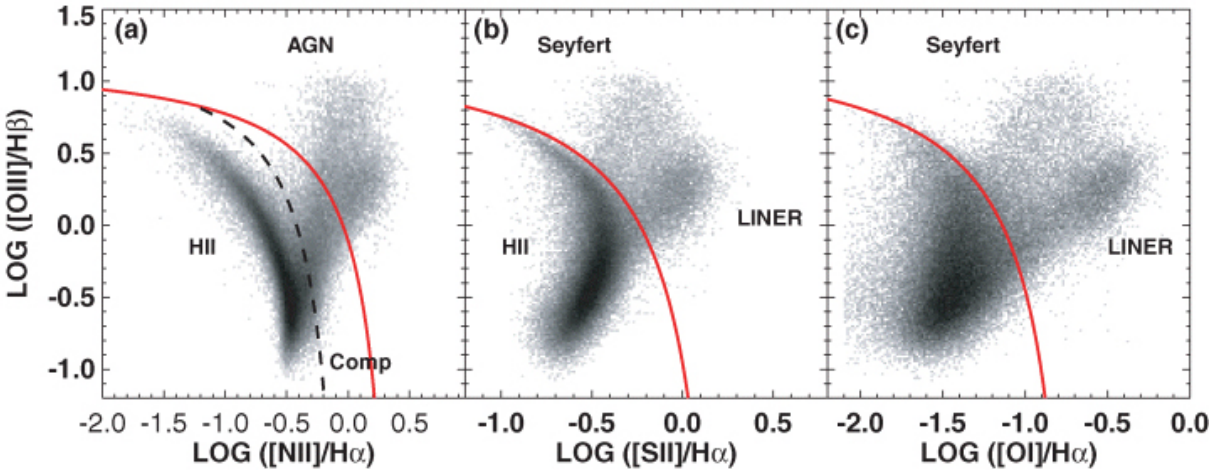

Figure 1.4: [NII]λ6584/Hα vs. [OIII]λ5007/Hβ (left), [SII]$\lambda\lambda$6717,6731/Hα vs. [OIII]λ5007/Hβ (middle) and [OI]λ6300/Hα vs. [OIII]λ5007/Hβ (right) diagrams, with the SF-AGN separation line from Kewley et al. (2001) (full red line) and the SF-Composite separation line from Kauffmann et al. (2003) (dashed black line). Taken from Figure 1 of Kewley et al. (2006).

Additionally, Kauffmann et al. (2003) found that the BPT diagram had a particular sensitivity to SF galaxies that simultaneously had an AGN contribution. This helped clarify galaxy classification by adding a second category besides SF and AGN: a Composite category, which is a blend of both. This sensitivity comes from the fact that the [NII]λ6584/Hα emission line flux ratio saturates at high metallicities, due to nebular contributions, causing the SF sequence to be nearly vertical at

$\log([\mathrm{NII}]\lambda 6584/\mathrm{H}\alpha) \sim -0.5$, as can be seen by Figure 1.4. This means that, for this ratio to be pushed to higher values, there must be another mechanism inside the galaxy that drives galaxies to skew to the right in this BPT diagram. This can be explained through AGN contributions in the SED.

Further analysis from Kewley et al. (2006), showed that the AGN regions of the emission line ratio diagrams were divided in two clusters, one with higher ionisation values than the other. Galaxies in the branch with higher ionisation, in the upper parts of the AGN region, were subclassified as Seyfert-type galaxies and the galaxies in the branch with lower ionisation values were subclassified as LINER (Low Ionisation Narrow Emission-line Region) type galaxies (Figure 1.5). This distinction can be explained by the generalised AGN model (Figure 1.3). The LINER classification, originally coined by Heckman (1980), however, has been suffering a reanalysis due to its ambiguous nature, through programs like the MaNGA (Mapping Nearby Galaxies at Apache Point Observatory) survey, because it is based on integral field unit data that can resolve the inner structure of local galaxies (Belfiore et al., 2016).

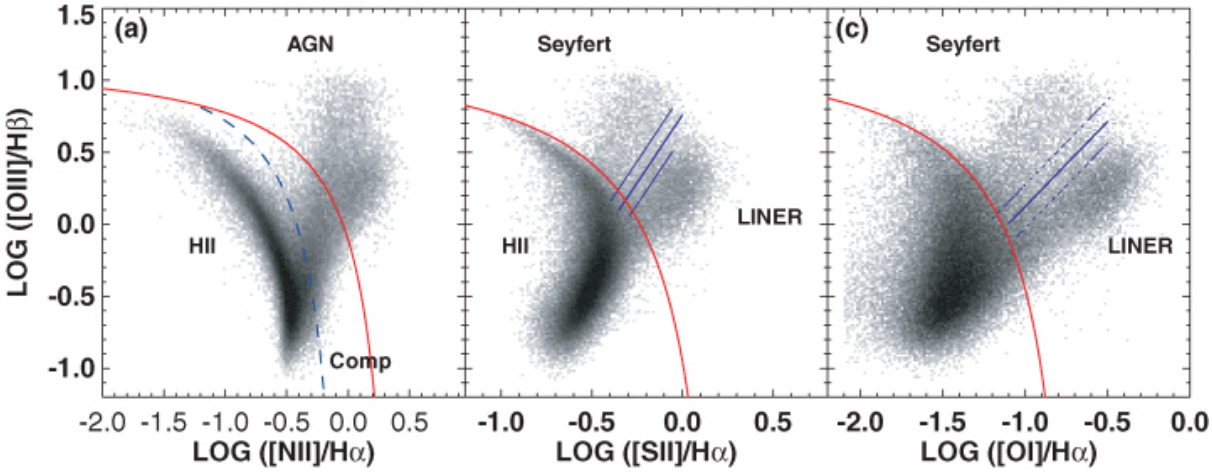

Figure 1.5: $[\mathrm{NII}]\lambda 6584/\mathrm{H}\alpha$ vs. $[\mathrm{OIII}]\lambda 5007/\mathrm{H}\beta$ (left), $[\mathrm{SII}]\lambda\lambda 6717,6731/\mathrm{H}\alpha$ vs. $[\mathrm{OIII}]\lambda 5007/\mathrm{H}\beta$ (middle) and $[\mathrm{OI}]\lambda 6300/\mathrm{H}\alpha$ vs. $[\mathrm{OIII}]\lambda 5007/\mathrm{H}\beta$ (right) BPT diagrams, with the SF-AGN separation line from Kewley et al. (2001) (full red lines), the SF-Composite separation line from Kauffmann et al. (2003) (dashed blue line) and the Seyfert-LINER separation line from Kewley et al. (2006) (solid blue lines). Taken from Figure 4 of Kewley et al. (2006).

1.2 Stellar Population Synthesis

To extract the emission line flux from these galaxies, as well as other relevant properties, we fit models onto the observed spectra of galaxies. Modelling these SEDs is of extreme importance as, through these models, we can garner a look at the history of these galaxies, how they might have evolved and still might evolve. One way to model them is through stellar population synthesis (SPS) codes, summarised in Figure 1.6.

Essentially, SPS codes have two main steps: the construction of simple stellar populations, SSPs, and the construction of the composite stellar population, CSP (Conroy, 2013).

A simple stellar population is described as the time evolution of a single stellar population where all its components have the same age, metallicity and abundance patterns (or the amount of chemical elements present in the stellar population other than hydrogen or helium). They are generally represented by the equation below:

$$f_{\mathrm{SSP}}(t,Z) = \int_{m_{\mathrm{low}}}^{m_{\mathrm{up}}(t)} f_{\mathrm{star}}[T_{\mathrm{eff}}(M), \log g(M)|t,Z]\,\Phi(M)\,dM \tag{1.1}$$

where $f_{\rm star}$ is the spectrum of a star with a specific effective temperature, $T_{\rm eff}$, surface gravity, $\log g$, initial (or zero age main sequence) mass, M, time, t, and metallicity, Z; $\Phi(M)$ represents the initial mass function (IMF) and $f_{\rm SSP}$ represents the spectrum of the SSP, which is both time and metallicity dependent. The integral limits are usually between the hydrogen burning limit (0.1 solar masses, or 0.1 $M_{\odot}$) and the upper limit is dictated by stellar evolution (hence the time dependency).

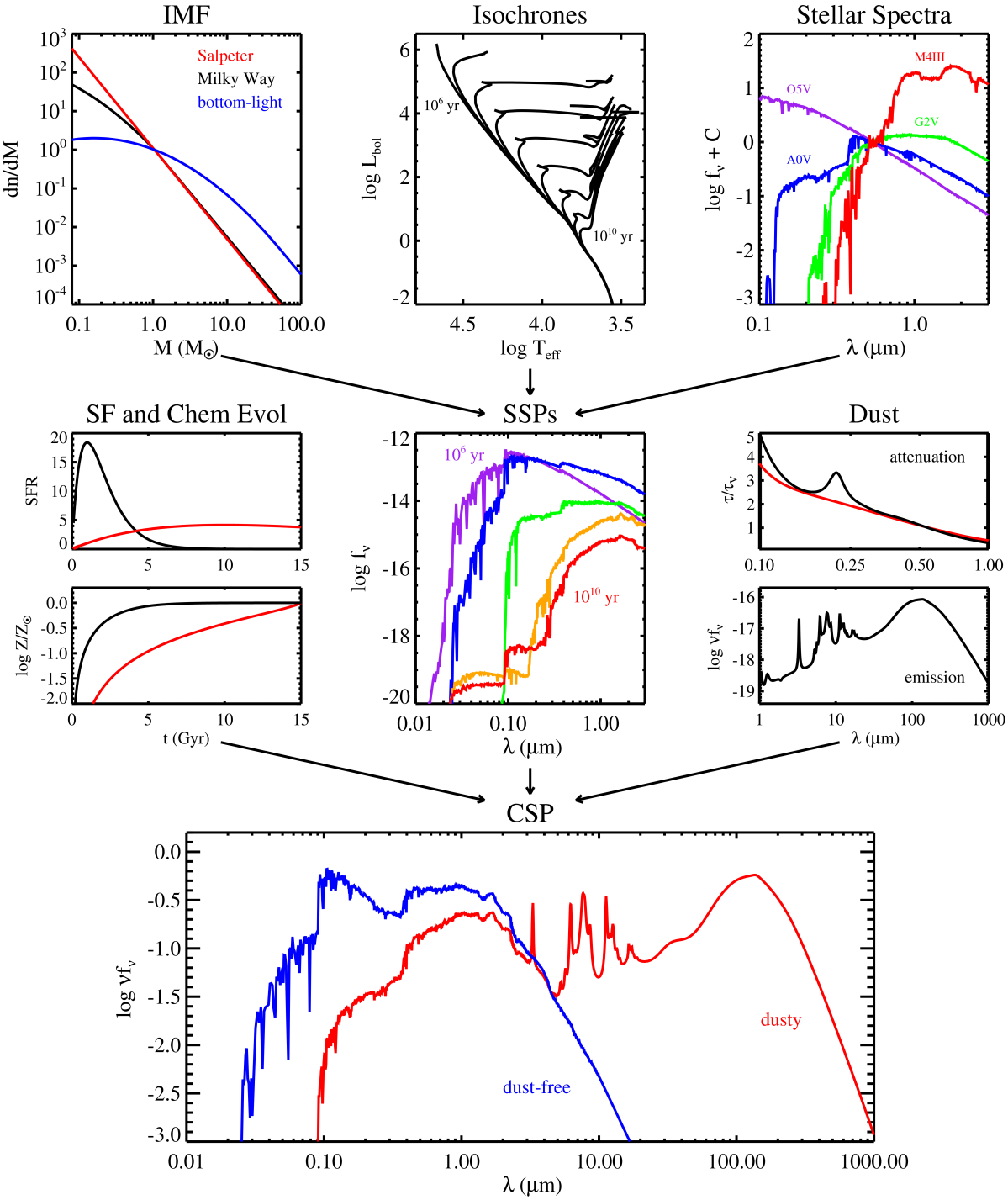

Figure 1.6: Overview of the SPS modelling technique. The upper panels represent the necessary components to form a simple stellar population (SSP), the middle panels the same but for a composite stellar population (CSP) and the bottom row represents the final CSP with and without a dust model. Taken from Figure 1 of Conroy (2013).

This informs us that, to construct an SSP, we need three major components: an IMF, a library of

stellar spectra and a model that dictates how stars evolve over time, simply known as isochrone lines (Conroy, 2013). Let us go into detail into each component.

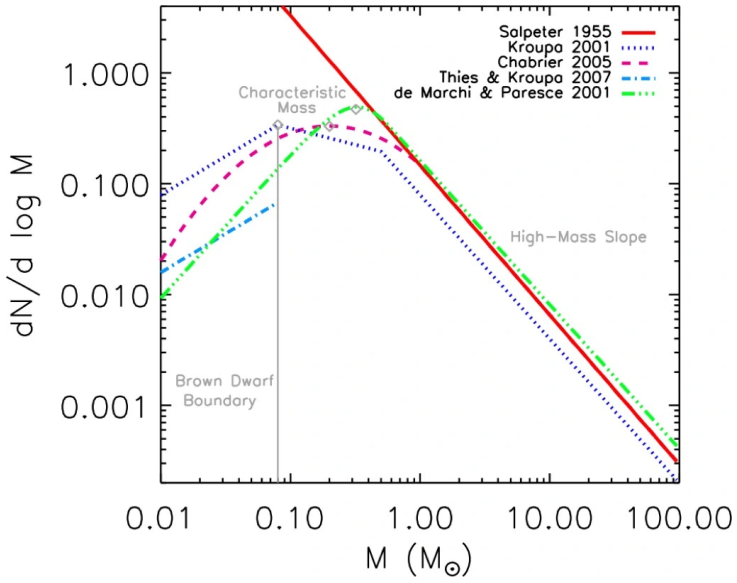

Figure 1.7: IMF functions by several authors. Taken from Figure 1 of Offner et al. (2014).

The IMF tells us the mass distribution of stars over a range of masses. Since more stars are formed with lower masses, then the IMF tends to be higher at the low mass range (as can be seen by Figure 1.7), although the cut-off point of these low mass stars is still hotly debated (Kroupa et al., 2013). Overall, the significance of the IMF to an SPS code is: the determination of the normalisation of the mass-to-light ratio, which allows us to extract the stellar mass of a galaxy through their luminosity; the determination of the luminosity evolution for passively evolving stellar populations; and how much it affects the SED of a galaxy. For the SSPs, specifically, the IMF has a very small effect on their SED but, for CSPs, it has a pronounced effect, due to the wide range of masses that it works with. Examples include the Salpeter (1955), Chabrier (2003) and Kroupa (2001) IMFs.

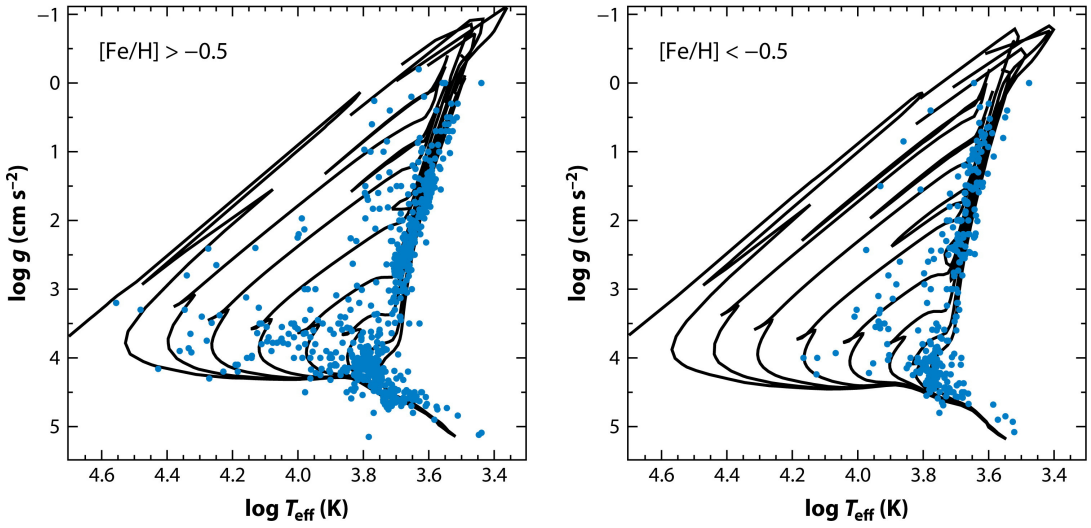

Figure 1.8: Physical properties of the MILES (Sánchez-Blázquez et al., 2006) empirical spectral library (blue dots), separated by metallicity ([Fe/H]). The black lines are isochrones from the Padova group (i.e. Marigo et al. 2008), from $t = 3 \times 10^6 - 10^{10}$ yr. Figure taken from Figure 2 of Conroy (2013).

Isochrones are lines that give us information about how stars with identical ages and metallicities can

evolve in the Hertzsprung–Russell diagram (HR diagram). This is incredibly important to implement in SPS codes, as through the isochrones we can have a relationship between $T_{\rm eff}$, $\log g$ and M for a stellar population with a specific age and metallicity. A notable example are the PARSEC-COLIBRI isochrones (Bressan et al., 2012; Marigo et al., 2013).

Finally, it is also important to have a stellar spectrum library, so we can have the spectra of stars at certain phases of their lives, depending on the physical properties mentioned above. Unfortunately, there is no one library that covers all the parameter space of the spectrum of stars, so usually several libraries are stitched together, be they theoretical or empirical. An example of a theoretical stellar spectra library is the PHOENIX library (Husser et al., 2012), while an example of an empirical stellar spectra library is the MILES library (seen in Figure 1.8, Sánchez-Blázquez et al. 2006).

With all these components mixed together, we now have a wide range of SSPs to work with. To construct the CSP, we must add three important components to the SSPs so we can emulate the SED of a galaxy: the dust, star formation history (SFH) and chemical evolution (or the metallicity evolution) of a galaxy, which can be visualised in the equation below (Conroy, 2013):

$$f_{\rm CSP}(t,Z) = \int_{t=0}^{t=t'} \int_{Z=0}^{Z_{\rm max}} \Big({\rm SFR}(t-t') P(Z,t-t') f_{\rm SSP}(t',Z) e^{-\tau_d(t')} + A f_{\rm dust}(t',Z) \Big)\, dt' dZ \tag{1.2}$$

where $P(Z,t-t')$ represents the metallicity distribution over a given time, usually as a δ-function, where for a specific time a single metallicity is considered; $e^{-\tau_d(t')}$ represents a generic time dependent dust attenuation, with τ_d representing the optical depth of the galactic medium and $A f_{\rm dust}(t',Z)$ representing the spectra of the dust emission with a normalisation constant, A, that balances the luminosity absorbed by the dust with the luminosity re-radiated by it. The time integration limits are between when the galaxy is "born" and up until the 14 or 15 Gyr mark (though this can vary), with the same principle being applied to the metallicity. Let us describe the dust and SFH contributions.

The contributions of dust are modelled in SPS codes with two principal processes: attenuation and emission. Dust attenuation is the phenomenon that obscures light by absorption and scattering of starlight through grains of dust. This should not be confused with dust extinction for two important reasons: the dust in galaxies can scatter light both out of and into the line of sight and their geometrical distribution strongly affects the SED of a galaxy. The inclusion of this dust attenuation in SPS codes is usually done by fixing the shape of the attenuation, usually through a "law", such as the Milky Way extinction curve, seen in Figure 1.9 (Fitzpatrick, 2004).

Dust emission comes from reheated dust particles, and usually dominates the SED of a galaxy with wavelengths higher than 10 μm (see middle and bottom row of Figure 1.6). There are many models that can be implemented in SPS codes that combine grain size distributions and optical properties to account for this emission, but some simpler models can be based on only one variable, such as the bolometric infrared emission of a galaxy (Chary & Elbaz, 2001) or a series of modified blackbodies (da Cunha, Charlot, & Elbaz, 2008).

The SFH of the SPS model can be as complex as the creator chooses it to be, although simple forms such as the exponentially decreasing SFR, where ${\rm SFR} \propto t^{\beta} e^{-t/\tau}$, are usually adopted for simplicity.

With all these elements together, the final stage of the SPS codes is fitting the models onto observed data. The data can vary in terms of complexity, ranging from data that covers a wide range of the electromagnetic spectrum (otherwise known as broadband SED), to just a band (such as the optical or the near-infrared), to even spectral indices (that describes the slope characteristics of the spectra over a range of wavelengths). The user defines parameters such as the SFH, metallicity and dust attenuation

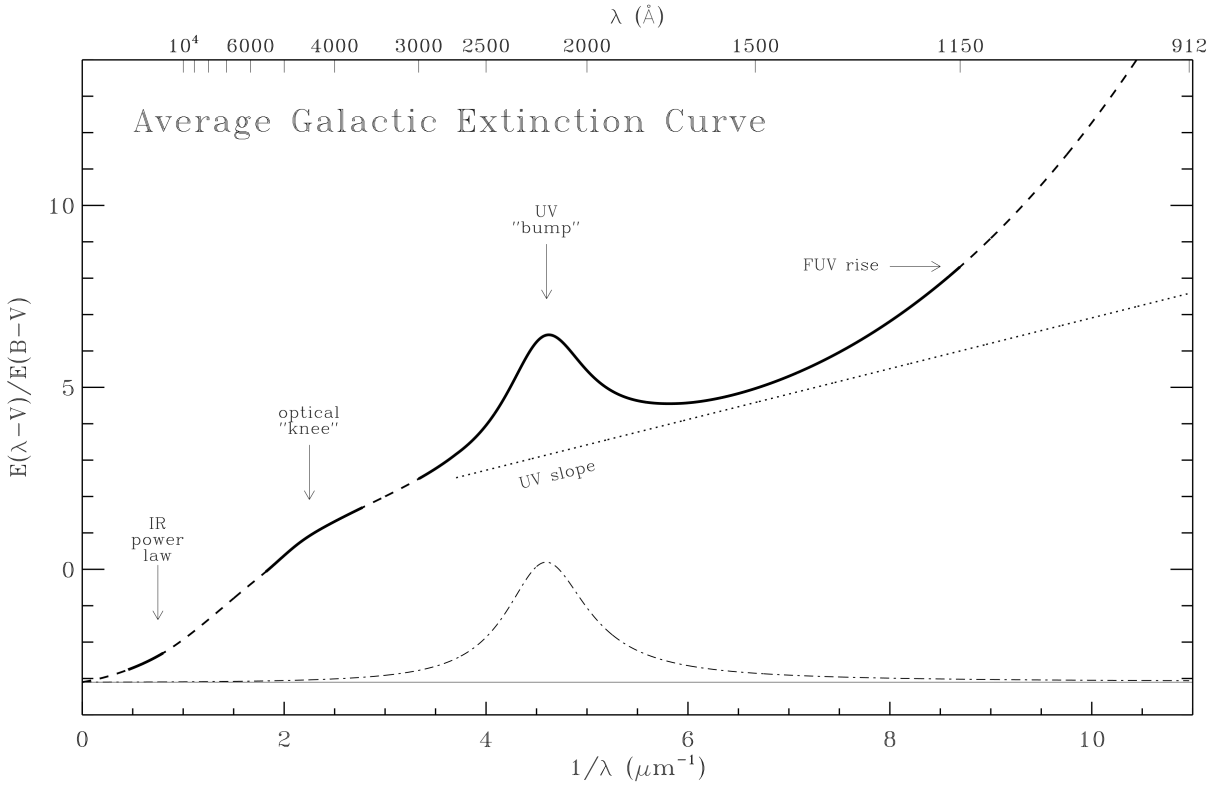

Figure 1.9: Average Milky Way extinction curve, with a ratio of V-band extinction to colour excess in the optical region, $R_V = 3.1$. Figure taken from Figure 2 of Fitzpatrick (2004).

curve. Afterwards, there is a need to define a fitting technique onto the data. The most common method is through the χ^2 minimisation techniques, but Markov Chain Monte Carlo methods become more efficient as more parameters are added to the model. Since the SED shape is the only constraint for the model parameters, parameters such as the mass-to-light ratio and the specific SFR (sSFR $\equiv$ SFR/M) must be multiplied by the luminosity and stellar mass to obtain the physical properties we desire: the stellar mass and SFR (Conroy, 2013; Walcher et al., 2011). More recently, machine learning algorithms, such as convolutional neural networks and deep learning methods, have been implemented in galaxy spectra to extract physical properties, with results that match, or even improve on, previous methods (Lovell et al., 2019; Teimoorinia et al., 2022).

There is a major component that SPS models usually neglect, however: the contribution of the nebular component. This is important, especially for galaxies with high SFRs or at high redshifts, due to the contributions of the ionised gas in the SEDs of galaxies. This does not mean that there are not any nebular continuum and emission models. CLOUDY (Ferland et al., 1998) and the previously mentioned MAPPINGS III (Groves, Dopita, & Sutherland, 2004), for example, are photoionisation codes whose purpose is to understand how chemical elements affect the continuum and emission of nebular clouds. However, most SPS models neglect to self-consistently add these contributions to their models, meaning that the SEDs produced will not be representative of the full spectra of a galaxy. The consequences of this is that the SFH and the stellar mass of galaxies can be poorly estimated. Galaxies such as Green Peas suffer from this neglect more harshly. Since they can be used as proxies for galaxies at high redshifts, understanding their SFH, chemical and stellar mass evolution is essential, due to the fact that they can give us insight into the evolution of galaxies from the primordial Universe up until recent epochs. Modelling their SEDs with and without the nebular contribution has been proven to lead to different results of these physical properties, as these galaxies suffer through period of intense star-forming activity, otherwise

known as starburst episodes. During these episodes, the nebular emission and absorption significantly alter the SED of these galaxies, meaning that neglecting their contribution will lead to consequential errors (Amorín et al., 2012; Fernández et al., 2022).

To address this issue, FADO[1], Fitting Analysis using Differential evolution Optimization (Gomes & Papaderos, 2017), an SPS code developed at the Institute of Astrophysics and Space Sciences[2], self-consistently adds the nebular contribution to its model of the spectra of galaxies, instead of separately adding this important component of an SED (Figure 1.10). This will change equation 1.2 to one of the following type:

$$f_{\rm CSP}(t,Z) = \int_{t=0}^{t=t'} \int_{Z=0}^{Z_{\rm max}} [{\rm SFR}(t-t')P(Z,t-t')\times \left(f_{\rm SSP}(t',Z)e^{-\tau_d(t')} + f_{\rm neb}(t',Z)e^{-\tau_{\rm neb}(t')}\right) + A f_{\rm dust}(t',Z)]dt'dZ \tag{1.3}$$

where $f_{\rm neb}$ represents the nebular spectrum and $e^{-\tau_{\rm neb}(t')}$ the nebular extinction associated.

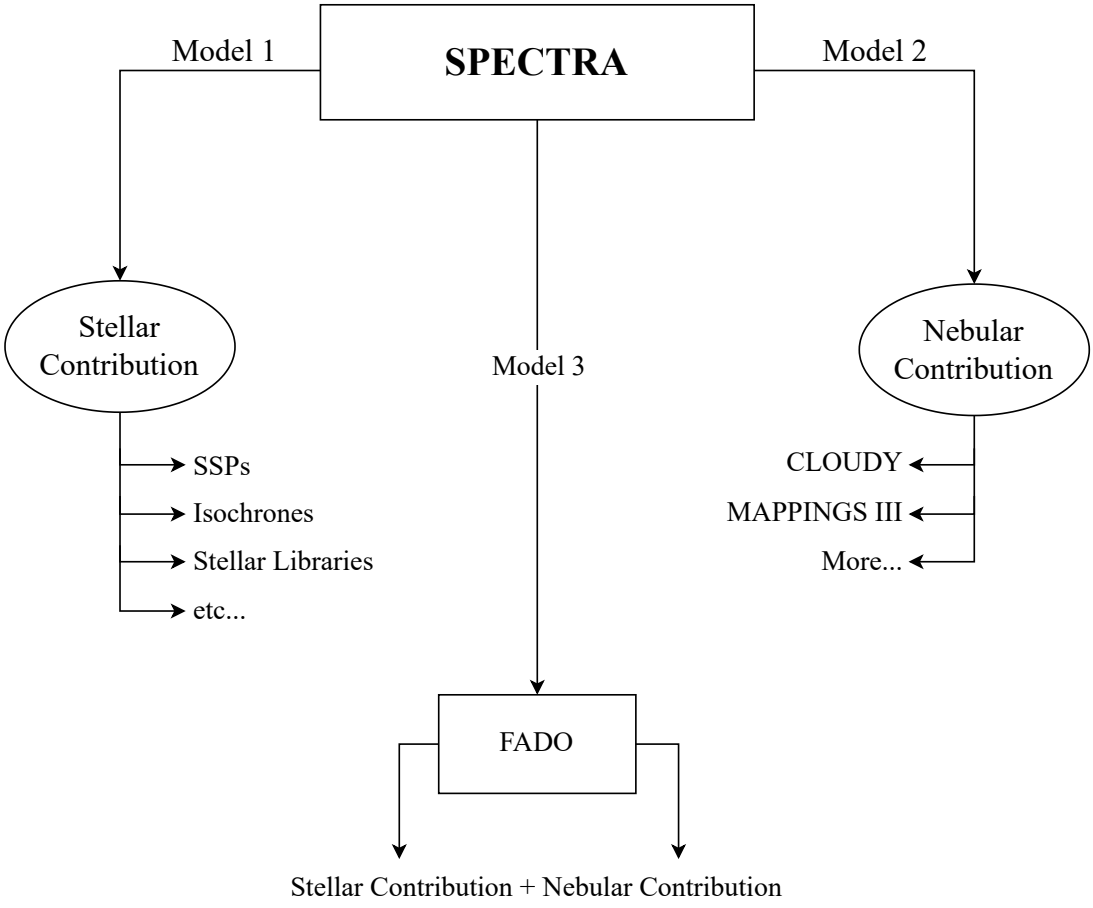

Figure 1.10: To extract the emission line fluxes of galaxies, firstly we have to apply a model for the stellar contribution (Model 1) and another for the nebular contribution (Model 2). This can lead to a lack of consistency between both components. This schematic shows the philosophy of FADO: instead of having two models that are both inherently important for the SED of a galaxy, FADO fits them both self-consistently (Model 3).

[1] https://spectralsynthesis.org/fado
[2] http://www.iastro.pt/index.html

1.3 Main sequence of galaxies

With the physical properties extracted from the SPS models, we can now compare them between a large number of galaxies, due to large scale surveys of extragalactic objects. Many authors found that, specifically, the logarithm of the SFR and the stellar mass were linearly correlated (e.g. Brinchmann et al. 2004; Eales et al. 2017; Noeske et al. 2007; Rodighiero et al. 2011).

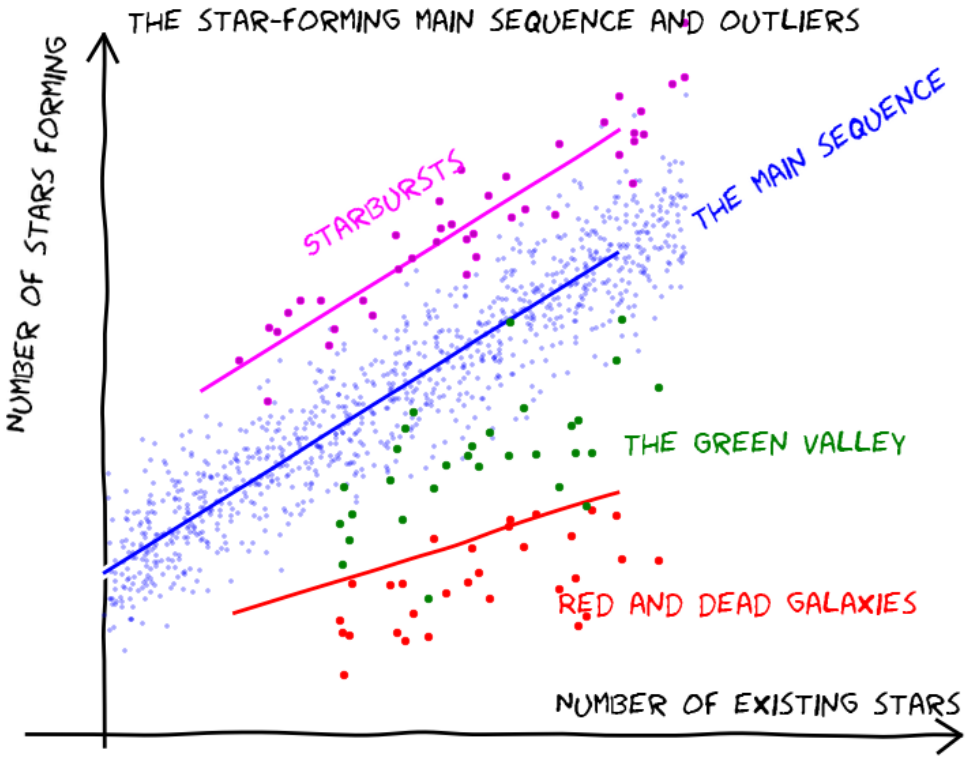

Figure 1.11: Most galaxies seem to form stars at a rate that is proportional to the number of stars that they already have. Some astronomers are calling this the main sequence of star-forming galaxies. Other galaxies fall off the sequence. The red and dead ones or quenched or quiescent ones are not forming many stars at all. On the other hand there are some galaxies forming stars at much higher rates, which we call starbursts. Then there are a few galaxies that are still forming stars, but at lower rates than on the main sequence. These populate the green valley. Caption and figure taken from the CANDEL website.

Figure 1.11 shows that there are three main types of galaxies that are correlated, and one transitional object:

- SF galaxies, which have the tightest correlation between SFR and stellar mass. This relationship was dubbed the star-forming main sequence due to the nature of the galaxies present in them;
- Starburst galaxies, outliers of the star-forming main sequence, which are galaxies that, as previously stated, are suffering through a violent episode of star-formation, hypothesised to be due to galaxy mergers;
- Passive, or retired, galaxies, or galaxies that are gas-poor and have a low capacity of forming stars, hence their names;
- Green valley galaxies, or galaxies that are not producing stars at the same rate as the galaxies in the main sequence, but have not been quenched enough to be considered passive objects.

This diagram, then, is a tool that allows us to contrast and compare galaxies with different star-forming activity.

With this in mind, our main objective in this work is to understand whether the self-consistent addition of the nebular contribution in FADO changes the way we classify galaxies through standardised emission line flux ratio diagrams,or the BPT diagrams, as well as through the main sequence, by comparing FADO results with purely stellar SPS models.

This dissertation is structured as such: in Chapter 2, we describe the data analysed by FADO and a purely stellar SPS model, the Max Planck Institute for Astrophysics and the Johns Hopkins University (MPA-JHU) models, exploring the differences between these datasets, as well as how they extracted the data. In Chapter 3, we select a subsample and then explore the differences between FADO and MPA-JHU in the BPT diagram. In Chapter 4, we explore the differences between these datasets with more emission line flux ratio diagrams, to understand if the changes seen in Chapter 3 are similar. In Chapter 5, we describe and analyse a new classification scheme that we have found when comparing the equivalent width of specific emission lines. In Chapter 6, we take a brief look at the main sequence of galaxies for the sample defined in the previous Chapters.

Chapter 2

Data Description

In this chapter, we describe and compare the results of the FADO and MPA-JHU[1] analysis of the SDSS Data Release 7[2] (DR7) galaxy sample, namely for the the flux and equivalent width of the BPT diagram emission lines.

2.1 SDSS DR7

The SDSS is a survey that covers a wide area of the sky, both in imaging and spectral mode. In this work, we are interested in specific emission lines from the SEDs of galaxies and, as such, we want the spectroscopy of the SDSS galaxies. The seventh Data Release covers 9 380 deg^2 of the sky, where 929 555 spectra were measured (Abazajian et al., 2009).

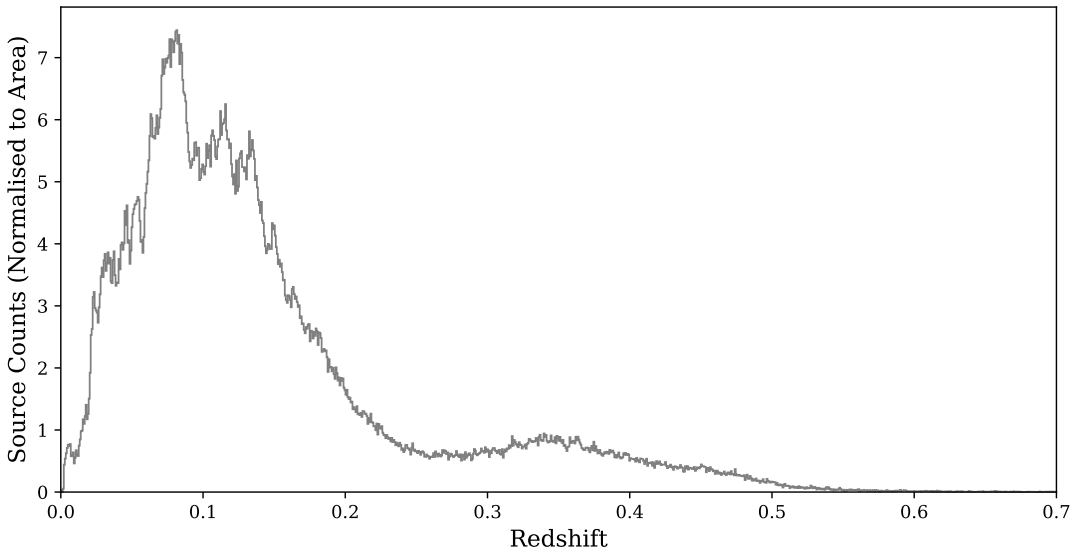

Figure 2.1: Histogram of the redshift estimated by MPA-JHU in the common sample of 916 118 galaxies between FADO and MPA-JHU of SDSS DR7.

Both FADO and MPA-JHU have processed SDSS DR7, however they have a different total number of galaxies, because of the way that each dataset handled this Data Release (Cardoso et al. 2022 for FADO and Brinchmann et al. 2004 for MPA-JHU). As such, from these datasets, we selected the objects

[1] https://wwwmpa.mpa-garching.mpg.de/SDSS/DR7/
[2] https://classic.sdss.org/dr7/

in common, which comes to a total of 916 118 galaxies, or $\approx 98.5\%$ of the previously mentioned DR7 spectra. This sample has a median redshift of $z = 0.116$, with a peak at $z = 0.081$, where 30% of the sample lies below this value and 70% lies above (Figure 2.1).

From this sample, FADO and MPA-JHU estimated the physical properties of galaxies with different processes. In this Chapter, we are going to focus on how these models processed the flux and equivalent width (EW) of the galaxies. The metallicity estimation is explained in Chapter 5 (although only MPA-JHU has a metallicity estimation) and the SFR and stellar mass estimations are explained in Chapter 6.

FADO estimates the flux of an emission lines through the following equation (Gomes & Papaderos, 2017):

$$F_\lambda = \sum_{i=1}^{N_\star} \overbrace{M_{i,\lambda_0} L_{i,\lambda} \times 10^{-0.4 A_V q_\lambda} \otimes S(v_\star, \sigma_\star)}^{\text{Stellar Component}} + \overbrace{\Gamma_\lambda(n_e, T_e) \times 10^{-0.4 A_V^{\text{neb}} q_\lambda} \otimes N(v_\eta, \sigma_\eta)}^{\text{Nebular Component}} \tag{2.1}$$

where F_λ represent the flux at a specific wavelength λ; $N_\star$ represents the number of spectra adopted from the stellar libraries; M_{i,λ_0} represents the stellar mass of the i^{th} spectral component, normalised at a wavelength λ_0; $L_{i,\lambda}$ the luminosity contribution of the i^{th} spectral component; A_V the V-band extinction, with q_λ as the conversion factor between the extinction at any wavelength, A_λ and the V-band extinction, represented as a ratio between these two extinctions; $S(v_\star, \sigma_\star)$ represents the stellar kinematics of a galaxy, with the stellar systemic velocity, $v_\star$, and velocity dispersion, $\sigma_\star$; $\Gamma_\lambda(n_e, T_e)$ represents the nebular continuum; A_V^{neb} represents the V-band nebular extinction; and $N(v_\eta, \sigma_\eta)$ represents a simulation of nebular kinematics, with the nebular systemic velocity, v_η, and velocity dispersion, σ_η.

Essentially, the observed SED of a galaxy is fit through a linear combination of not only discrete spectral components (the first part of equation 2.1), but also the addition of the nebular continuum (the second part of the equation). The stellar component is as described in Chapter 1, with the combination of the IMF, isochrones and stellar spectrum libraries. The nebular component is equally as simple, essentially just adding a photoionisation model to the fit of the SED, instead of doing it separately. FADO's photoionisation model is based on the FIVEL (Five-Level Atom) model (De Robertis, Dufour, & Hunt, 1987; Shaw & Dufour, 1994; Shaw & Dufour, 1995), where the nebular continuum is modelled assuming that all stellar photons with wavelengths below 911.76 Å are absorbed and re-emitted as nebular emission under case B recombination, and the electron density and temperature (n_e and T_e, respectively) is computed on the fly by FADO. In the case that FADO cannot calculate these values, they are assumed to be $n_e = 10^2\,\text{cm}^{-3}$ and $T_e = 10^4\,\text{K}$. The nebular extinction is calculated through the Balmer decrement (as explained in Chapter 3), when it can be reliably determined.

The inclusion of the nebular component does not only affect the emission line fluxes, but also the continuum of the SED. Instead of it being measured only through the spectral component, it is rather measured from the bottom of the absorption feature of emission lines, if they exist. This, in turn, affects the EW: for lines that have prominent absorption features, such as the Balmer emission lines, the continuum is going to be lower and, therefore, the EWs will be larger.

MPA-JHU is simpler by comparison, using only a variation of the stellar component of equation 2.1 to calculate the flux of the emission lines, and does not consider the absorption features in the calculation of the continuum. It is of note that, as the website states, the flux normalisation is not the same as the SDSS calibration. They do this because the spectrophotmetric correction is done through point-spread functions (PSFs) of stars, which means that the SDSS calibration is good for point sources but not for extended sources. As such, they renormalised the flux to match the photometric fiber magnitude in the

r-band of SDSS to correct for this inaccuracy.

To ensure that the FADO and MPA-JHU fluxes are comparable (due to the different normalisations) we made sure to divide the MPA-JHU emission line fluxes by the renormalisation factor that the dataset provides, as well as to multiply the FADO emission lines fluxes by its normalisation factor.

2.2 Flux comparison

The BPT diagram is the most common emission line ratio diagram used, so, as a first look, we focus on comparing the emission line fluxes and the signal-to-noise ratios (S/N) of $[\mathrm{NII}]\lambda6584, \mathrm{H}\alpha, [\mathrm{OIII}]\lambda5007$ and $\mathrm{H}\beta$ between FADO and MPA-JHU.

Table 2.1: *Left:* Median differences (in dex) between FADO and MPA-JHU in emission line flux and S/N for the 916 118 galaxies in the common sample. *Right:* Median uncertainties associated to each emission line for FADO and MPA-JHU for the same sample, in dex.

Emission line	FADO − MPA-JHU Flux (dex)	FADO − MPA-JHU S/N (dex)
$\mathrm{H}\alpha$	0.076	0.055
$\mathrm{H}\beta$	0.189	-0.347
$[\mathrm{NII}]\lambda6584$	0.088	-0.069
$[\mathrm{OIII}]\lambda5007$	0.200	-0.268

Emission line	FADO σ_{F_λ} (dex)	MPA-JHU σ_{F_λ} (dex)
$\mathrm{H}\alpha$	0.078	0.061
$\mathrm{H}\beta$	0.414	0.122
$[\mathrm{NII}]\lambda6584$	0.118	0.078
$[\mathrm{OIII}]\lambda5007$	0.380	0.172

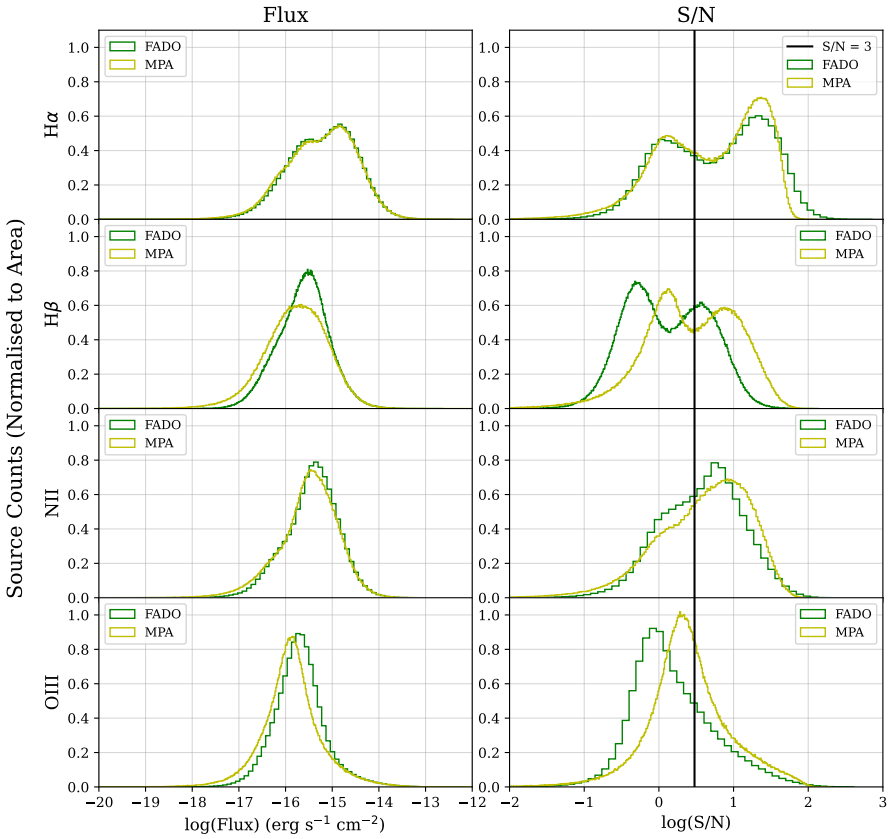

Figure 2.2: Histogram of the fluxes and S/N of the FADO and MPA-JHU datasets for the $\mathrm{H}\alpha$, $\mathrm{H}\beta$, $[\mathrm{NII}]\lambda6584$ and $[\mathrm{OIII}]\lambda5007$ emission lines, in the common sample of 916 118 galaxies. The black vertical line represents when S/N = 3 for each emission line.

From the common sample defined in the previous section, we checked the flux (F_λ) and the signal-to-noise ratio of these four emission lines. We define the S/N as the ratio between the emission line flux with its uncertainty:

$$\text{S/N} = \frac{F_\lambda}{\sigma_{F_\lambda}} \tag{2.2}$$

Figure 2.2 and Table 2.1 compare these properties between the two datasets. As can be seen, the emission line fluxes have the highest number of galaxies in the $10^{-18} < F_\lambda\,[\text{erg s}^{-1}\text{cm}^{-2}] < 10^{-12}$ interval ($\approx 88\%$ of the common sample for all emission lines), and the differences between FADO and MPA-JHU are smaller in Hα and [NII]λ6584 ($\approx 19\%$ and $\approx 22\%$, respectively) than Hβ and [OIII]λ5007 ($\approx 55\%$ and $\approx 58\%$, respectively).

Table 2.2: Extinction law differences between FADO (equation 2.3) and MPA-JHU (equation 2.4), for the emission lines considered, normalised for Hα.

Emission Line	FADO − MPA-JHU
Hα	0
Hβ	0.148
[NII]λ6584	-0.002
[OIII]λ5007	0.134

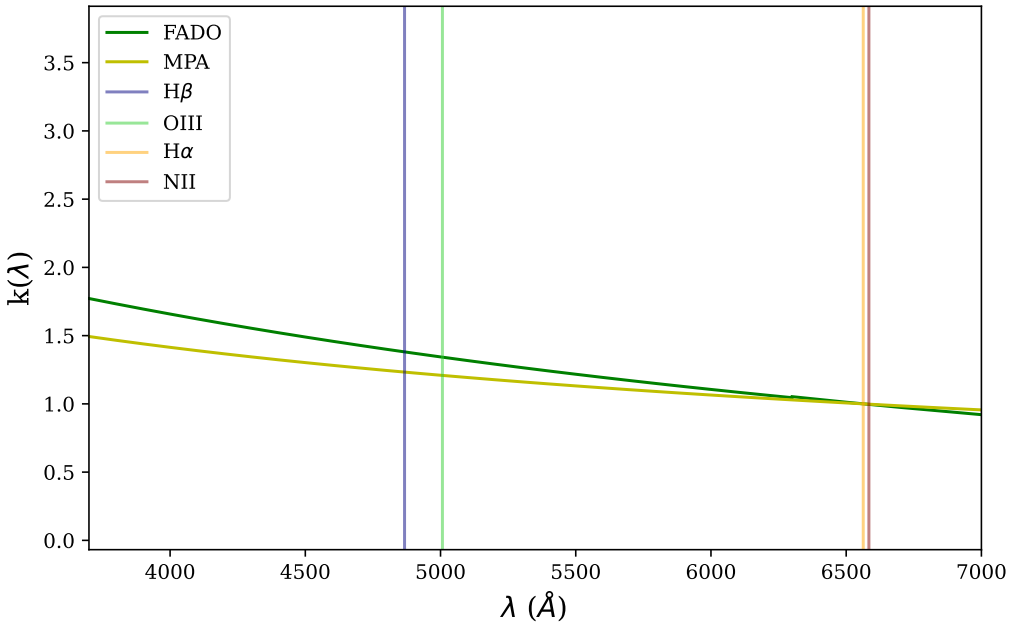

Figure 2.3: Extinction law comparison between FADO (equation 2.3) and MPA-JHU (equation 2.4), normalised for Hα. The relevant emission lines have also been annotated.

These differences arise due to the extinction laws that each dataset considered when modelling the SED: FADO considered the Calzetti et al. (2000) extinction law (equation 2.3), while MPA-JHU considered the Charlot and Fall (2000) extinction law (equation 2.4).

$$\begin{aligned} k(\lambda) &= 2.659\left(-2.156 + \frac{1.509}{\lambda} - \frac{0.198}{\lambda^2} + \frac{0.011}{\lambda^3}\right) + 4.05, \text{ for } 0.12\,\mu\text{m} \leq \lambda < 0.63\,\mu\text{m} \\ k(\lambda) &= 2.659\left(-1.857 + \frac{1.040}{\lambda}\right) + 4.05, \text{ for } 0.63\,\mu\text{m} \leq \lambda \leq 2.20\,\mu\text{m} \end{aligned} \tag{2.3}$$

$$k(\lambda) = \left(\frac{\lambda}{5500}\right)^{-0.7}, \text{ with } \lambda \text{ in Å} \tag{2.4}$$

When we compare them (Figure 2.3), we can see that the bluer emission lines (in this case, Hβ and [OIII]λ5007) have a wider gap between extinction laws than the redder emission lines (Hα and [NII]λ6584). The differences between FADO and MPA-JHU are larger at shorter wavelengths (Table 2.2). This explains why Hβ and [OIII]λ5007 have higher differences in flux than Hα and [NII]λ6584. Another possible reason for this difference is due to the flux calculation of each model, but that will be discussed in the next section, as it has more of an impact on the equivalent widths.

As is expected, these differences propagate to the S/N, with Hα and [NII]λ6584 having smaller differences ($\approx 14\%$ and $\approx -17\%$, respectively) than Hβ and [OIII]λ5007 ($\approx -122\%$ and $\approx -85\%$, respectively). FADO has significantly lower S/Ns than MPA-JHU in most emission lines, however, and this is due to each model's calculation of errors. FADO's error comes from the standard deviation of the emission line flux for each galaxy (Cardoso et al., 2022). MPA-JHU's error calculations, on the other hand, are more complex, where the full likelihood distribution of these emission line fluxes are generated, allowing for rigorous confidence intervals to be calculated, yielding lower values (Brinchmann et al., 2004). It is of note that, since the flux uncertainties compensate for the differences in flux, then we can safely state that the fluxes between FADO and MPA-JHU match, despite their differences.

Table 2.3: Galaxy classification according to the FADO and MPA-JHU categories, in the common sample of 916 118 galaxies.

FADO			MPA-JHU		
Criteria	Galaxy #	Percentage	Criteria	Galaxy #	Percentage
Pure SF	219 757	24	SF	202 624	22
SF	91 236	10	Low S/N SF	177 971	20
Composite	178 583	20	Composite	47 385	5
LINER	126 547	14	Low S/N LINER	67 932	7
Seyfert	69 980	7	AGN Non-LINER	22 704	3
Unclassifiable	230 015	25	Unclassifiable	397 502	43
Total	916 118	-	Total	916 118	-

Pressing on, galaxies in FADO and MPA-JHU are classified according to different criteria. In FADO we have six categories, according to the placement of the galaxy in the BPT diagram (Gomes & Papaderos, 2017): Pure SF, SF, Composite, Seyfert, Liner and Unclassifiable (in Chapter 3 the BPT placements will be discussed in more detail). Analogously, MPA-JHU also has six categories, thoroughly described in Section 2.1 of Brinchmann et al. (2004), with much of the same features as FADO, albeit with more S/N separation: they are SF, Low S/N SF, Composite, Low S/N LINER, AGN Non-LINER and Unclassifiable. The distribution of the galaxies in the common sample can be seen in Table 2.3. Overall, FADO has less Unclassificable galaxies than MPA-JHU, due to the fact that the dataset does not restrict galaxies according to S/N, meaning that these galaxies will be redistributed among all categories. It would be expected, then, that FADO had more galaxies across all classifications, but this is only true for Composites and AGNs (Seyferts and LINERs). There are more SF galaxies in MPA-JHU than FADO, and this is due to the lax classification of the Low S/N SF category, that is based on a vertical line with low S/N galaxies (when $\log([\text{NII}]\lambda 6584/\text{H}\alpha) = -0.22$), rather than the usual hyperboles represented in the BPT diagram.

Joining the SF and AGN categories in FADO and MPA-JHU to create two simplified categories, we can see the distribution of these galaxies, in relation to the common sample of 916 118 galaxies, in Figure

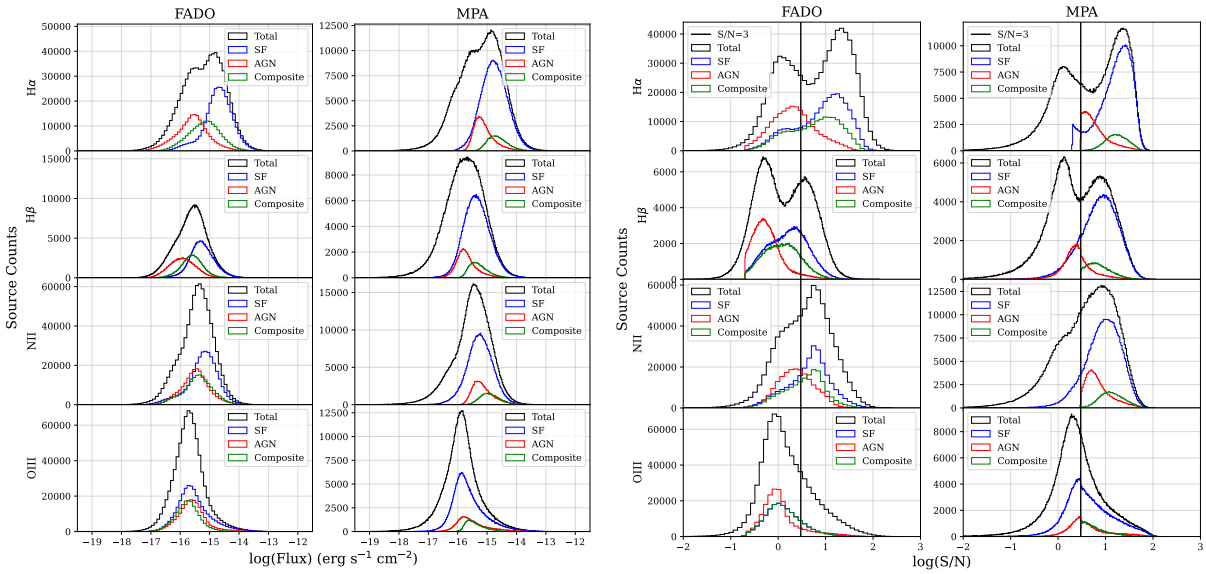

Figure 2.4: Histogram of the fluxes (left) and S/N (right) of the FADO and MPA-JHU datasets for the Hα, Hβ, [NII]λ6584 and [OIII]λ5007 emission lines, in the common sample of 916 118 galaxies. The black histogram represents the full common sample, the blue histogram the SF sample, the red histogram the AGN sample and the green histogram the Composite sample. The histograms may not completely match due to the existence of galaxies that are considered unclassifiable, and the effect is much more pronounced in the MPA-JHU sample, as there are more unclassified galaxies. The black vertical line represents when S/N = 3 for each emission line.

2.4 (without the Unclassifiable galaxies). This explains a phenomenon that we could see in the Hα flux and most of the S/N of the emission lines in Figure 2.2: a bimodality in the distribution of galaxies. In most of the emission lines, the SF galaxies have higher fluxes and S/N than the AGNs, with Composites lying in between these two categories, creating a double peak in the histogram.

2.3 Equivalent Width comparison

Equivalent widths are, essentially, the division between the flux and the continuum at a specific wavelength on the SED of galaxy. Since they can tell us the difference in the continuum estimations between FADO and MPA-JHU in the case where the fluxes are similar (which we are considering as such), they are another property that is relevant to compare between datasets. The EW has an uncertainty associated from each dataset, which we will define in general as σ_{EW}.

Table 2.4: Median differences (in dex) between FADO and MPA-JHU in EW and σ_{EW} of the EW for the 916 118 galaxies in the common sample.

Emission line	FADO − MPA-JHU EW (dex)	FADO − MPA-JHU σ_{EW} (dex)
Hα	0.112	0.382
Hβ	0.242	0.566
[NII]λ6584	0.038	0.271
[OIII]λ5007	0.141	0.343

Figure 2.5 and Table 2.4 show the distribution and differences of the EWs of the BPT diagram emission lines, as well as the differences between FADO and MPA-JHU. The median differences between datasets are larger in the EW measurements than the flux measurements of the Balmer lines: Hα has a

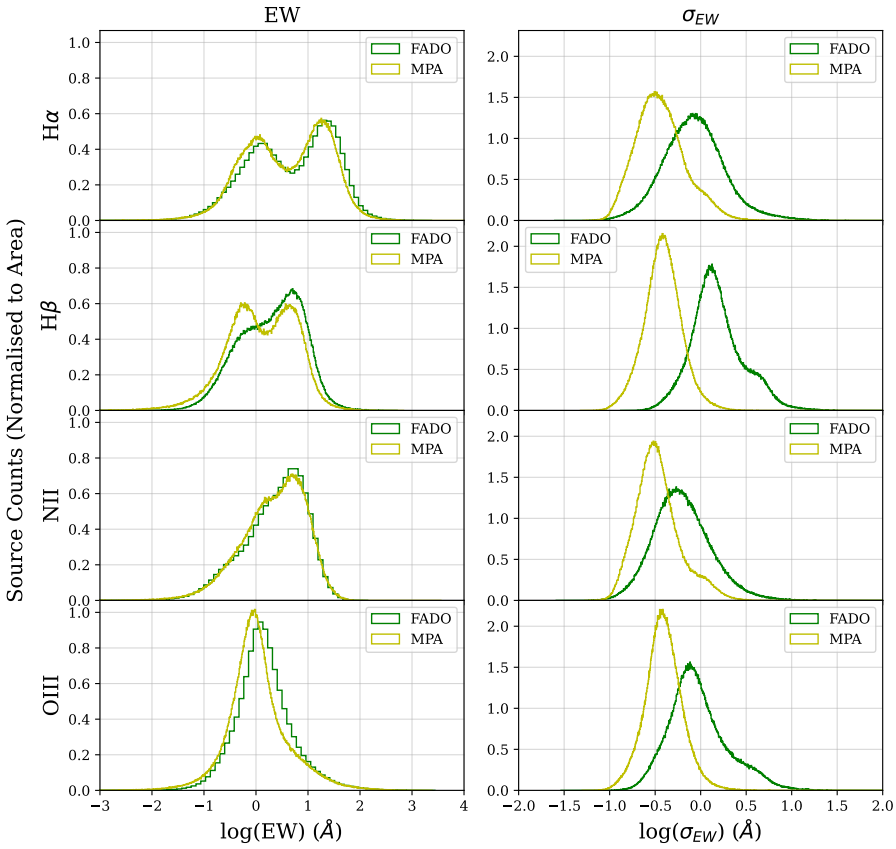

Figure 2.5: Histogram of the EW and uncertainty of the EW of the FADO and MPA-JHU datasets for the Hα, Hβ, [NII]λ6584 and [OIII]λ5007 emission lines, in the common sample of 916 118 galaxies.

difference of $\approx 29\%$ and Hβ with $\approx 75\%$, both higher than their flux counterparts. For the forbidden lines of [OIII]λ5007 and [NII]λ6584, both have lower differences than their flux counterparts ($\approx 38\%$ and $\approx 9\%$, respectively).

While this difference could affirm that FADO has lower continuums than MPA-JHU, this is not the case, because MPA-JHU measures the flux from the continuum (Brinchmann et al., 2004), while FADO measures it from the absorption feature of the emission line in question (Gomes & Papaderos, 2017). In the flux of an emission line, the impact is negligible and, if there exists any differences, they can be justified by other reasons (like the extinction law difference explained in the previous section). However, in emission lines with notable absorption features (such as the Balmer lines), the continuum between FADO and MPA-JHU is different. As such, we expect the Balmer lines to have larger differences in the EWs between datasets than their fluxes. The opposite is true of the forbidden lines, since they do not have prominent absorption lines. Comparing Tables 2.4 and 2.1 proves these statements, as the the EW differences between datasets is higher than the flux differences for the Balmer lines, with the reverse happening to the forbidden lines.

When it comes to the uncertainties of the EWs, we can see from Table 2.4 that the differences between datasets are high for all emission lines: Hα has a difference of $\approx 141\%$, Hβ has a $\approx 268\%$ difference, and [NII]λ6584 and [OIII]λ5007 have a difference of $\approx 87\%$ and $\approx 120\%$, respectively. This happens for the same reasons as explained in the previous section: FADO and MPA-JHU calculate the errors with significantly different methods and, because of this, the latter dataset has lower errors than the former - this is compounded with the fact that the EW is more complicated to calculate than the flux, meaning that the errors in FADO are substantially higher than MPA-JHU's. This also implies that, with the addition

of the uncertainties, the EWs of these datasets also match, much like the fluxes.

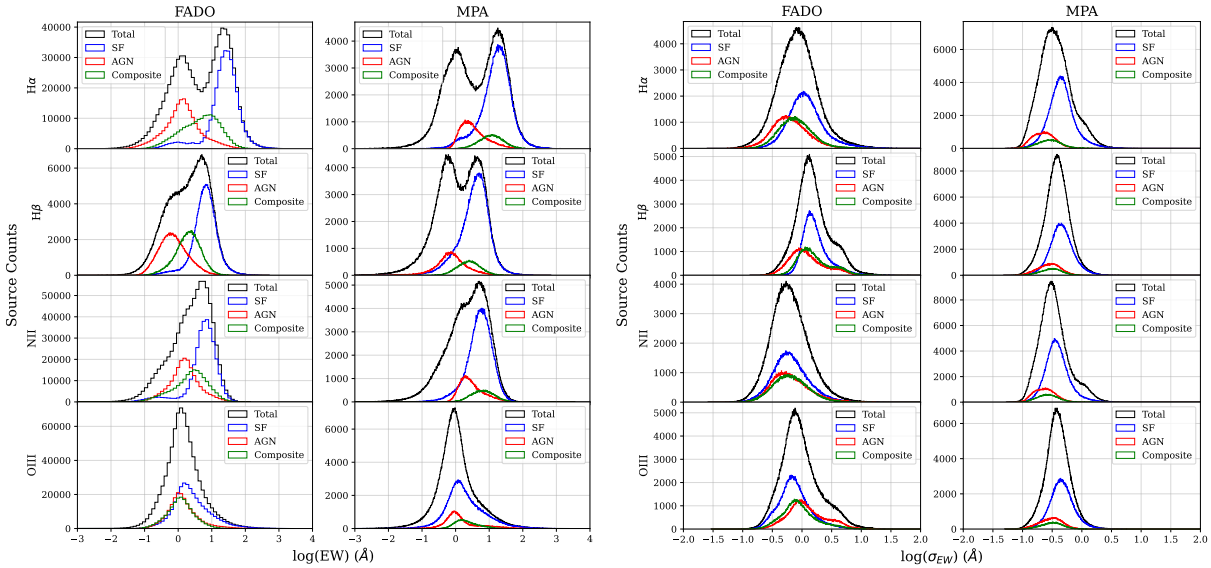

Figure 2.6: Histogram of the EWs (left) and uncertainty of the EWs (right) of the FADO and MPA-JHU datasets for the Hα, Hβ, [NII]λ6584 and [OIII]λ5007 emission lines, in the common sample of 916 118 galaxies. Each histogram represents the same as in Figure 2.4.

Additionally, the bimodality present in the fluxes can also be seen in the EWs of the emission lines selected (Figure 2.6). Noticeably, while in the fluxes of Hβ and [NII]λ6584 the bimodality isn't prevalent, in these emission lines' EWs the bimodality is much more pronounced. It is also of note that [OIII]λ5007 appears to have no sensibility to the different classifications of galaxies, either in the flux or the EW.

Chapter 3

BPT Diagram with FADO and MPA-JHU

There are many diagrams one can use to classify galaxy types using optical emission lines (Baldwin, Phillips, & Terlevich, 1981; Kauffmann et al., 2003; Kewley et al., 2001; Kewley et al., 2006; Veilleux & Osterbrock, 1987), but, as a first look, and to keep the classification consistent with both Brinchmann et al. 2004 (the classification for MPA-JHU) and Gomes and Papaderos 2017 (the classification for FADO), we start classifying the galaxy sample with the comparison between the emission line flux ratios of $[\mathrm{NII}]\lambda 6584/\mathrm{H}\alpha$ and $[\mathrm{OIII}]\lambda 5007/\mathrm{H}\beta$, or the BPT diagram.

3.1 Sample Selection

From the common sample described in the previous Chapter (Table 2.3), we have decided to restrict the galaxies in FADO and MPA-JHU with two criteria:

a) The S/N of each emission line ($\mathrm{H}\alpha$, $\mathrm{H}\beta$, $[\mathrm{NII}]\lambda 6584$ and $[\mathrm{OIII}]\lambda 5007$) is higher than 3;

b) The emission line flux from all the emission lines mentioned is between $10^{-18} < F_\lambda\,[\mathrm{erg\ s^{-1}cm^{-2}}] < 10^{-12}$.

We have done this to make sure that the galaxies we selected don't suffer from contamination and unusually high uncertainties, as well as due to the fact that most of the galaxies are between the range of fluxes mentioned (see Figure 2.2).

Table 3.1: Galaxy classification according to FADO and MPA-JHU categories. This takes into account limitations in the S/N and flux (S/N > 3 and $-18 < \log_{10}(F\,[\mathrm{erg\ s^{-1}cm^{-2}}]) < -12$) for the $\mathrm{H}\alpha$, $\mathrm{H}\beta$, $[\mathrm{NII}]\lambda 6584$ and $[\mathrm{OIII}]\lambda 5007$ emission lines.

FADO			MPA-JHU		
Criteria	Galaxy #	Percentage	Criteria	Galaxy #	Percentage
Pure SF	114 099	70	SF	202 623	74
SF	23 083	14	Low S/N SF	0	0
Composite	16 990	11	Composite	47 385	17
LINER	831	< 1	Low S/N LINER	0	0
Seyfert	6 083	4	AGN Non-LINER	22 702	8
Unclassifiable	1	< 1	Unclassifiable	0	0
Total	161 087	-	Total	272 710	-

After applying the previously mentioned criteria, we can see that the sample goes from the 916 118 sources to 161 087 galaxies in FADO ($\approx 17\%$ of the common sample) and 272 710 galaxies in MPA-JHU ($\approx 30\%$ of the common sample). The distribution of these across the FADO and MPA-JHU categorisation

of galaxies can be seen in Table 3.1. Our cut-off criteria also has the consequence of removing all of the low S/N galaxies in the MPA-JHU dataset, as can be seen by comparing Tables 2.3 and 3.1.

To simplify our efforts, we have decided to join these six categories into four, more easy to understand categories:

- The Pure SF and SF categories in FADO are joined together to become the SF category. The same logic is applied to the SF and the Low S/N SF categories in MPA-JHU.
- The Composite category in FADO and MPA-JHU remains the same.
- The LINER and Seyfert categories in FADO are joined together to become the AGN category. The same logic is applied to MPA-JHU with the Low S/N LINER and the AGN Non-LINER categories. Despite the fact that the mechanisms of emission in LINER and Seyfert are different, we joined them together because we wanted to simplify classification into three categories: SF, Composite and AGN.
- The Unclassificable category in FADO and MPA-JHU remains the same.

This changes the previous Table to a simplified version, that can be seen in Table 3.2. However, FADO and MPA-JHU have a significantly different number of total galaxies, despite the same criteria being applied to both datasets. Figure 2.2 and Table 2.1 shows that the S/Ns of FADO are lower than MPA-JHU's, meaning that the cutoff at S/N > 3 is going to have a larger effect on the former rather than the latter dataset, explaining the difference in galaxy counts.

Table 3.2: Galaxy classification according to the categories defined in this work. This takes into account limitations in the S/N and flux (S/N > 3 and $-18 < \log_{10}(F[\text{erg s}^{-1}\text{cm}^{-2}]) < -12$) for the H$\alpha$, H$\beta$, [NII]$\lambda$6584 and [OIII]$\lambda$5007 emission lines.

	FADO		MPA-JHU	
Criteria	Galaxy #	Percentage	Galaxy #	Percentage
SF	137 182	85	202 623	74
Composite	16 990	11	47 385	17
AGN	6 914	4	22 702	8
Unclassifiable	1	< 1	0	0
Total	161 087	-	272 710	-

Afterwards, we applied the criteria on both datasets simultaneously, so we can have the same galaxies in FADO and MPA-JHU and, from which, we can compare how they classify them. This results in a total of 156 750 galaxies (or $\approx 17\%$ of the common sample), whose distribution over the four previously mentioned categories can be seen in Table 3.3.

Table 3.3: Galaxy classification according to the categories defined in this work, taking into account the limitations in the S/N and flux (S/N > 3 and $10^{-18} < F_\lambda[\text{erg s}^{-1}\text{cm}^{-2}] < 10^{-12}$) for the H$\alpha$, H$\beta$, [NII]$\lambda$6584 and [OIII]$\lambda$5007 emission lines in both FADO and MPA-JHU simultaneously.

	FADO		MPA-JHU	
Criteria	Galaxy #	Percentage	Galaxy #	Percentage
SF	133 595	85	135 345	86
Composite	16 428	10	14 891	9
AGN	6 726	4	6 514	4
Unclassifiable	1	< 1	0	0
Total	156 750	-	156 750	-

This is the sample we will be working with for the rest of this chapter, where the flux and S/N distribution can be seen in Figure 3.1 and the differences between the fluxes and S/N between both datasets can be seen in Table 3.4. The differences in the flux become minute compared to the sample with 916 118 galaxies: Hα has a difference of $\approx -0.2\%$, Hβ has a $\approx -0.7\%$ difference, and [NII]λ6584 and [OIII]λ5007 have $\approx 1\%$ and $\approx 5\%$ differences, respectively. Despite this, the S/N differences remain mostly the same as the common sample: Hα and [NII]λ6584 remain with the lowest differences ($\approx 20\%$ and $\approx -28\%$, respectively), with Hβ and [OIII]λ5007 having noticeably higher gaps in S/N ($\approx -130\%$ and $\approx -50\%$, respectively).

Table 3.4: Median differences (in dex) between FADO and MPA-JHU in emission line flux and S/N for the 156 750 galaxies in the common sample.

Emission line	FADO − MPA-JHU Flux (dex)	FADO − MPA-JHU S/N (dex)
Hα	-0.001	0.079
Hβ	-0.003	-0.363
[NII]λ6584	0.006	-0.109
[OIII]λ5007	0.022	-0.172

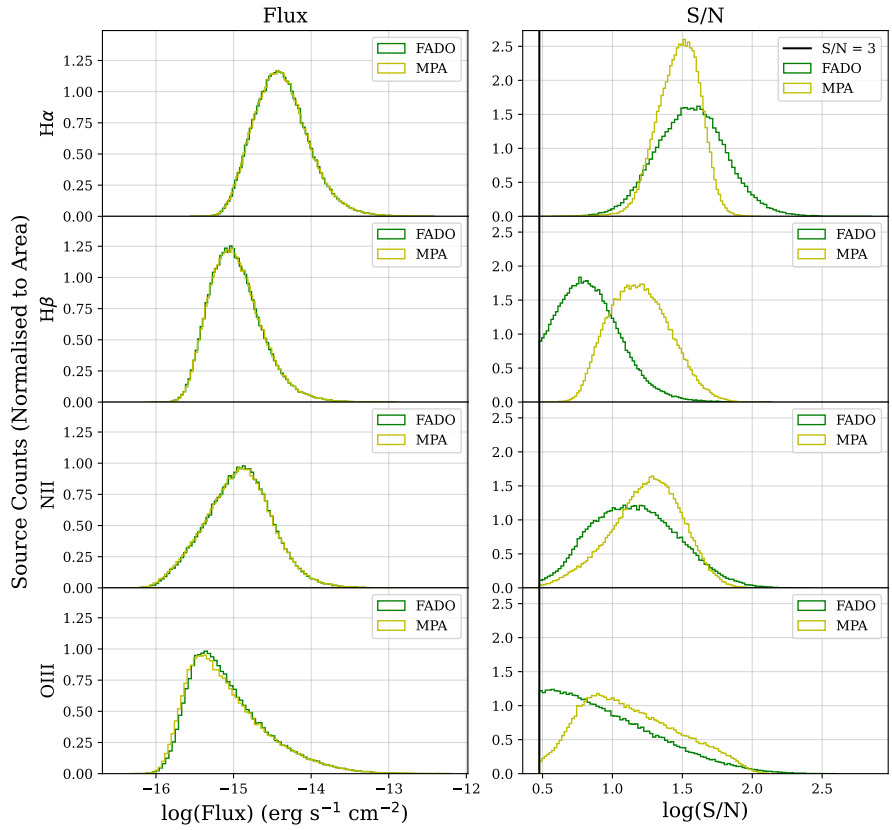

Figure 3.1: Histogram of the fluxes and S/N of the FADO and MPA-JHU datasets for the Hα, Hβ, [NII]λ6584 and [OIII]λ5007, in the common sample of 156 750 galaxies (Table 3.3). The black vertical line represents when the S/N= 3.

Figure 3.2 and Table 3.5 show the distribution of the EWs of the sample. They also follow the same rule of thumb as the fluxes: the differences between the EW of each dataset become smaller, with Hα and Hβ having smaller differences than the common sample ($\approx 25\%$ and $\approx 35\%$, respectively), and an added consequence of the differences in the forbidden lines becoming minute, nearly negligible (0% for

[NII]λ6584 and $\approx$ 6% for [OIII]λ5007). Despite this, the difference of uncertainties between FADO and MPA-JHU remain as large as the common sample, with the Balmer lines having higher differences (Hα with a $\approx$ 139% difference and Hβ with a $\approx$ 252% difference) than the forbidden lines ($\approx$ 66% difference for [NII]λ6584 and $\approx$ 58% difference for [OIII]λ5007).

It is also noteworthy that the bimodality in both the flux and the EW disappears, due to the fact that the AGN galaxies have a much lower source count than the SF galaxies and because of the S/N cut. As such, we will stop having two peaks in the distribution.

Table 3.5: Median differences (in dex) between FADO and MPA-JHU in EW and σ_{EW} for the 156 750 galaxies in the common sample.

	FADO − MPA-JHU	
Emission line	EW (dex)	σ_{EW} (dex)
Hα	0.098	0.379
Hβ	0.129	0.546
[NII]λ6584	0	0.221
[OIII]λ5007	0.025	0.190

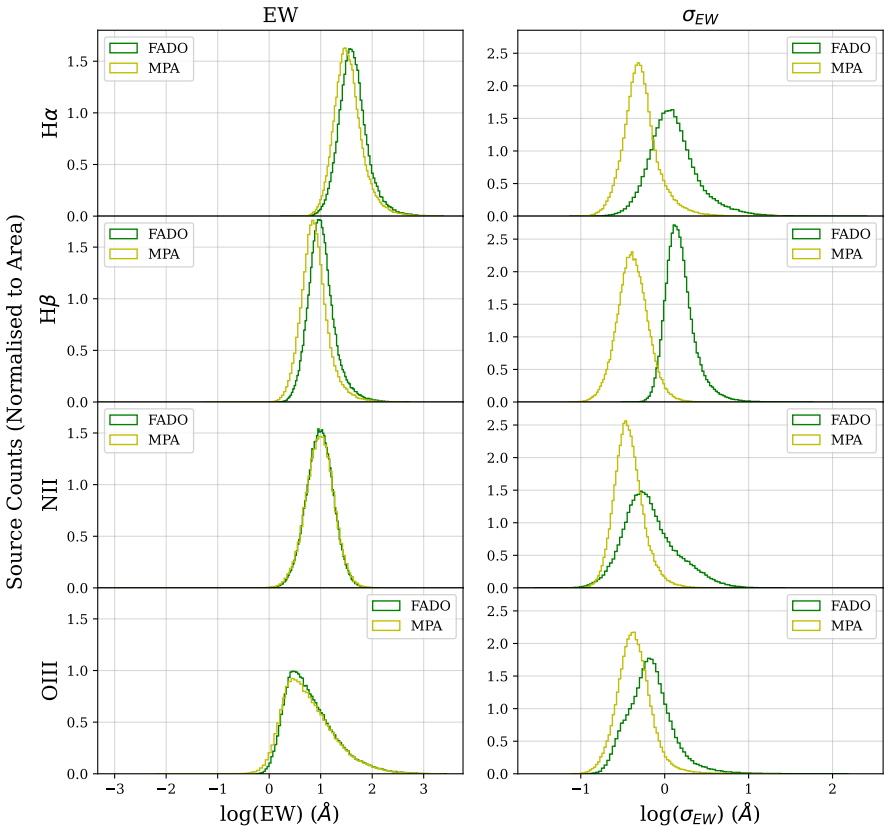

Figure 3.2: Histogram of the EWs and uncertainties of the EWs of the FADO and MPA-JHU datasets for the Hα, Hβ, [NII]λ6584 and [OIII]λ5007, in the common sample of 156 750 galaxies (Table 3.3).

3.2 Extinction correction

Before we move on to plot the BPT diagram, we must first correct the emission line fluxes for line of sight extinction, where the dust and gas present between the observer (us) and the objects (the galaxies)

affect their spectra, reddening them and lowering the fluxes from the emission lines. To correct for this effect, we must first quantify the differences between the uncorrected and corrected fluxes for a specific wavelength, something that will henceforth be called extinction, and defined as:

$$m(\lambda) - m^0(\lambda) = A(\lambda) \tag{3.1}$$

where $m(\lambda)$ represents the apparent magnitude of a galaxy as we see it (i.e., with the dust and gas present between the observer and the object), $m^0(\lambda)$ the apparent magnitude of a galaxy if there were not any extinction effects and $A(\lambda)$ the extinction. Since the difference in magnitudes corresponds to a ratio of fluxes, then we can also state that:

$$m(\lambda) - m^0(\lambda) = -2.5 \log\left(\frac{F_\lambda}{F_\lambda^0}\right) \tag{3.2}$$

where F_λ represents extinction-uncorrected flux and F_λ^0 represents the extinction-corrected flux. Joining together equations 3.1 and 3.2, we get that:

$$A(\lambda) = -2.5 \log\left(\frac{F_\lambda}{F_\lambda^0}\right) \Leftrightarrow F_\lambda^0 = F_\lambda \, 10^{0.4A(\lambda)} \tag{3.3}$$

or, in other words, the corrected flux is a defined by a power law with the extinction playing a major factor. We are considering this power law approximation because we are assuming a uniform layer of gas and dust between us and the galaxies and that the extinction does not depend on geometry.

Extinction does not affect every wavelength in the same way, however - shorter wavelengths suffer more from this effect than longer wavelengths. As such, we must also take this phenomenon into account, quantifying it and adding it to the extinction. One way to do this is to consider the Balmer decrement between Hα and Hβ, or the difference between the extinctions of the flux of these emission lines, since Hβ has a shorter wavelength than Hα and is, therefore, more affected by the effects of dust and gas. As such:

$$\begin{aligned} A(\mathrm{H}\beta) - A(\mathrm{H}\alpha) &= -2.5 \log\left(\frac{F(\mathrm{H}\beta)}{F(\mathrm{H}\beta)^0}\right) + 2.5 \log\left(\frac{F(\mathrm{H}\alpha)}{F(\mathrm{H}\alpha)^0}\right) = \\ &= 2.5 \log\left(\frac{F(\mathrm{H}\alpha)/F(\mathrm{H}\beta)}{F(\mathrm{H}\alpha)^0/F(\mathrm{H}\beta)^0}\right) = 2.5 \log\left(\frac{(\mathrm{H}\alpha/\mathrm{H}\beta)_{obs}}{(\mathrm{H}\alpha/\mathrm{H}\beta)_{int}}\right) \end{aligned} \tag{3.4}$$

where $(\mathrm{H}\alpha/\mathrm{H}\beta)_{obs}$ is the observed Balmer decrement, or the ratio between the fluxes we extract from the models, and $(\mathrm{H}\alpha/\mathrm{H}\beta)_{int}$ is the intrinsic Balmer decrement, which represents the ratio between the Hα and Hβ fluxes that should exist if there was no extinction. From Osterbrock and Ferland (2006), we know that the intrinsic Balmer decrement is $(\mathrm{H}\alpha/\mathrm{H}\beta)_{int} = 2.85$ for SF galaxies and $(\mathrm{H}\alpha/\mathrm{H}\beta)_{int} = 3.1$ for AGNs, which are the same values used by Kewley et al. (2006). This assumes an electron density of $n = 100\,\mathrm{cm}^{-3}$ and electron temperature of $T = 10^4$ K for case B recombination, which is when the gas is optically thick for ionising radiation, meaning that the electron emitted after recombination is absorbed locally. The differences between Balmer decrements arise because AGNs have an additional contribution in Hβ, due to collisional excitation.

Composites are galaxies that are SF but also have traces of AGN activity - however, even though the differences between the two corrected fluxes with different intrinsic Balmer decrements are notable, rounding with a 50% difference (Figure 3.3 and Table 3.6), we cannot confidently discern the impact of AGNs on these galaxies. As such, we opted to use the SF intrinsic Balmer decrement, where

$(\mathrm{H}\alpha/\mathrm{H}\beta)_{int} = 2.85$.

Table 3.6: Median differences (in dex) between the two corrected fluxes with different intrinsic Balmer decrements in the Composite galaxies defined in Table 3.3.

$(\mathrm{H}\alpha/\mathrm{H}\beta)_{int} = 2.85 - (\mathrm{H}\alpha/\mathrm{H}\beta)_{int} = 3.1$		
	FADO	MPA-JHU
Hα	0.156	0.156
Hβ	0.193	0.193
[NII]λ6584	0.156	0.149
[OIII]λ5007	0.189	0.189

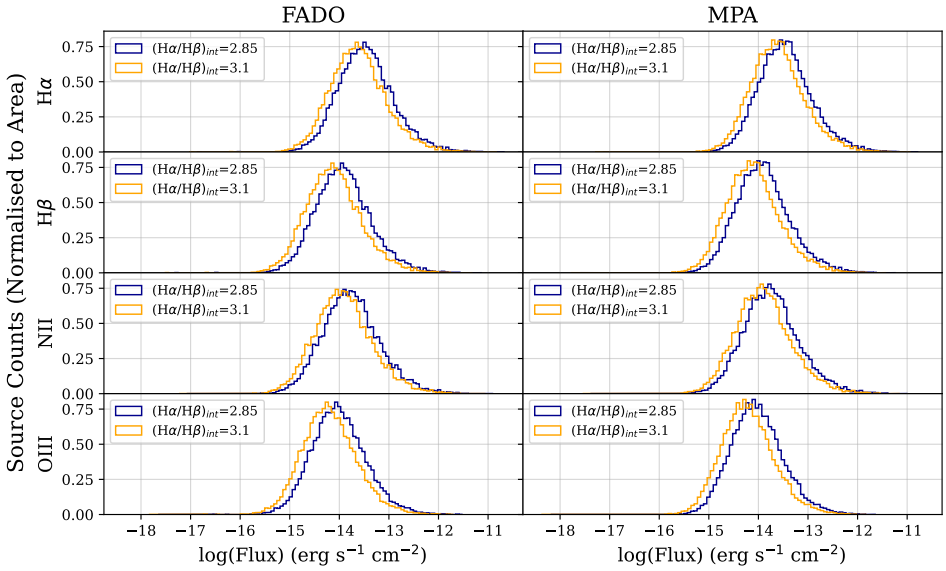

Figure 3.3: Histogram of the fluxes from FADO and MPA-JHU for the Composite galaxies defined in Table 3.3, for Hα, Hβ, [NII]λ6584 and [OIII]λ5007, corrected for line of sight extinction, taking into account two intrinsic Balmer decrements: 2.85 and 3.1.

Extinction can also be defined as a combination of two factors: the colour excess, $E(B-V)$, and the extinction law $k(\lambda)$, as such:

$$A(\lambda) = k(\lambda)E(B-V) \tag{3.5}$$

We can apply this thread of reasoning to the Balmer decrement, where we'll have that

$$A(\mathrm{H}\beta) - A(\mathrm{H}\alpha) = (k(\mathrm{H}\beta) - k(\mathrm{H}\alpha))\,E(B-V) \tag{3.6}$$

where $k(\mathrm{H}\alpha)$ and $k(\mathrm{H}\beta)$ represent the extinction law for Hα and Hβ, respectively. Joining together equations 3.6 and 3.4, we have the final expression for the colour excess:

$$E(B-V) = \frac{2.5}{k(\mathrm{H}\beta) - k(\mathrm{H}\alpha)} \log\left(\frac{(\mathrm{H}\alpha/\mathrm{H}\beta)_{\mathrm{obs}}}{(\mathrm{H}\alpha/\mathrm{H}\beta)_{\mathrm{int}}}\right) \tag{3.7}$$

Finally, all that's left is to define the extinction law. For this, we will use the Charlot and Fall (2000) extinction law, to be consistent with Brinchmann et al. (2004) procedures:

$$k(\lambda) = \left(\frac{\lambda}{5500}\right)^{-0.7} \tag{3.8}$$

where λ is in ångströms. It is of note that, although most of the MPA-JHU fitting procedure comes from Brinchmann et al. (2004), this dataset does not, in fact, correct for the line of sight extinction.

Table 3.7: Median differences (in dex) between FADO and MPA-JHU for each emission line flux in the BPT diagram, for the sample defined in Table 3.3.

	FADO−MPA-JHU	
Hα	Uncorrected	Corrected
SF	0.001	0.010
Composite	-0.011	-0.009
AGN	-0.018	-0.043
NII	Uncorrected	Corrected
SF	0.005	0.012
Composite	-0.023	-0.020
AGN	-0.022	-0.052

	FADO−MPA-JHU	
Hβ	Uncorrected	Corrected
SF	-0.001	0.010
Composite	-0.007	-0.009
AGN	-0.004	-0.043
OIII	Uncorrected	Corrected
SF	0.022	0.040
Composite	0.013	0.016
AGN	-0.008	-0.032

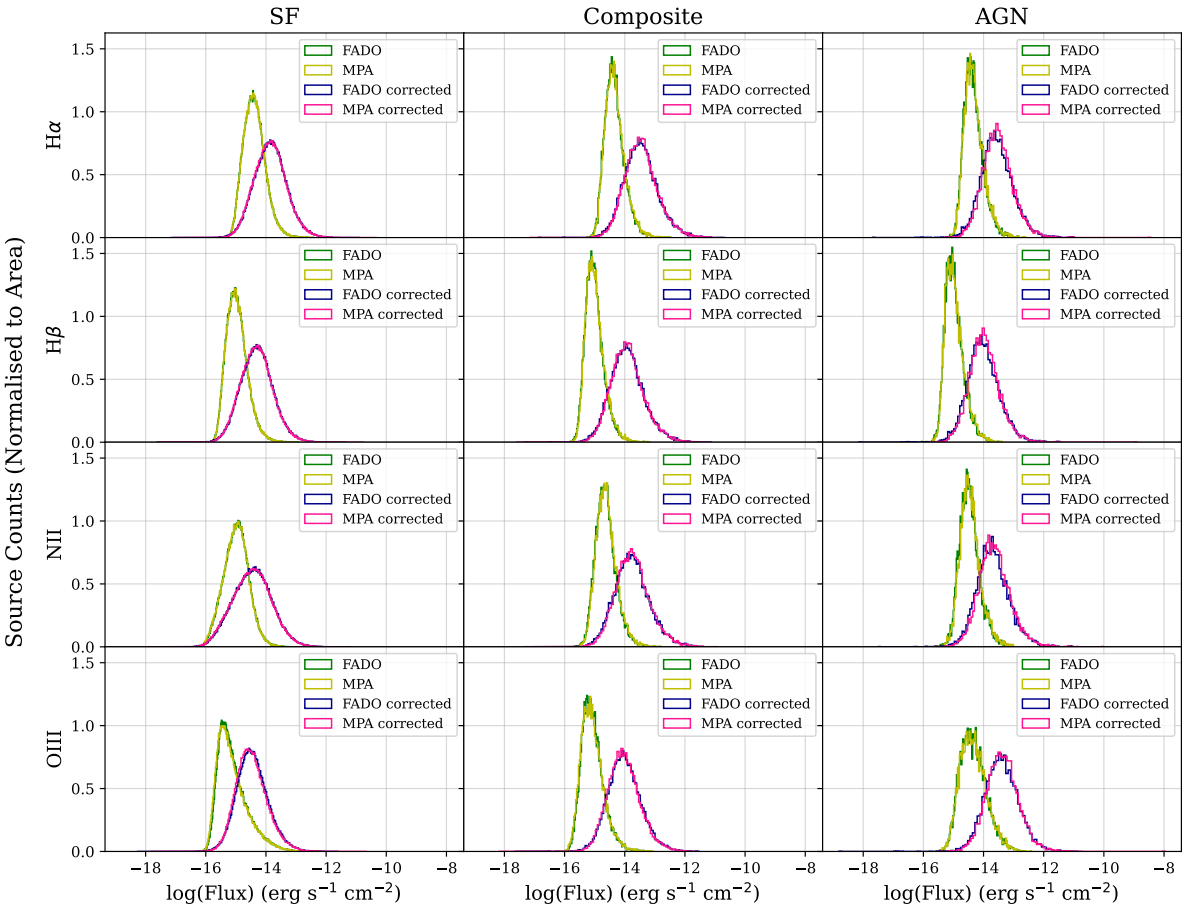

Figure 3.4: Histogram of the fluxes from the FADO and MPA-JHU datasets for the sample defined in Table 3.3. It contains the uncorrected and corrected for line of sight extinction fluxes.

To correct the emission line fluxes for extinction, the procedure is as follows:

1. Calculate the difference between the observed and intrinsic Balmer decrement as seen in equation 3.4;
2. Calculate the colour excess as seen in equation 3.7;
3. Apply the previous values and the extinction law (3.8) in equation 3.5;

4. Finally, calculate the flux in the absence of dust and gas, F_λ^0, with equation 3.3.

As we can see from Figure 3.4, when we correct the emission line fluxes of these galaxies we can see that the distribution broadens and the peak is shifted towards higher fluxes, which is what we expected. It is also of note that the differences between FADO and MPA-JHU remain mostly the same between the uncorrected and corrected fluxes, as can be seen in Table 3.7.

3.3 Plotting the BPT diagram

Now that we have our sample defined and the emission line fluxes corrected, we can finally plot the BPT diagram. We will start first by using each dataset's classification and, after explaining the differences that arise, we will move on to show a consistent method of classification of the galaxies, so we can better compare FADO and MPA-JHU.

3.3.1 FADO and MPA-JHU classification

With the sample defined in Table 3.3, we plotted the BPT diagram with the categories as defined by FADO and MPA-JHU in Figure 3.5.

From this Figure, we can see that FADO separates the galaxies according to their placement in the diagram, taking into account the following diagnostic lines (Kauffmann et al., 2003; Kewley et al., 2001; Kewley et al., 2006):

$$\log \frac{[OIII]\lambda 5007}{H\beta} = \frac{0.61}{\log \frac{[NII]\lambda 6583}{H\alpha} - 0.05} + 1.3 \tag{3.9}$$

$$\log \frac{[OIII]\lambda 5007}{H\beta} = \frac{0.61}{\log \frac{[NII]\lambda 6583}{H\alpha} - 0.47} + 1.19 \tag{3.10}$$

$$\log \frac{[OIII]\lambda 5007}{H\beta} = 2.1445 \log \frac{[NII]\lambda 6583}{H\alpha} + 0.465 \tag{3.11}$$

If the ratio of the flux of the emission lines is below equation 3.9, the galaxy is considered SF; if it is above equation 3.10, the galaxy is considered an AGN; if it is between equations 3.9 and 3.10, the galaxy is considered to be a Composite galaxy; if a galaxy is below equation 3.11 and above 3.10, it is considered a LINER-type galaxy and if it is above both previous equations it is considered a Seyfert-type galaxy. It is of note that 3.11 comes from Kauffmann et al. (2003), and is a line that has a 25 degree angle when compared with equation 3.9.

The MPA-JHU data follows a different classification scheme, explained in the previous Chapter - however, due to our cuts in the S/N and the flux, both datasets have galaxies inside the limits defined previously.

The contour plots between FADO and MPA-JHU also have slightly different shapes, but this can be explained by the number of galaxies present in each category, as MPA-JHU has overall more SF galaxies, while FADO has more Composites and AGNs.

3.3.2 Consistent classification

Although our sample of galaxies in both FADO and MPA-JHU fit well into the established limits of the BPT diagram, we can still see from Table 3.3 that galaxy counts among classifications are not exactly the same, with approximately a 2% difference. With this in mind, instead of using each dataset

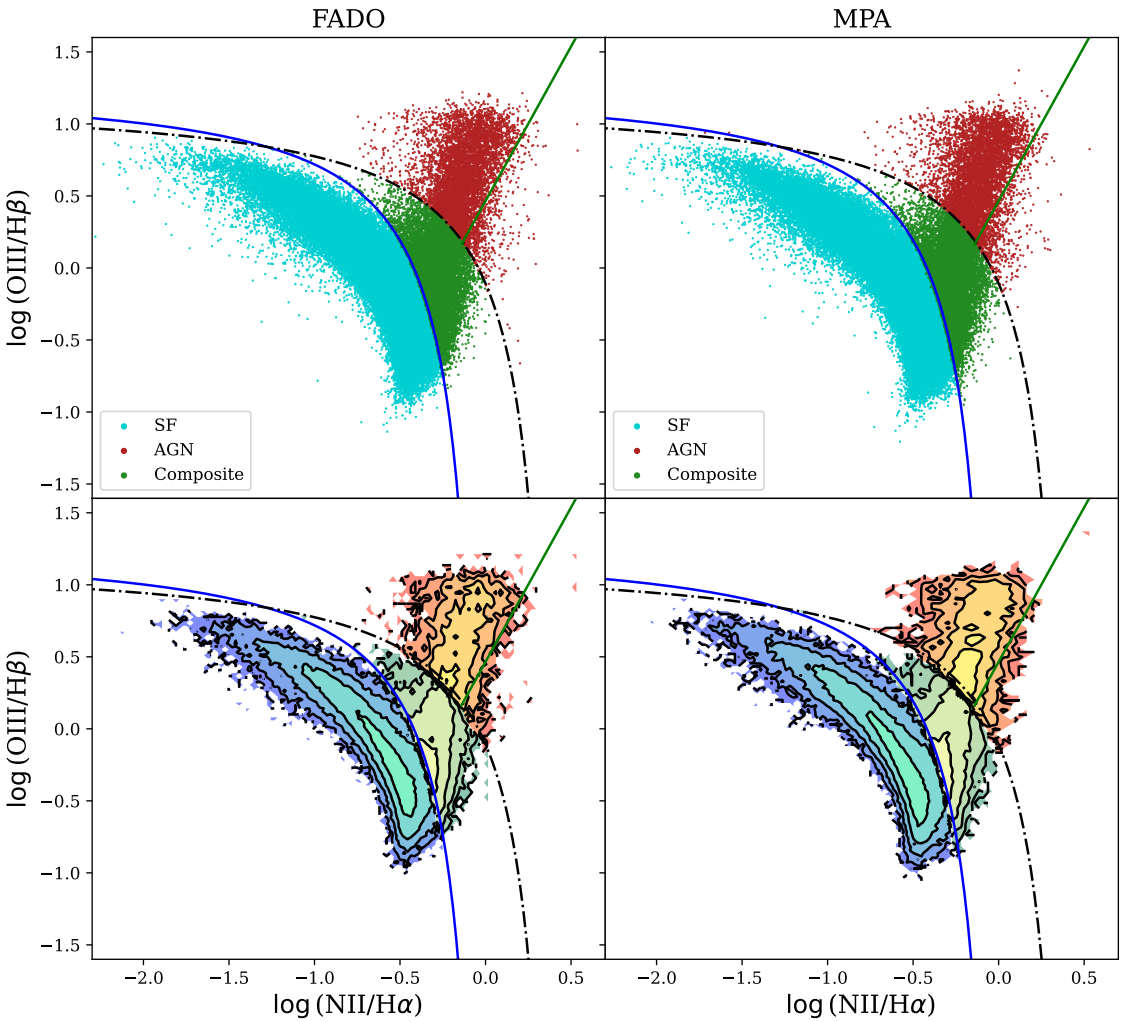

Figure 3.5: BPT diagram of the sample defined in Table 3.3 for both FADO and MPA-JHU, in the form of a scatter and contour plot. The blue dots represent the SF galaxies, the red dots the AGN galaxies and the green dots the Composite galaxies. Each contour represents 20% more of each sample in each category. Furthermore, the blue line represents equation 3.9, the black dot-dashed line represents equation 3.10 and the green line represents equation 3.11.

classification, we enforced the BPT diagram classification scheme (given by equations 3.9 and 3.10) to both datasets simultaneously, to understand the origin of these differences.

From Table 3.8 we can see that neither FADO nor MPA-JHU suffer any major differences when we enforce the classification through the diagnostic lines mentioned previously. This tracks, as they already adhered to the BPT diagram diagnostic lines faithfully. However, despite the fact that the total number of galaxies is the same between FADO and MPA-JHU, the differences in galaxy counts for each classification remain. FADO has more Composite galaxies and AGNs than MPA-JHU, but MPA-JHU has more SF galaxies than FADO.

We filtered these galaxies so that in each dataset we would have the same galaxies with the same classification according to the diagnostic lines from equations 3.9 and 3.10. This gives rise to a total of 153 760 galaxies in common where the classification match, or about 98% of the sample we defined in Table 3.8. The remaining 2% (which surmises a total of 2 987 galaxies) are where the classification does not match in both datasets, which tells us where the differences in Table 3.3 come from. This information is summarised in Table 3.9.

Table 3.8: Galaxy classification according to the FADO and MPA-JHU classification (represented by the name of each dataset) and classification as defined by the diagnostic lines from equations 3.9 and 3.10 (here presented as DL, representing Diagnostic Lines). This takes into account limitations in the S/N and flux (S/N > 3 and $10^{-18} < F_\lambda\,[\text{erg s}^{-1}\text{cm}^{-2}] < 10^{-12}$) for the H$\alpha$, H$\beta$, [NII]$\lambda$6584 and [OIII]$\lambda$5007 emission lines.

Criteria	FADO	DL
SF	133 595	133 594
Composite	16 428	16 429
AGN	6 726	6 725
Unclassified	1	-
Total	156 750	156 748

Criteria	MPA	DL
SF	135 345	135 345
Composite	14 891	14 890
AGN	6 514	6 513
Unclassified	0	-
Total	156 750	156 748

Table 3.9: Galaxy classification of FADO and MPA-JHU according to the diagnostic lines mentioned in equations 3.9 and 3.10. This represents the galaxies from both datasets that have the same classification and the ones where the classification does not match, according to the methods explained. There are 153 760 galaxies with the same classification and 2 987 with different classification. This takes into account limitations in the S/N and flux (S/N > 3 and $10^{-18} < F_\lambda\,[\text{erg s}^{-1}\text{cm}^{-2}] < 10^{-12}$) for the H$\alpha$, H$\beta$, [NII]$\lambda$6584 and [OIII]$\lambda$5007 emission lines.

		MPA-JHU		
	Criteria	SF	Composite	AGN
FADO	SF	133 206	381	7
	Composite	2 123	14 177	128
	AGN	16	332	6 377

3.4 Analysis of the BPT diagram

Before we move on to analyse the galaxies with the same classification, we must first understand the galaxies where the classification is mismatched, that is, where the classification in FADO is different from MPA-JHU according to the diagnostic lines.

3.4.1 Galaxies with mismatched classification

Before looking for the placement of these mismatched galaxies in the BPT diagram, we first calculated the uncertainties associated with the ratio of the emission line fluxes. We propagate the flux ratio errors using equation 3.12 and calculated the median error for these galaxies, which can be seen in Table 3.10.

$$\begin{aligned}\sigma\left[\log\left(\frac{F_{\lambda_1}}{F_{\lambda_2}}\right)\right] &= \left|\frac{\partial\left[\log\left(\frac{F_{\lambda_1}}{F_{\lambda_2}}\right)\right]}{\partial F_{\lambda_1}}\right|\sigma_{F_{\lambda_1}} + \left|\frac{\partial\left[\log\left(\frac{F_{\lambda_1}}{F_{\lambda_2}}\right)\right]}{\partial F_{\lambda_2}}\right|\sigma_{F_{\lambda_2}} = \\ &= \left|\frac{1}{\ln(10)F_{\lambda_1}}\right|\sigma_{F_{\lambda_1}} + \left|-\frac{1}{\ln(10)F_{\lambda_2}}\right|\sigma_{F_{\lambda_2}}\end{aligned} \tag{3.12}$$

Table 3.10: Median error (in dex) of the FADO and MPA-JHU BPT diagrams for the 2 987 mismatched galaxies.

FADO		MPA-JHU	
$\log(\text{NII}/\text{H}\alpha)$ (dex)	$\log(\text{OIII}/\text{H}\beta)$ (dex)	$\log(\text{NII}/\text{H}\alpha)$ (dex)	$\log(\text{OIII}/\text{H}\beta)$ (dex)
0.064	0.177	0.041	0.095

With this, we plotted the diagram for these mismatched galaxies (Figure 3.6), and found that, taking the error bars into account, 2 963 galaxies in FADO (99% of the mismatched galaxies) and 2 933 galaxies

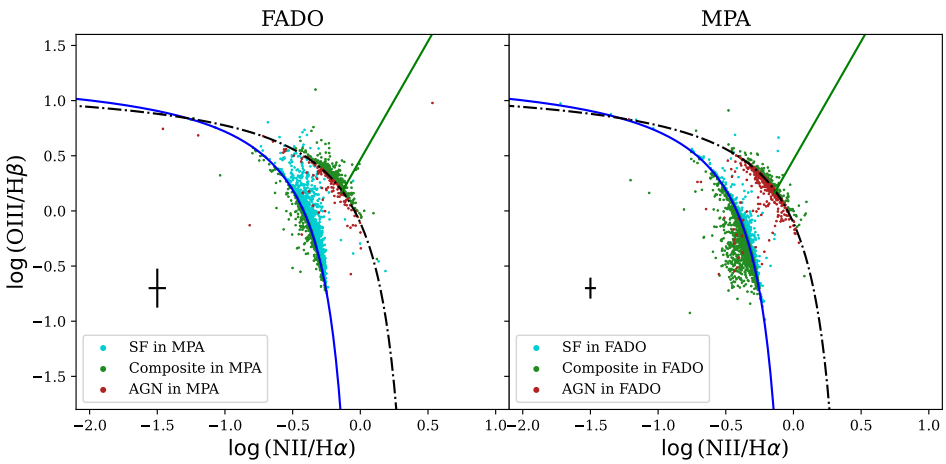

Figure 3.6: BPT diagram of the 2 987 mismatched galaxies, for both FADO and MPA-JHU. In each diagram, we have that the colours represents galaxies from the classification of the other dataset, i.e.: in the FADO BPT diagram, the blue dots represent the SF galaxies in MPA-JHU. The black error bars represent the median errors for each dataset.

in MPA-JHU (98% of the mismatched galaxies), could belong to more than one classification, as they lie in the border. As such, we can choose to not use these mismatched galaxies in the next steps of our analysis, as the error bars account for most of the differences between the galaxies. As for the remaining galaxies, the ones where the error bars do not compensate for the different classifications, they are certainly interesting objects of study. However, since we cannot really discriminate if the differences arise from the different extinction corrections of each dataset, the different methods of flux calculation, or if one has a correct classification and the other does not, then we discarded them from our analysis as well.

3.4.2 Galaxies with the same classification

For the 153 760 galaxies that have the same classification in FADO and MPA-JHU, we proceeded to plot the BPT diagram, adding the uncertainties as dictated by equation 3.12 for each category (SF, Composite and AGN), as well as the overall errors for all galaxies. We can see this in Figure 3.7, but to better understand the differences between these two datasets, comparing their countour plots is essential, as from them we can understand if the shift of the galaxies between FADO and MPA-JHU is significant or not.

Table 3.11: Median differences (in dex) between the FADO and MPA-JHU BPT diagrams for the 153 760 galaxies with the same classification.

Criteria	FADO − MPA-JHU	
	$\log(\mathrm{NII}/\mathrm{H}\alpha)$ (dex)	$\log(\mathrm{OIII}/\mathrm{H}\beta)$ (dex)
SF	0.008	0.018
Composite	-0.001	0.043
AGN	-0.011	0.013

The countour plots are compared in Figure 3.8, as well as the median differences between FADO and MPA-JHU that are in Table 3.11, with its associated errors in Table 3.12. The differences between them are small, with the highest ones being about 10%, which isn't very relevant, especially if we take into account the error bars represented in Figure 3.7, as the galaxies have essentially the same distribution for

Table 3.12: Median error (in dex) of the FADO and MPA-JHU BPT diagrams for the 153 760 galaxies with the same classification.

Criteria	FADO		MPA-JHU	
	$\log(\mathrm{NII}/\mathrm{H}\alpha)$ (dex)	$\log(\mathrm{OIII}/\mathrm{H}\beta)$ (dex)	$\log(\mathrm{NII}/\mathrm{H}\alpha)$ (dex)	$\log(\mathrm{OIII}/\mathrm{H}\beta)$ (dex)
SF	0.041	0.122	0.040	0.067
Composite	0.060	0.160	0.035	0.076
AGN	0.062	0.118	0.032	0.049
All	0.044	0.126	0.039	0.067

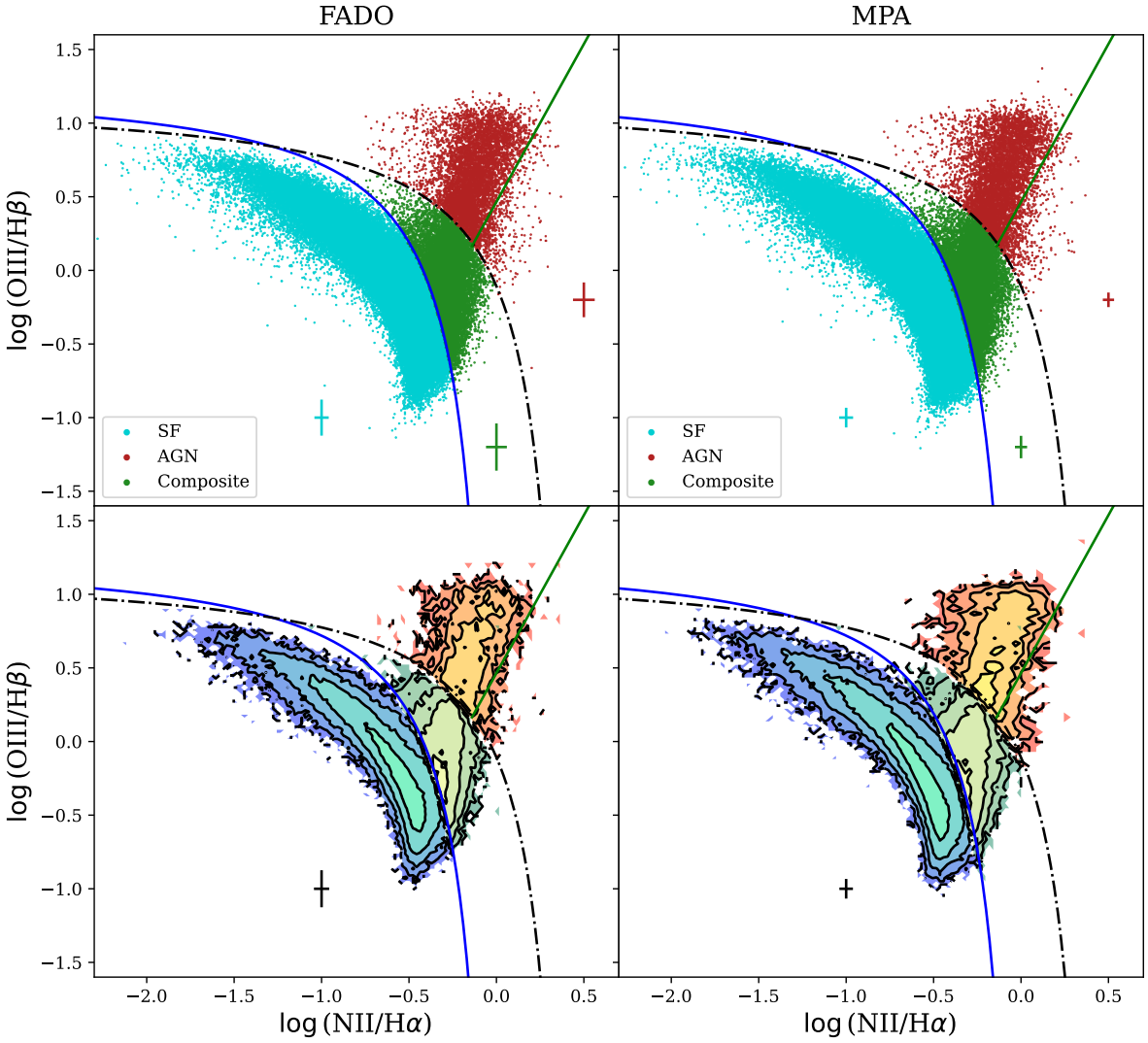

Figure 3.7: BPT diagram of 153 760 galaxies that have the same classification in FADO and MPA-JHU. The blue error bars represent the uncertainties with the SF galaxies, the green with the Composite galaxies and the red with the AGNs. The black error bars represent the overall error for all the galaxies represented. The remaining elements represent the same as explained in Figure 3.5.

each category and each dataset.

This information is meaningful, as it tells us that the the self-consistent inclusion of nebular emission in FADO does not affect galaxy classification through the standard BPT diagram. This, in turn, means that to understand whether a galaxy has an AGN or is SF, we can base ourselves on simpler (or purely stellar) models.

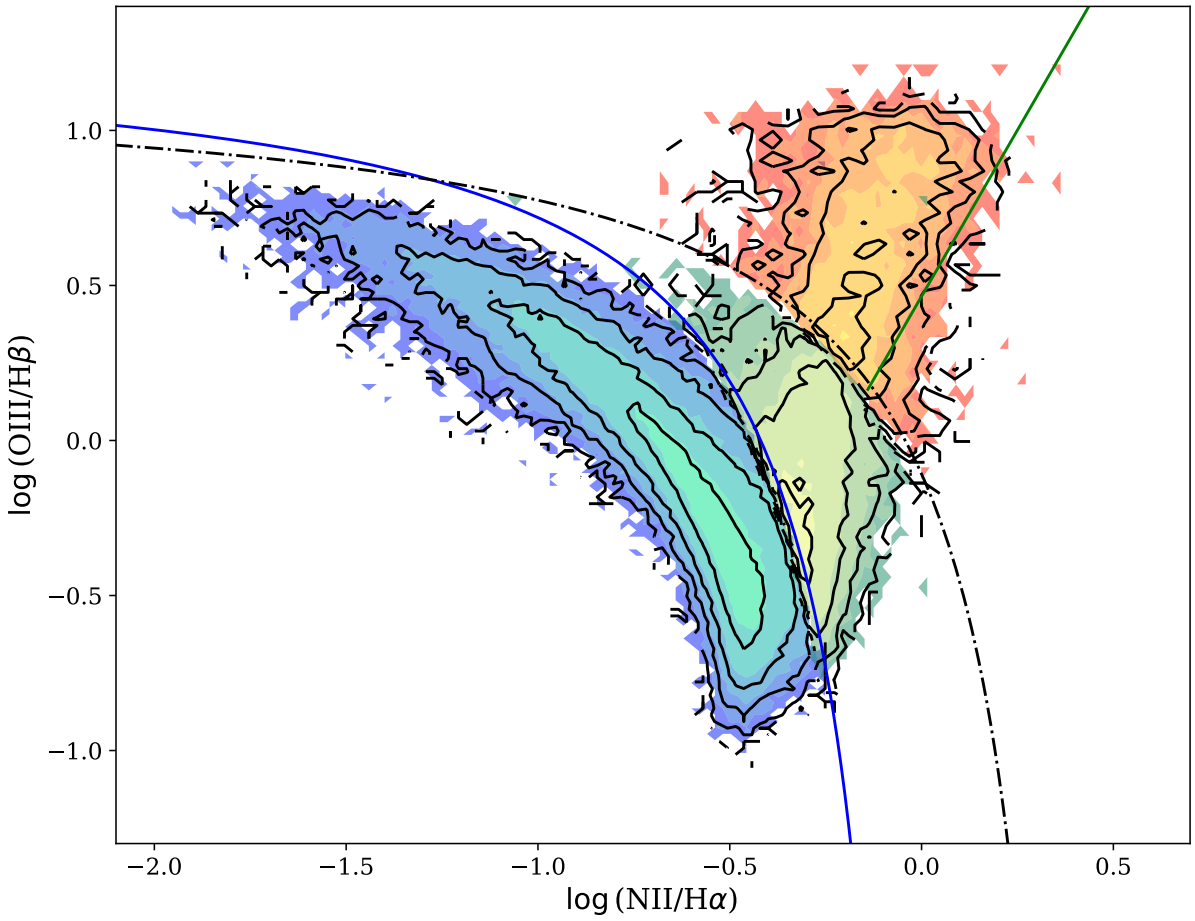

Figure 3.8: BPT diagram of 153 760 galaxies that have the same classification in FADO and MPA-JHU. Here, the countours with colours represent FADO and the black lines represent MPA-JHU. The remaining elements represent the same as in Figure 3.5.

Chapter 4

Other classification schemes with FADO and MPA-JHU

In this chapter, we will analyse the sample defined previously with more emission line flux ratios diagrams - notably, the $[\mathrm{SII}]\lambda\lambda 6717,6731/\mathrm{H}\alpha$ vs $[\mathrm{OIII}]\lambda 5007/\mathrm{H}\beta$ and $[\mathrm{OI}]\lambda 6300/\mathrm{H}\alpha$ vs $[\mathrm{OIII}]\lambda 5007/\mathrm{H}\beta$ (or, respectively, the SII and OI diagrams) - and comparing them to the BPT diagram, to understand whether or not these classifications are consistent.

4.1 Sample selection

To compare all three diagrams between FADO and MPA-JHU, we must first select galaxies that are common in both datasets and, as such, we are going to follow the same methodology of the previous chapter's sample selection. Firstly, we are going to restrict our sample to galaxies that have:

a) S/N of each emission line ($\mathrm{H}\alpha$, $\mathrm{H}\beta$, $[\mathrm{NII}]\lambda 6584$, $[\mathrm{OIII}]\lambda 5007$ and $[\mathrm{SII}]\lambda\lambda 6717,6731$) higher than 3;

b) Flux from all emission lines mentioned above between $10^{-18} < F_\lambda\,[\mathrm{erg\ s^{-1}cm^{-2}}] < 10^{-12}$.

Although we could cut the S/N and flux of $[\mathrm{OI}]\lambda 6300$, the differences between the inclusion and exclusion of this restriction are quite severe (as can be seen by the S/N in Figure 4.1). Moreover, this emission line can be contaminated by sky lines. As such, we have decided to follow a similar methodology as Kewley et al. (2006), where we applied the previously mentioned criteria to all emission lines except for $[\mathrm{OI}]\lambda 6300$.

Table 4.1: *Left:* Median differences (in dex) between FADO and MPA-JHU in emission line flux and S/N for the 916 118 galaxies in the common sample. *Right:* Median uncertainties associated to each emission line for FADO and MPA-JHU, in dex.

Emission line	FADO − MPA-JHU Flux (dex)	FADO − MPA-JHU S/N (dex)
$[\mathrm{SII}]\lambda 6717$	0.035	-0.229
$[\mathrm{SII}]\lambda 6731$	0.072	-0.229
$[\mathrm{OI}]\lambda 6300$	0.307	-0.240

Emission line	FADO σ_{F_λ} (dex)	MPA-JHU σ_{F_λ} (dex)
$[\mathrm{SII}]\lambda 6717$	0.243	0.090
$[\mathrm{SII}]\lambda 6731$	0.308	0.107
$[\mathrm{OI}]\lambda 6300$	0.479	0.212

Since we already analysed the differences between the BPT diagram emission lines, we can focus on the differences of $[\mathrm{SII}]\lambda\lambda 6717,6731$ and $[\mathrm{OI}]\lambda 6300$ between FADO and MPA-JHU. From Table 4.1,

the flux differences between the doublet of the SII lines are small ($\approx 8\%$ for [SII]λ6717 and $\approx 18\%$ for [SII]λ6731), with stark differences for [OI]λ6300 ($\approx 103\%$). Because [OI]λ6300 is an intrinsically weak line, then extracting this flux is going to be harder for these models. MPA-JHU usually has higher S/N ratios than FADO, but even this model struggles to have values higher than 3. The SII doublet, however, are stronger emission lines and, as such, have both higher fluxes and S/Ns. The differences in S/N of each emission line are high ($\approx -70\%$ for all emission lines) due to the different methods of calculation of the uncertainty in each dataset, as explained previously. As seen previously for the BPT diagram emission line fluxes, the uncertainties present in these emission lines also compensate for the differences between datasets, meaning that the fluxes overall match.

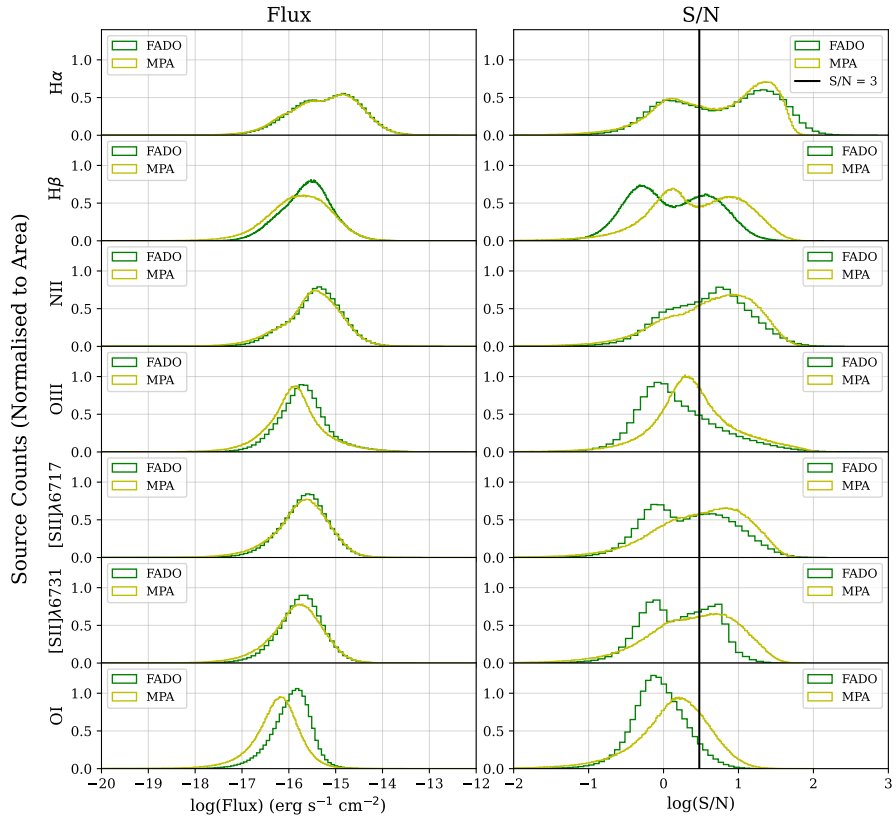

Figure 4.1: Histogram of the fluxes and S/N of the FADO and MPA-JHU datasets for the Hα, Hβ, [NII]λ6584, [OIII]λ5007, [SII]$\lambda\lambda$6717, 6731 and [OI]λ6300 emission lines, in the common sample of 916 118 galaxies. The black vertical line represents when S/N $= 3$.

Table 4.2: Galaxy classification according to the categories defined in this work, taking into account the limitations in the S/N and flux (S/N > 3 and $10^{-18} < F_\lambda\,[\text{erg s}^{-1}\text{cm}^{-2}] < 10^{-12}$) for the H$\alpha$, H$\beta$, [NII]$\lambda$6584, [OIII]$\lambda$5007 and [SII]$\lambda\lambda$6717, 6731 emission lines, in both FADO and MPA-JHU simultaneously. It has a total of 143 179 galaxies.

	FADO		MPA-JHU	
Criteria	Galaxy #	Percentage	Galaxy #	Percentage
SF	123 110	86	124 567	87
Composite	13 978	10	12 721	9
AGN	6 090	4	5 891	4
Unclassifiable	1	< 1	0	0
Total	143 179	-	143 179	-

Applying the previously mentioned criteria to both datasets simultaneously, we get a total sample of 143 179 galaxies (or $\approx 16\%$ of the common sample), whose distribution can be seen in Table 4.2

and Figure 4.2, with the simplified categories as defined in Chapter 3. As expected, although the galaxy counts between classification differ slightly, the highest differences are of 1%, meaning that FADO and MPA-JHU are mostly in agreement. Once again, since we have already analysed the differences of the BPT diagram emission lines, we are going to focus on explaining the differences between the SII doublet and the [OI]λ6300 emission lines between FADO and MPA-JHU. From Table 4.3, we can see that the already low flux differences for the common sample in the SII doublet become minute in this restricted sample ($\approx 1\%$ for [SII]λ6717 and $\approx 2\%$ for [SII]λ6731) and the same applies to [OI]λ6300, albeit with a higher difference ($\approx 34\%$), no doubt due to the fact that we have not restricted its S/N. When it comes to the S/N, the differences remain high ($\approx -34\%$ for [SII]λ6717 and $\approx -100\%$ for [SII]λ6731 and [OI]λ6300), for reasons already discussed.

It is also of note that emission line fluxes of these three emission lines were also corrected for line-of-sight extinction with the same methodology as described in section 3.2.

Table 4.3: Median differences (in dex) between FADO and MPA-JHU in emission line flux and S/N for the sample defined in Table 4.2.

Emission line	FADO − MPA-JHU	
	Flux (dex)	S/N (dex)
[SII]λ6717	0.006	-0.128
[SII]λ6731	0.010	-0.310
[OI]λ6300	0.127	-0.299

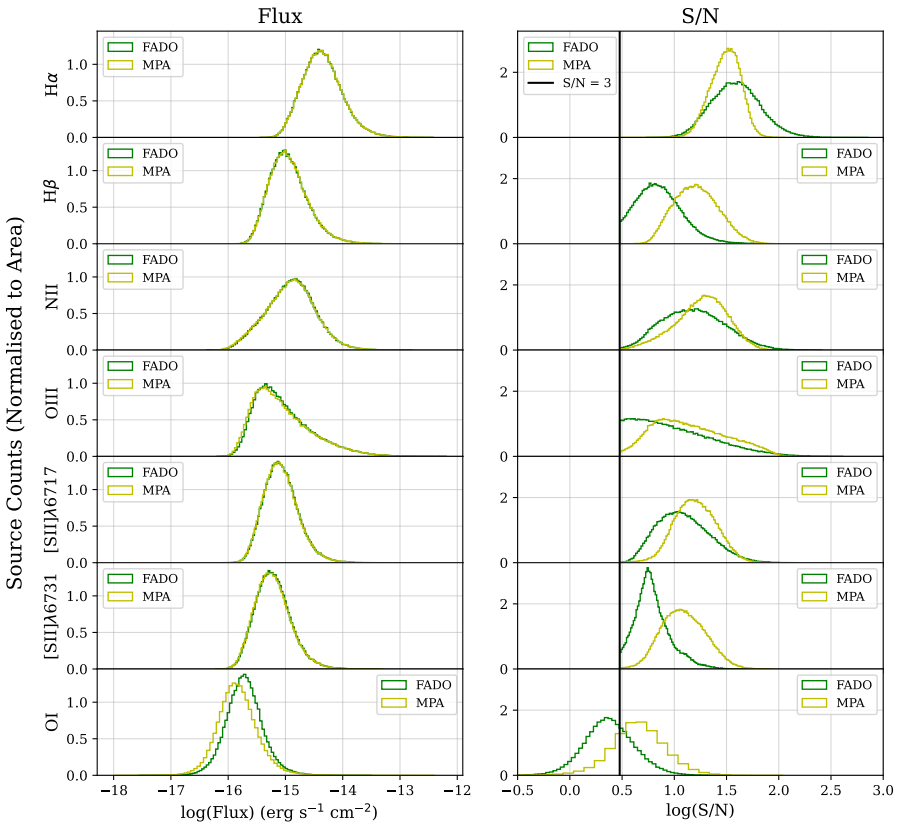

Figure 4.2: Histogram of the fluxes and S/N of the FADO and MPA-JHU datasets for the Hα, Hβ, [NII]λ6584, [OIII]λ5007, [SII]$\lambda\lambda$6717, 6731 and [OI]λ6300 emission lines, in the sample defined in Table 4.2. The black vertical line represents when S/N = 3.

This sample, however, considers each dataset's classification, and not a consistent one, as we have done in section 3.3.2. As such, we must first apply the same reasoning as we have done before, enforcing the diagnostic lines of the BPT diagram (equations 3.9 and 3.10) on our sample. This changes the galaxy distribution as can be seen in Table 4.4 and, as previously, neither FADO nor MPA-JHU suffer significant alterations to galaxy classification, with the total galaxy counts suffering minute changes.

Table 4.4: Galaxy classification according to the FADO and MPA-JHU classification (represented by the name of each dataset) and the BPT classification as defined by the diagnostic lines from equations 3.9 and3.10 (here presented as DL, representing Diagnostic Lines). This takes into account limitations in the S/N and flux (S/N > 3 and $10^{-18} < F_\lambda\,[\text{erg s}^{-1}\text{cm}^{-2}] < 10^{-12}$) for the H$\alpha$, H$\beta$, [NII]$\lambda$6584, [OIII]$\lambda$5007 and [SII]$\lambda\lambda$6717,6731 emission lines. The total number of galaxies in common across both datasets is 143 177 galaxies.

Criteria	FADO	DL
SF	123 110	123 110
Composite	13 978	13 978
AGN	6 090	6 090
Unclassified	1	-
Total	143 179	143 178

Criteria	MPA-JHU	DL
SF	124 567	124 567
Composite	12 721	12 720
AGN	5 891	5 890
Unclassified	0	-
Total	143 179	143 177

Finally, we now must understand from these galaxies which ones have a matching classification. As we have seen in the previous chapter, the galaxies with mismatched classification lie mostly in the border between classifications in the BPT diagram and, as such, they will be disregarded from our analysis. This gives us the distribution seen in Table 4.5, where 140 758 galaxies have classifications that match ($\approx 98\%$ of the sample defined in Table 4.4) and 2 419 galaxies where the classification does not match ($\approx 2\%$ of the sample defined in Table 4.4). The galaxies with matching classification is the sample we will be using to analyse all three diagrams.

Table 4.5: Galaxy classification of FADO and MPA-JHU according to the diagnostic lines mentioned in equations 3.9 and 3.10. This represents the galaxies from both datasets that have the same classification and the ones where the classification does not match. There are 140 758 galaxies with the same classification and 2 419 with different classification. This takes into account limitations in the S/N and flux (S/N > 3 and $10^{-18} < F_\lambda\,[\text{erg s}^{-1}\text{cm}^{-2}] < 10^{-12}$) for H$\alpha$, H$\beta$, [NII]$\lambda$6584, [OIII]$\lambda$5007 and [SII]$\lambda\lambda$6717,6731.

		MPA-JHU		
	Criteria	SF	Composite	AGN
FADO	SF	122 812	293	5
	Composite	1 744	12 147	86
	AGN	11	280	5 799

4.2 Plotting the diagrams

Before plotting the SII and OI diagrams, we must define the diagnostic lines that distinguish the several types of galaxies. From the equations mentioned below, neither the SII nor the OI diagram have a line that separates SF and AGN from Composites. As explained in Chapter 1, this is due to the fact that the [NII]λ6584/Hα emission line ratio is more sensitive to metallicity changes and can saturate (which is why the SF limit is nearly vertical at [NII]λ6584/H$\alpha \sim -0.5$). After this saturation point, another mechanism must contribute to the shift of this emission line ratio, and a possible explanation for this is AGN activity (Kauffmann et al., 2003; Kewley et al., 2006). The [SII]$\lambda\lambda$6717,6731/Hα or [OI]λ6300/Hα emission line ratios do not possess this sensitivity for the metallicity in SF galaxies and, as can be seen by Figure 4.3, the Composite galaxies (which are defined as SF with possible AGN

activity) are spread through both the SF and AGN categories, with no clear separation from each of them, unlike the BPT diagram.

We define two diagnostic lines for the SII diagram (Kewley et al., 2001; Kewley et al., 2006):

$$\log \frac{[OIII]\lambda 5007}{H\beta} = \frac{0.72}{\log \frac{[SII]\lambda\lambda 6717,6731}{H\alpha} - 0.32} + 1.30 \tag{4.1}$$

$$\log \frac{[OIII]\lambda 5007}{H\beta} = 1.89 \log \frac{[SII]\lambda\lambda 6717,6731}{H\alpha} + 0.76 \tag{4.2}$$

If the ratio of the flux of the emission lines of a galaxy is below equation 4.1, then the galaxy should be considered SF; if it is below equation 4.2 and above 4.1, the galaxy is a LINER-type and if it is above both, then it is a Seyfert-type galaxy.

For the OI diagram, we once again define two diagnostic lines (Kewley et al., 2001; Kewley et al., 2006):

$$\log \frac{[OIII]\lambda 5007}{H\beta} = \frac{0.73}{\log \frac{[OI]\lambda 6300}{H\alpha} + 0.59} + 1.33 \tag{4.3}$$

$$\log \frac{[OIII]\lambda 5007}{H\beta} = 1.18 \log \frac{[OI]\lambda 6300}{H\alpha} + 1.30 \tag{4.4}$$

If the ratio of the flux of the emission lines of a galaxy is below equation 4.3, then the galaxy should be considered SF; if it is below equation 4.4 and above 4.3, the galaxy is a LINER-type and if it is above both, then it is a Seyfert-type galaxy.

Thus, with the 140 758 galaxies with matching classifications in FADO and MPA-JHU through the BPT diagram, we get all three diagrams plotted in Figure 4.3.

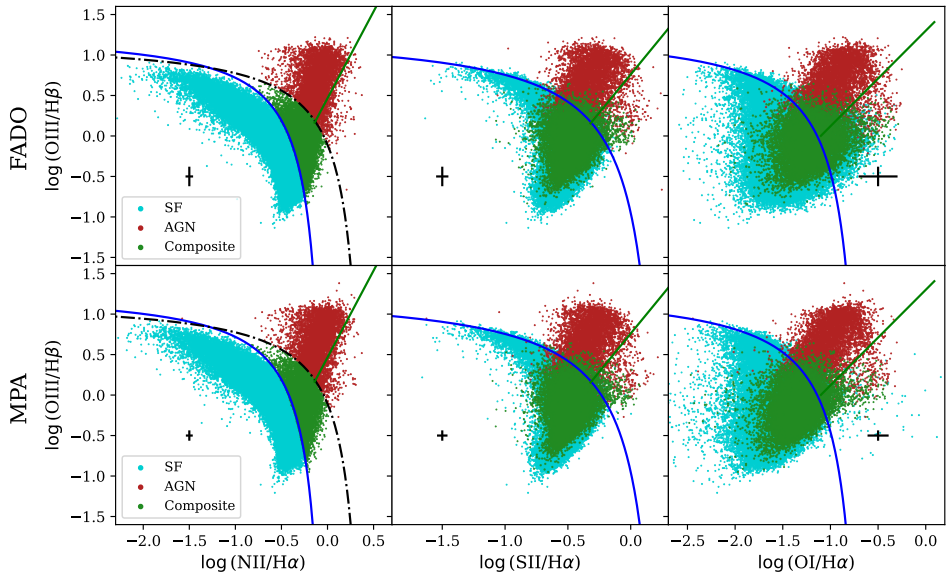

Figure 4.3: BPT (left), SII (centre) and OI (right) diagrams of the 140 758 galaxies defined in Table 4.5. In the BPT diagram, the blue, black dot-dashed and green lines the same as in Figure 3.5. In the SII diagram, the blue line represents equation 4.1 and the green line represents equation 4.2. In the OI diagram, the blue line represents equation 4.3 and the green line represents equation 4.4. The black error bars represent the median error over all the galaxies represented in each diagram. For all three diagrams, the blue dots represent the SF galaxies, green the Composite galaxies and red the AGNs, according to the BPT diagram classification.

As mentioned previously, despite the classifications in the BPT diagram being separated into three

distinct types, this does not happen in the SII or OI diagrams, where some of the SF and AGN galaxies bleed into each other's categories. In the SII diagram, $\approx 95\%$ of the Composite galaxies in both datasets lie below the SF line, while in the OI diagram they are more equally distributed across the SF and AGN regions - this can be seen in Figure 4.4 where $\approx 85\%$ (in MPA-JHU) and $\approx 65\%$ (in FADO) of the Composites are in the SF region, which is in agreement with Kewley et al. (2006).

Even with this contamination, the minimal differences between FADO and MPA-JHU in the BPT diagram, which where seen in the previous Chapter, also apply to the SII diagram, with the highest differences being $\approx 4\%$. When it comes to the OI diagram, the differences are higher ($\approx 37\%$), but this is a result of the fact that we did not cut neither the flux or the S/N of $[\mathrm{OI}]\lambda 6300$ (Table 4.6).

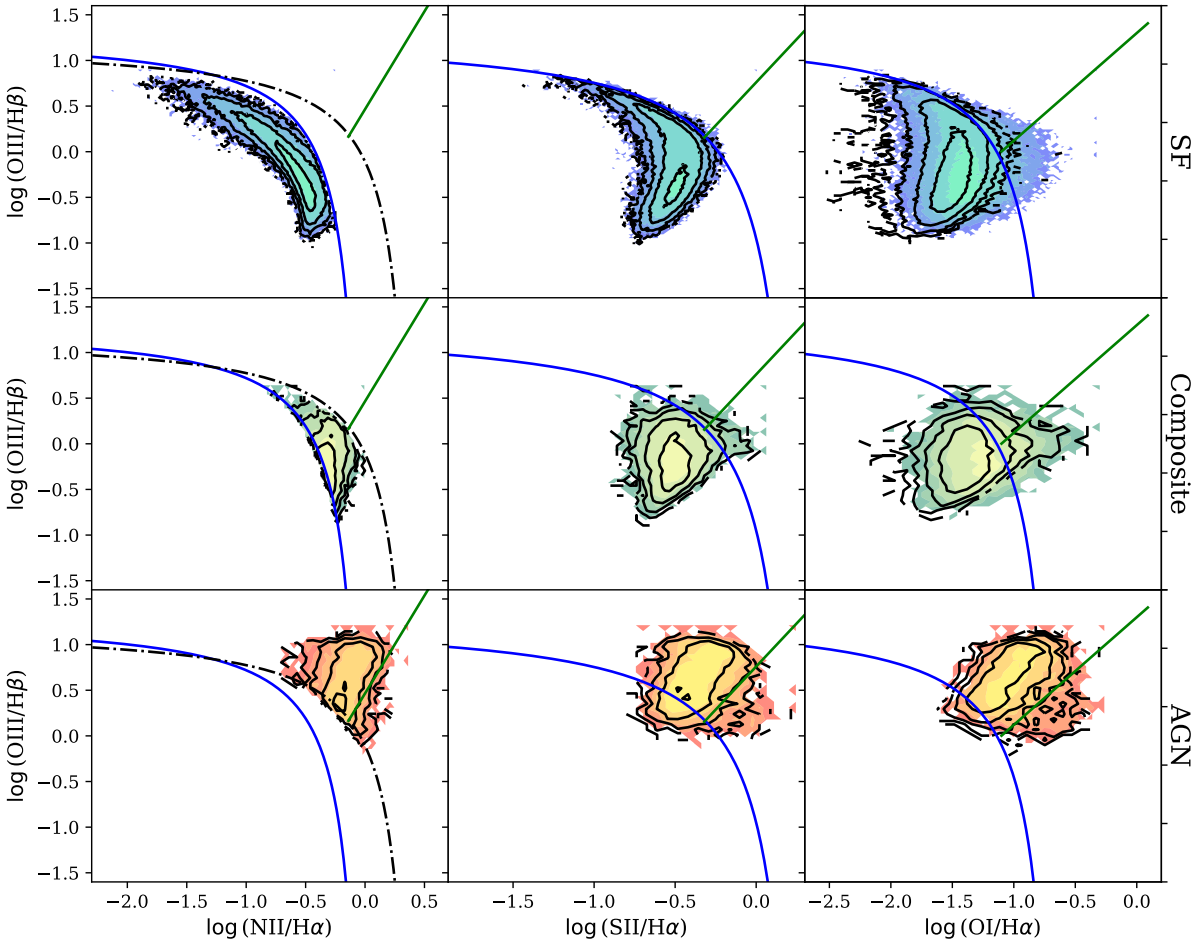

Figure 4.4: BPT (left), SII (centre) and OI (right) diagrams of the 140 758 galaxies defined in Table 4.5. The coloured contours represent the FADO data and the black lines the MPA-JHU data. The remaining elements are the same as Figure 4.3.

Table 4.6: Median differences (in dex) between the FADO and MPA-JHU emission line ratios of the BPT, SII and OI diagrams for the 140 758 galaxies with the same classification.

	FADO − MPA-JHU			
Criteria	$\log(\mathrm{NII}/\mathrm{H}\alpha)$	$\log(\mathrm{SII}/\mathrm{H}\alpha)$	$\log(\mathrm{OI}/\mathrm{H}\alpha)$	$\log(\mathrm{OIII}/\mathrm{H}\beta)$
SF	0.007	0.007	0.097	0.017
Composite	-0.001	0.007	0.136	0.004
AGN	-0.012	0.008	0.077	0.012

If galaxies are classified as one type by one diagram and a different type by another, what does this

mean? Galaxies with these properties were coined with the title of "Ambiguous", and are defined as such Kewley et al. (2006) if:

- They are considered SF in the BPT diagram and AGN in either the SII or OI diagram;
- They are considered AGN in the BPT diagram and SF in either the SII or OI diagram;
- They are considered Composite in the BPT diagram and AGN in either the SII or OI diagram;
- They are considered a LINER-type galaxy in the SII diagram and a Seyfert-type galaxy in the OI diagram, and vice versa.

We have also decided to adopt this method to classify galaxies as Ambiguous. Since individual emission line flux ratios of the galaxies are different in FADO and MPA-JHU, then it is not surprising that, as can be seen by Table 4.7, both these datasets have different counts of Ambiguous galaxies. FADO has 15 840 galaxies, or $\approx 11\%$ of the sample defined in Table 4.5, and MPA-JHU has 5 463, or $\approx 4\%$ of the same sample, which is in line with the results of Kewley et al. (2006).

Table 4.7: Number of galaxies classified as Ambiguous in FADO and MPA-JHU, as described previously in this work.

	FADO	MPA-JHU
BPT Classification	Galaxy #	
SF	11 762	3 144
Composite	3 473	1 521
AGN	605	798
Total	15 840	5 463

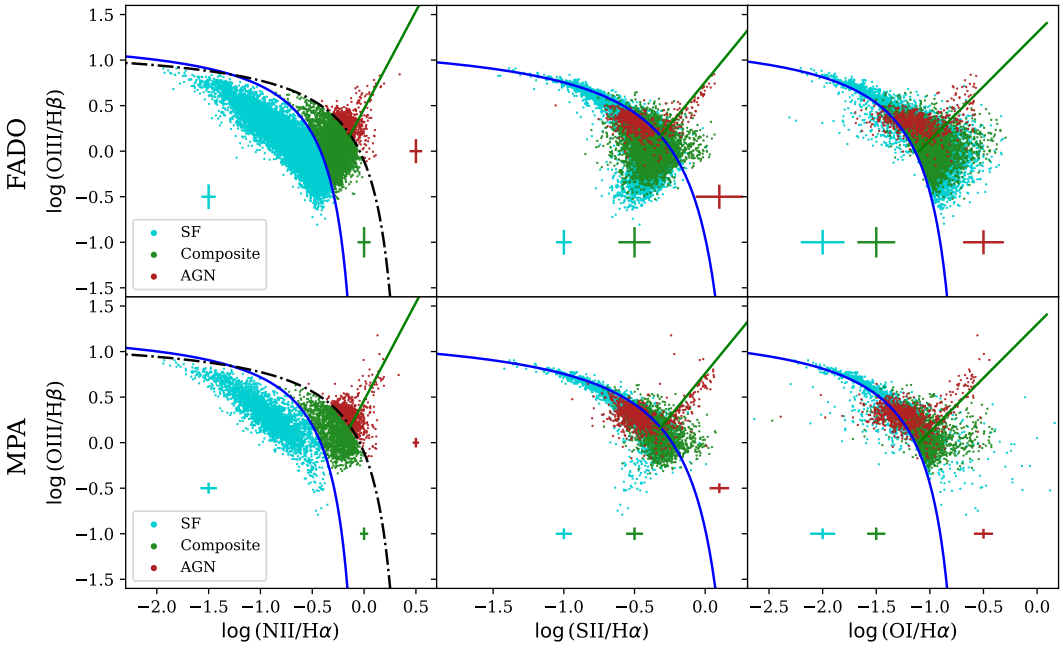

Figure 4.5: BPT (left), SII (centre) and OI (right) diagrams of the Ambiguous galaxies defined in Table 4.7. Each error bar corresponds to the median error of the colour associated with the galaxy classification. The remaining elements are the same as Figure 4.3.

4.3 Analysis of the three diagrams

Now that all of the diagrams have been plotted and the separation between categories explained, we will now proceed with the analysis of its two major categories: the Ambiguous and Non-Ambiguous galaxies.

4.3.1 Ambiguous galaxies

As defined in the previous section we have two sets of Ambiguous galaxies, one for each dataset. In Figure 4.5, we can see how these galaxies are distributed. The difference of galaxy counts between these datasets is due to the differences between the emission line fluxes: in the same galaxies, although the larger trends remain the same (as can be seen by Figure 4.3), the emission line flux ratios are different, meaning that we will have a different count of ambiguous galaxies between FADO and MPA-JHU.

In the OI diagram, we can see that most of these Ambiguous galaxies lie above the SF curve in FADO ($\approx 96\%$ of this dataset's Ambiguous sample) and MPA-JHU ($\approx 88\%$ of this dataset's Ambiguous sample), which does not happen in the SII diagram ($\approx 11\%$ of the FADO Ambiguous above the SF line and $\approx 22\%$ of the MPA-JHU Ambiguous galaxies above the SF line). To understand this effect, we investigated the metallicity and stellar mass of the FADO Ambiguous galaxies.

The gas-phase metallicity, or the metallicity of the interstellar medium, is taken from the MPA-JHU data. This data is estimated from simultaneous fits of prominent nebular emission lines (Tremonti et al., 2004). However, this model considers only the SF galaxies, as nebular emission lines that are measured in Composites and AGNs originate from different physical mechanisms besides nebular emission, meaning that it will no longer solely trace the metallicity. The stellar mass is discussed in more detail in Chapter 6, although it is of note that FADO and MPA-JHU calculate similar stellar masses for all galaxies, when taking into account its associated uncertainties. Also of note is that, from this point onward, we are classifying galaxies with low mass if they have stellar masses below $10^9\,M_\odot$, and with low metallicity if they have sub-solar metallicity (where the solar metallicity is $12+\log(O/H) = 8.69$).

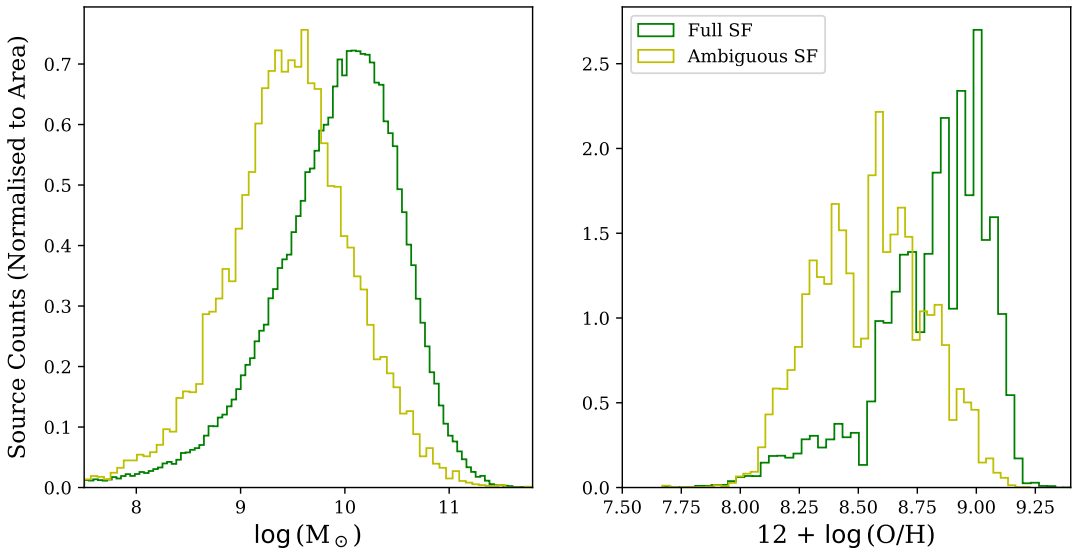

Figure 4.6: Histogram of the stellar mass (left) and metallicity (right) distributions of the full and Ambiguous SF sample. The median differences between the stellar masses is 0.509 dex and between metallicites is 0.291 dex.

After acquiring the metallicity and stellar mass of the SF sample, we compared the distribution of the full and Ambiguous SF sample of these properties in Figure 4.6. When it comes to the metallicity,

the Ambiguous SF galaxies have less metallicity than the full SF sample (a median $\approx 95\%$ difference), as well as having lower stellar masses, with a higher difference (median of $\approx 223\%$).

From the full SF sample, there are 32 236 with sub-solar metallicity galaxies, out of which 26% are Ambiguous. When it comes to stellar masses, from the 10 616 low mass galaxies, 23% are Ambiguous. Analogously, Ambiguous galaxies represent 4% of the solar and super-solar metallicities and 8% of galaxies with masses higher than $10^9\,M_\odot$. From these values, as well as the histograms mentioned, it is clear to see that there is a bias in the BPT diagram when classifying SF galaxies - sub-solar metallicity and low mass galaxies have a higher chance of being misidentified by other optical line emission ratio diagrams.

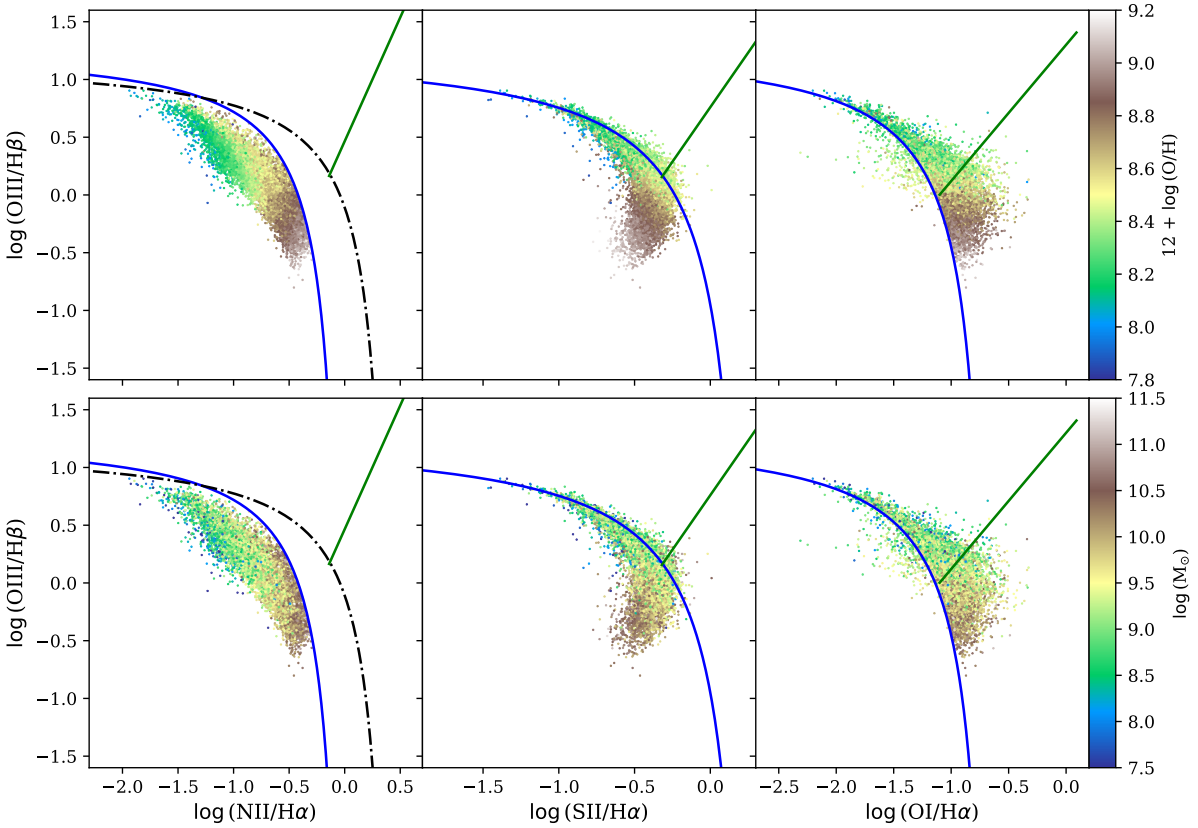

Figure 4.7: BPT (left), SII (middle) and OI (right) diagrams of the FADO Ambiguous SF galaxies, as defined by Table 4.7. The top row is colour-coded for the metallicity and the bottom row for the stellar mass. The remaining elements represent the same as Figure 4.3.

To illustrate this point, Figure 4.7 shows where these Ambiguous galaxies are placed, comparing them with the metallicity and stellar mass. In the SII diagram, of the 11 762 Ambiguous galaxies, the galaxies with sub-solar metallicity lie either in the AGN region or near the border between classifications. In the OI diagram, most of the SF galaxies already lie in the AGN region ($\approx 96\%$), meaning that the galaxies with sub-solar metallicity will be classified as AGN. This shows the metallicity sensitivity of these diagrams: the SII diagram is, like the BPT diagram, able to "trap" SF galaxies with higher metallicities in its SF region, but, unlike it, as this value lowers, they tend towards the AGN region. The OI diagram is essentially metallicity insensitive, as it is unable to distinguish between metal-rich and metal-poor SF galaxies, placing almost all of them in the AGN region with no discrimination, although it does have a vertical gradient in metallicity nonetheless.

When it comes to the stellar mass, the same properties are observed in both the SII and OI diagrams,

albeit with less rigidity than the metallicity, due to the fact that there are less low mass galaxies (21% of the Ambiguous SF sample) than sub-solar metallicity galaxies (71% of the same sample). This means that these Ambiguous galaxies are more correlated with differences in metallicity than in stellar mass, although the latter still plays an effect in classification.

To further test this BPT bias with sub-solar metallicity and low mass galaxies, we have decided to compare the Ambiguous SF galaxies with the Wide-field Infrared Survey Explorer, WISE (Wright et al., 2010), data in the mid-infrared, where the satellite observes the sky in four bands: W1 (3.4 μm), W2 (4.6 μm), W3 (12 μm) and W4 (22 μm). We do this because the mid-infrared gives us the properties of dust in a galaxy and, since most of the optical emission comes from stars, we should be able to see these galaxies in a different light, to better understand if they are truly AGN or SF. Specifically, we made use of the AllWISE data (Cutri et al., 2021), which covered 747 million sources across the entire sky, covering from the solar neighbourhood up until luminous galaxies at redshifts higher than 1.

We cross correlated the spatial coordinates of the Ambiguous SF sample with the AllWISE data, and, from the 11 742 galaxies, 10 090 matched. With the W1, W2 and W3 bands, we can generate a colour-colour diagram, which has a polygonal region that identifies AGNs, delineated by equation 4.5 (Bond et al., 2012).

$$\begin{aligned} \mathrm{W2}-\mathrm{W3} &> 2.2 \\ \mathrm{W2}-\mathrm{W3} &< 4.2 \\ \mathrm{W1}-\mathrm{W2} &< 1.7 \\ \mathrm{W1}-\mathrm{W2} &> 0.1(\mathrm{W2}-\mathrm{W3})+0.38 \end{aligned} \tag{4.5}$$

Figure 4.8 shows that, from the 10 090 galaxies that WISE could detect, only 46 (less than 1%) are classified as AGN with the equation defined above. Although interesting objects of study, we have decided to discard them from our analysis due to their low counts. This means that, for the most part, the WISE classification is in agreement with the BPT classification, which means it might also inherit its biases.

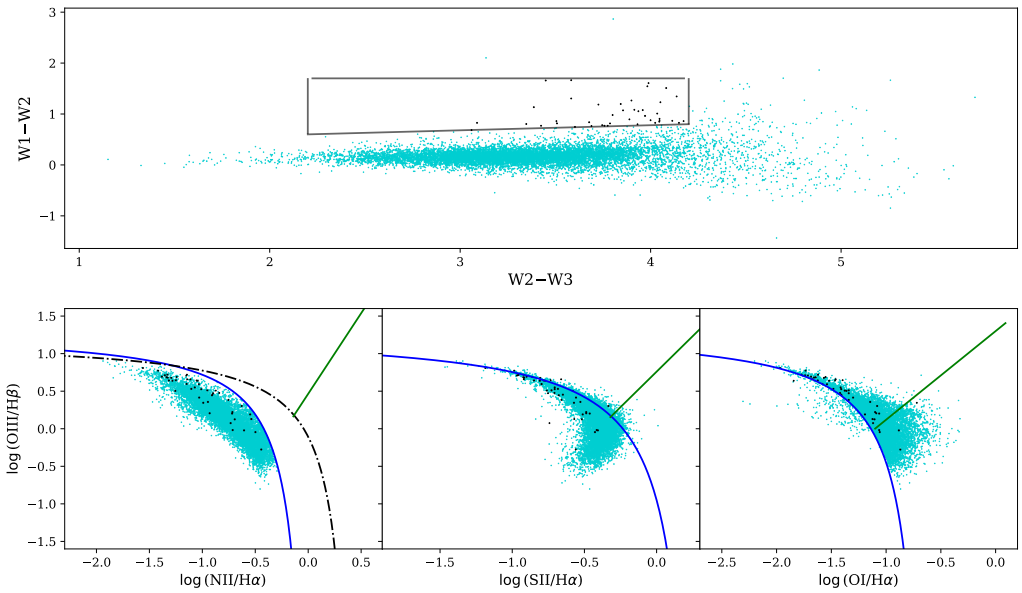

Figure 4.8: *Top row:* WISE colour-colour diagram between the W1, W2 and W3 bands. The black polygon represents the AGN region defined by equation 4.5. The black dots are the Ambiguous SF galaxies that are classified as AGN. *Bottom row:* BPT (left), SII (middle) and OI (right) diagrams of the FADO Ambiguous SF galaxies, as defined by Table 4.7. The black dots represent the AGN galaxies selected by the WISE colour-colour diagram. The remaining elements represent the same as Figure 4.3.

To prove this point, we have selected 20 000 random galaxies: 5 000 who are SF according to the BPT diagram; 5 000 who are SF according to the OI diagram; 5 000 who are AGN according to the BPT diagram; and 5 000 who are AGN according to the OI diagram. If the WISE diagram inherits the BPT biases, then the classification of these schemes should be similar, with the OI diagram having the biggest differences in classification. Of these 20 000 galaxies, 17 519 matched with WISE, with its distribution seen in Table 4.8.

Breaking down the WISE colour-colour diagram of the random galaxies (Figure 4.9), $\approx 22\%$ of the BPT AGN galaxies are also classified as AGN with WISE, more than three times higher than the OI AGN galaxies, with only $\approx 7\%$ of the galaxies matching. When it comes to the BPT and OI SF galaxies, it is expected that more BPT SF galaxies match with the WISE classification, since the OI SF galaxies include both Composites and AGNs, however, both have a $\approx 99\%$ match of the SF sample. Overall, the BPT classification matches with $\approx 59\%$ of the WISE classification and the OI diagram matches with $\approx 54\%$. Although this difference is small, it does point towards the fact that the BPT and WISE classifications can have difficulties classifying galaxies, since both in the AGN and SF types there are inconsistencies in classification when comparing the BPT and OI diagrams. These galaxies demand further study, to better understand if what drives them to different placements across emission line ratio diagrams, as well as diagrams at other wavelengths, is, in fact, an AGN in their core, or if it is something else.

Table 4.8: Randomly selected SF and AGN galaxies from the BPT and OI diagrams, as well as their matches with the WISE data.

Criteria	Galaxies selected	WISE matches
BPT SF	5 000	4 226
BPT AGN	5 000	4 330
OI SF	5 000	4 675
OI AGN	5 000	4 288
Total	20 000	17 519

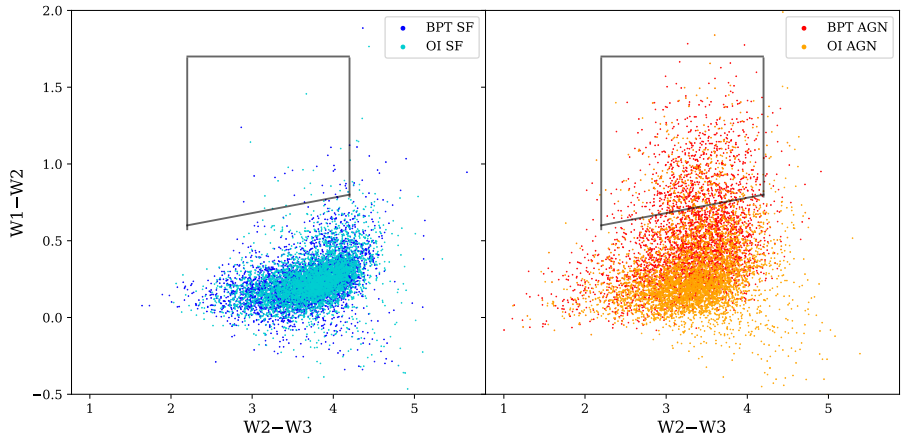

Figure 4.9: Random SF (left) and AGN (right) galaxies according to the BPT and OI diagrams, with the WISE colour-colour diagram. The black polygon represents the same as in Figure 4.8.

Ambiguous Composites and AGNs can only be compared with their stellar mass. Figure 4.10 shows that there does not appear to be any correlation with stellar mass and ambiguity, as all galaxies have higher stellar masses ($M_{\odot} > 10^{10}$) than their SF counterparts.

Ambiguous Composites, in the SII diagram, mostly lie in the SF region ($\approx 81\%$), with the OI diagram having almost all galaxies in the AGN region, much like the Ambiguous SF galaxies ($\approx 96\%$).

Composites are galaxies that are SF with a component (most likely, an AGN) that drives them over the metallicity saturation of $[\mathrm{NII}]\lambda 6584/\mathrm{H}\alpha$ which means, by their own definition, that they could be placed either in the SF or AGN category. Although we do not have the metallicity of these galaxies, we can infer that they are at least of the order of or higher than the metallicity of the galaxies near the bottom of the border in the SF region (Figure 4.7), making them have super-solar metallicity. As such, it would make sense for them to be "stuck" between the SF and AGN categories in the SII diagram and "trapped" in the AGN region of the OI diagram, seen as the former has some metallicity sensitivity and the latter does not have any.

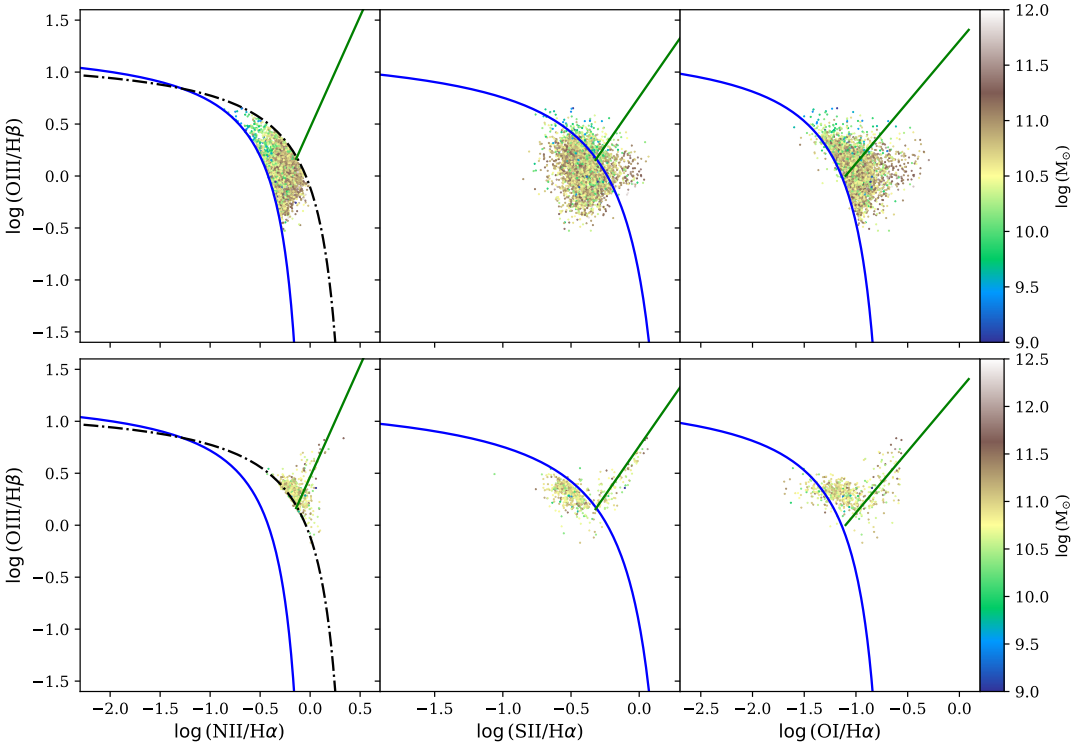

Figure 4.10: *Top Row:* BPT (left), SII (middle) and OI (right) diagrams of the FADO Ambiguous Composite galaxies, as defined by Table 4.7, colour coded by stellar mass. *Bottom row:* The same as the top row, only with the Ambiguous AGN galaxies.

Ambiguous AGNs are a more interesting case. Once again, the SII diagram has more Ambiguous galaxies in the SF region ($\approx 80\%$) while the OI diagram has more in the AGN region ($\approx 79\%$). In the SII diagram, the AGN galaxies all seem to be "trapped" in the SF region, however, they are all so close to the border between classifications (unlike the Composites), that we cannot certainly claim that the SII diagram classifies them as SF for sure. When it comes to the OI diagram, it is mostly in agreement with the BPT diagram, with the galaxies in the SF region suffering from the same flaws as the ones in the SII diagram. Overall, we can say that these Ambiguous AGNs are a product of the uncertainty of these diagrams, as even those that do not match the BPT classification lie so close to the border they could belong to this diagram's classification.

4.3.2 Non-Ambiguous galaxies

With the Ambiguous galaxy thoroughly studied, we will now compare the three diagrams without them, to see if the classification matches between FADO and MPA-JHU. Looking at Figure 4.11, we can see once again that the BPT diagram has three distinct areas, which is to be expected. Something to note

with the OI and SII diagrams, however, is that, without the Ambiguous galaxies, the Composite galaxies are inside the distribution of the SF galaxies: this tells us that, to distinguish these transitional objects, we need the BPT diagram as, without its more rigid classification scheme, we can't really understand their nature.

Table 4.9: Median differences (in dex) between the FADO and MPA-JHU emission line ratios of the BPT, SII and OI diagrams for the non-Ambiguous galaxies as defined in this work.

	FADO − MPA-JHU			
Criteria	$\log(\mathrm{NII}/\mathrm{H}\alpha)$ (dex)	$\log(\mathrm{SII}/\mathrm{H}\alpha)$ (dex)	$\log(\mathrm{OI}/\mathrm{H}\alpha)$ (dex)	$\log(\mathrm{OIII}/\mathrm{H}\beta)$ (dex)
SF	0.017	0.003	0.081	-0.005
Composite	-0.001	-0.005	0.089	0.009
AGN	-0.014	0.006	0.068	0

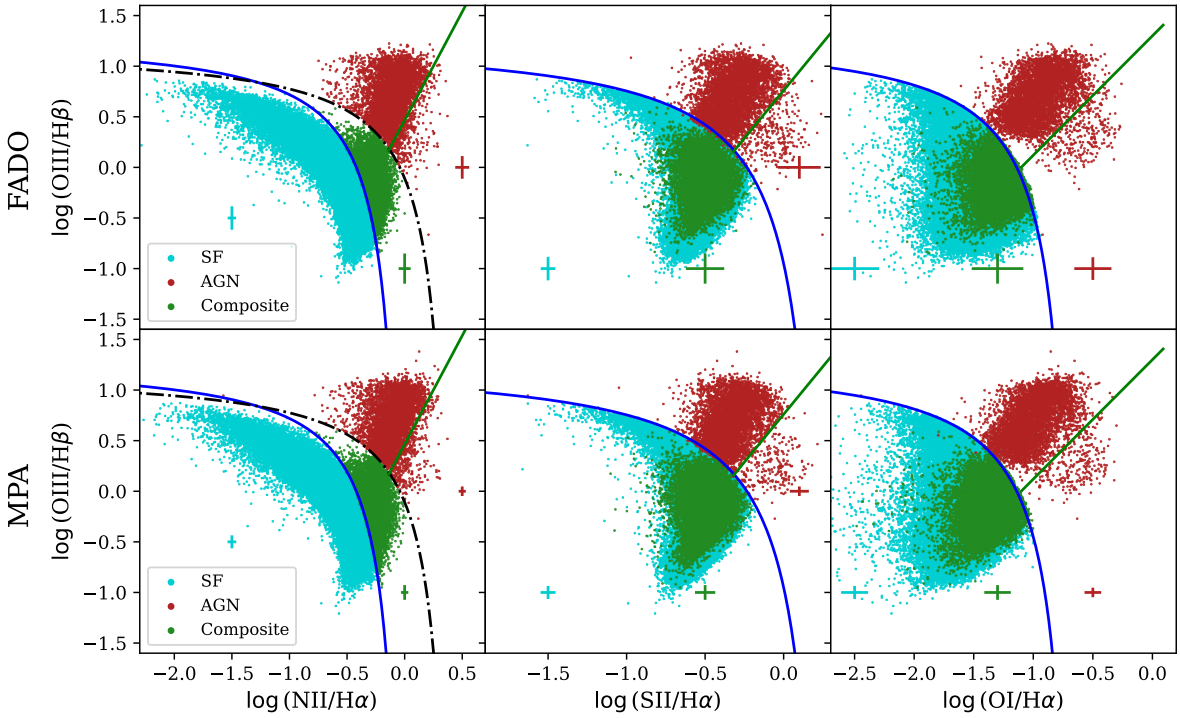

Figure 4.11: BPT (left), SII (centre) and OI (right) diagrams of the non-Ambiguous galaxies as defined previously in this work. The remaining elements are the same as Figure 4.3.

When it comes to differences between FADO and MPA-JHU, we can see, from Figure 4.12 and Tables 4.9 and 4.10, that they are minute in the BPT and SII diagrams ($\approx 4\%$), with the OI diagram having higher differences ($\approx 22\%$). We can explain these differences on the flux and S/N cut we have performed in the beginning of this Chapter. It is also of note that that uncertainties in these datasets compensate for the differences between them. With this in mind, what we have stated in the previous chapter maintains: the addition of self-consistent nebular emission in FADO does not translate into a change in galaxy classification, at least not in galaxies that are consistent in classification across all three diagrams.

Table 4.10: Median uncertainties (in dex) of the FADO and MPA-JHU emission line ratios for the BPT, SII and OI diagrams, in the non-Ambiguous sample of galaxies as defined in this work.

FADO	$\log(\mathrm{NII}/\mathrm{H}\alpha)$ (dex)	$\log(\mathrm{SII}/\mathrm{H}\alpha)$ (dex)	$\log(\mathrm{OI}/\mathrm{H}\alpha)$ (dex)	$\log(\mathrm{OIII}/\mathrm{H}\beta)$ (dex)
SF	0.037	0.047	0.207	0.116
Composite	0.052	0.122	0.216	0.149
AGN	0.061	0.136	0.156	0.112
MPA-JHU	$\log(\mathrm{NII}/\mathrm{H}\alpha)$ (dex)	$\log(\mathrm{SII}/\mathrm{H}\alpha)$ (dex)	$\log(\mathrm{OI}/\mathrm{H}\alpha)$ (dex)	$\log(\mathrm{OIII}/\mathrm{H}\beta)$ (dex)
SF	0.038	0.048	0.113	0.064
Composite	0.032	0.064	0.111	0.073
AGN	0.031	0.061	0.070	0.047

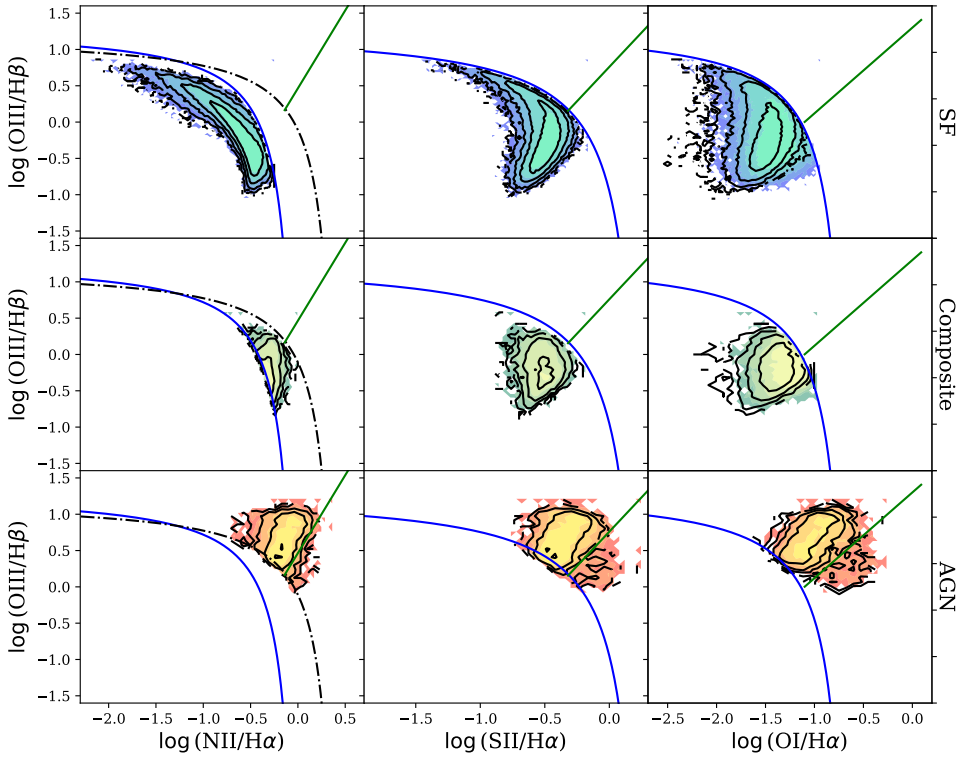

Figure 4.12: BPT (left), SII (centre) and OI (right) diagrams of the non-Ambiguous galaxies as defined previously in this work. The colourful countours represent the FADO data and the black lines the MPA data. The remaining elements are the same as Figure 4.3.

Chapter 5

New classification scheme

Comparing the emission line flux ratios with the equivalent widths of other emission lines, for the galaxies with common classification between FADO and MPA-JHU in the BPT diagram (Table 3.9), we detected an interesting correlation when plotting the emission line flux ratio of $[\mathrm{OIII}]\lambda 5007/\mathrm{H}\beta$ against the EW of $\mathrm{H}\alpha$ and $\mathrm{H}\beta$: galaxies seem to naturally separate themselves in two distinct clusters: one for AGN and another for SF and Composite (according to the BPT diagram classification, as can be seen by Figure 5.1). There are similar studies using the $\mathrm{H}\alpha$ and $[\mathrm{NII}]\lambda 6584$ emission lines (Cid Fernandes et al., 2011) and, motivated by these, we decided to analyse these diagrams. This section will be focused on the FADO data, as MPA-JHU EWs have the same properties as FADO, albeit with a slight shift in the position of galaxies.

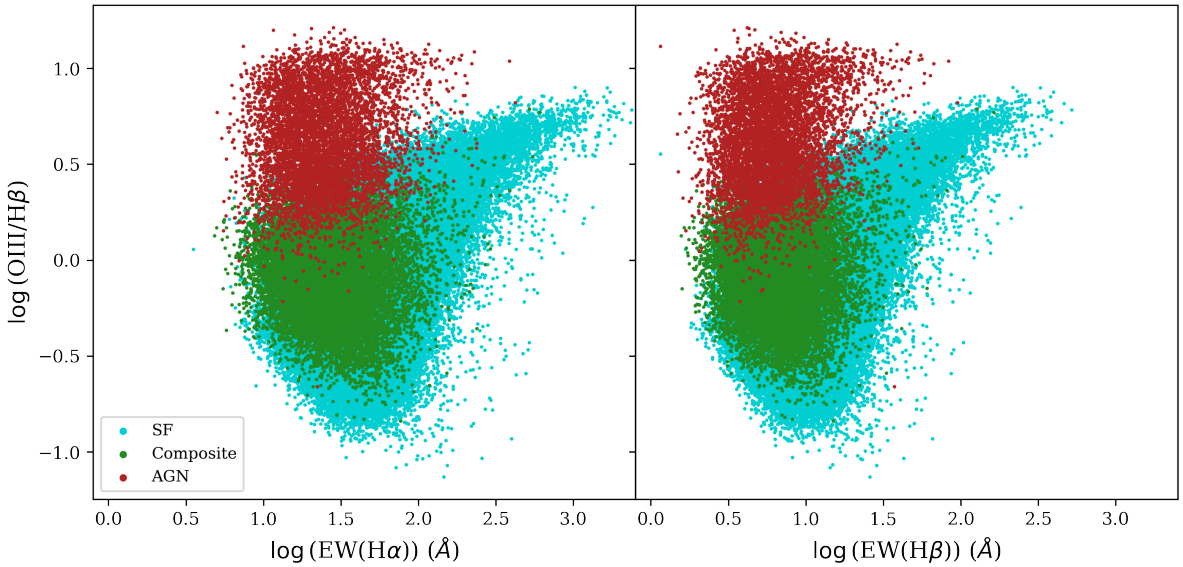

Figure 5.1: Comparison between the FADO EWs of $\mathrm{H}\alpha$ and $\mathrm{H}\beta$ with the emission line flux ratio of $[\mathrm{OIII}]\lambda 5007/\mathrm{H}\beta$, for the 153 760 galaxies with common classification in FADO and MPA-JHU. The blue points represent SF galaxies, the green points represent Composite galaxies and the red points are AGNs, according to the BPT diagram.

5.1 Emission line comparison

As a first look of this classification scheme, we compared the BPT diagram emission lines in four sets, both for the EW and the flux:

1. Comparing Hα with [OIII]λ5007/Hβ;
2. Comparing Hβ with [OIII]λ5007/Hβ;
3. Comparing [NII]λ6584 with [OIII]λ5007/Hβ;
4. Comparing the addition of Hα and [NII]λ6584 with [OIII]λ5007/Hβ.

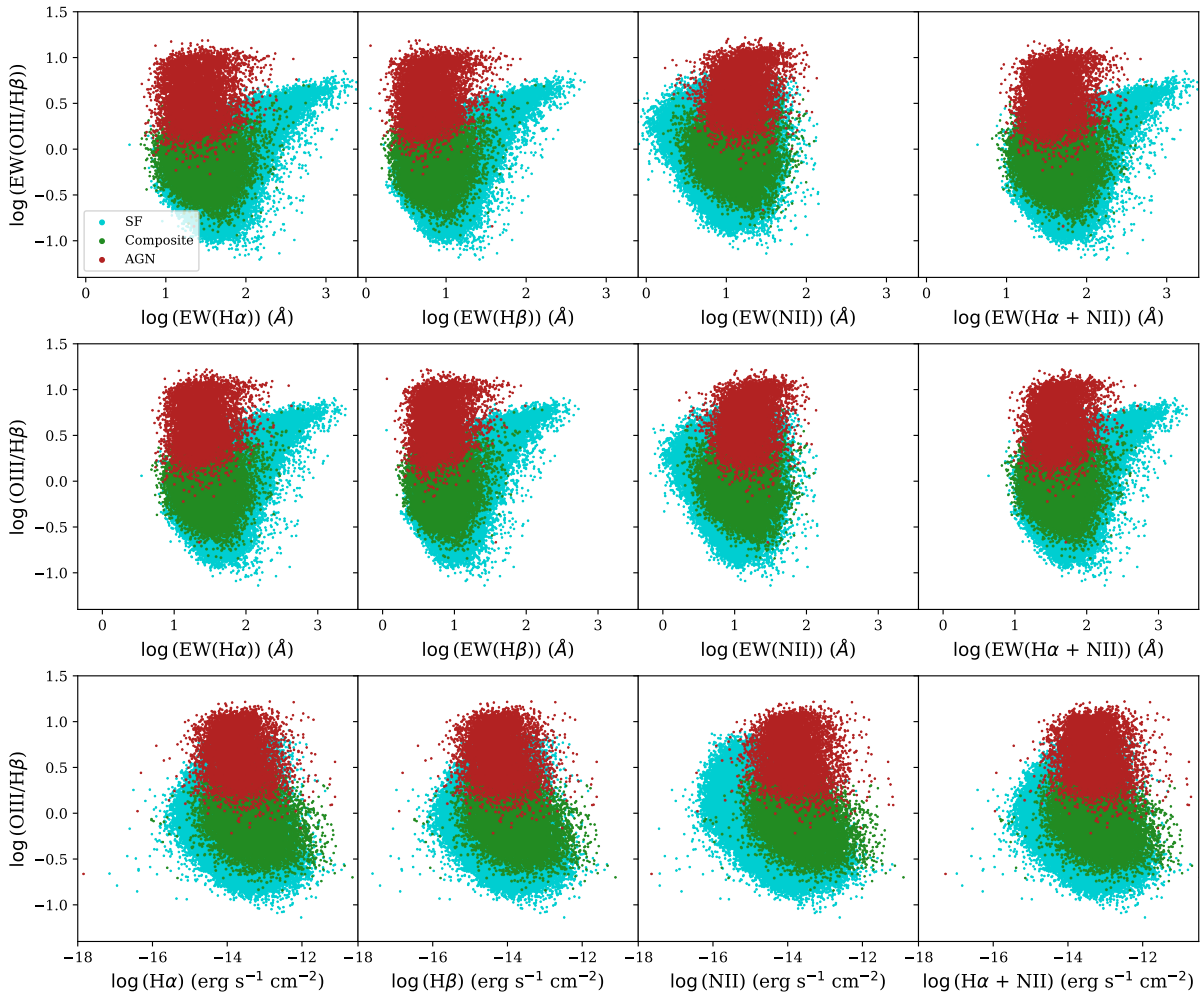

Figure 5.2: Comparison of the FADO EWs and fluxes of Hα, Hβ, [NII]λ6584 and the sum of Hα and [NII]λ6584 with the division of the EW and flux of [OIII]λ5007 and Hβ. The remaining elements represent the same as in Figure 5.1.

The reason we added the final comparison set is because, in some SEDs, especially in those that are very noisy and have low resolution, it is hard to parse between the emission lines of Hα and [NII]λ6584, seeing as they have very similar wavelengths (Figure 2.3). As such, we consider adding them together as a definition of a lower limit for the flux and EW of a blended emission line. If the separation occurs as in Figure 5.1, we do not need to resolve both of these emission lines to separate galaxy types.

From Figure 5.2, we can see that this separation is clearer in the EWs of these emission lines, rather than the fluxes. Furthermore, in the EWs, only the Balmer lines and the addition of the EW of Hα and the EW of [NII]λ6584 actually separate galaxies into two distinct clusters: one for AGNs and one for SF and Composite galaxies, as we have seen previously.

The distinction between the shapes of the clusters between the flux and the EWs can be explained due to the continuum of each galaxy. From Figure 5.3, we can see that, by tracing a vertical line on the plot

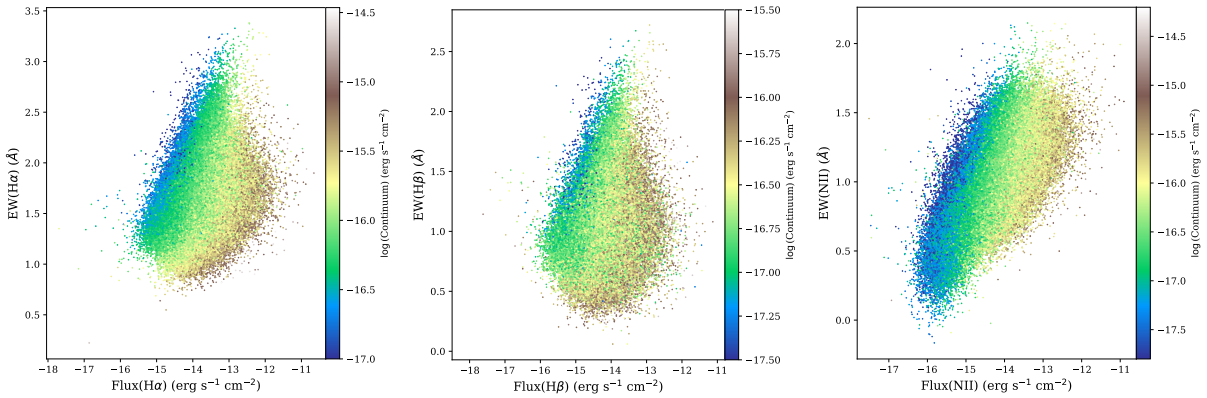

Figure 5.3: Comparison between the fluxes and EWs of Hα (left), Hβ (middle) and [NII]λ6584 (right), with the colour map representing the continuum flux for each emission line.

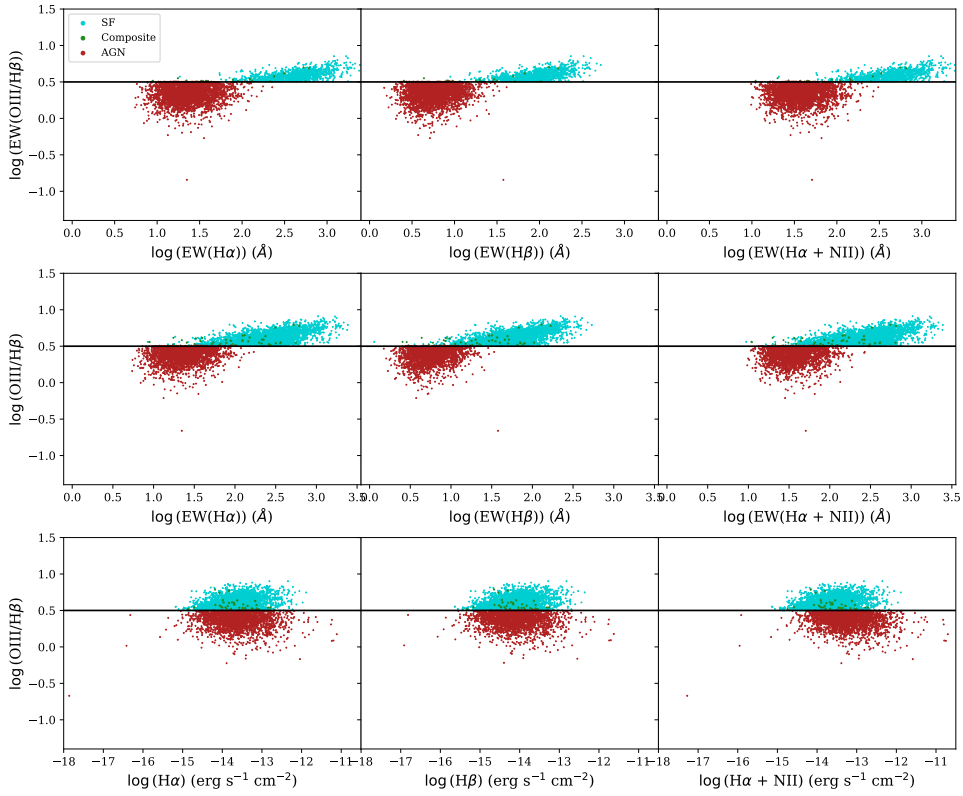

Figure 5.4: Comparison of the FADO EWs and fluxes of Hα, Hβ and the sum of Hα and [NII]λ6584 with the division of the EW and flux of [OIII]λ5007 and Hβ. Represented here are the galaxies that cross over the classification limit of $\log([\text{OIII}]\lambda5007/\text{H}\beta) = \log(\text{EW}([\text{OIII}]\lambda5007/\text{H}\beta)) = 0.5$. The remaining elements represent the same as Figure 5.1.

(or, for a constant flux), the values of the continuum of the galaxies do not change significantly, meaning that the flux is a tracer of galaxies with similar continuums. The same cannot be said about the EWs: by tracing a horizontal line, we can see that the values of the continuum for a constant EW vary from the lowest to highest in nearly all scenarios. This happens because, when calculating the EW, we divide the flux by the continuum of that emission line, meaning it will become independent of the continuum. This tells us why we have a clearer separation of classifications with the comparison between EWs rather than fluxes.

Besides this, the reason why the Balmer lines can distinguish between galaxy types and [NII]λ6584 cannot is due to the fact that the former are more sensitive to pure ionisation, and are also easier to detect

in galaxies. As such, this translates into the Balmer lines having only SF galaxies with high EWs, while the [NII]λ6584 emission line cannot reach those values and, therefore, is unable to separate galaxies in two distinct clusters.

Discarding the [NII]λ6584 emission line diagrams for the reasons mentioned above, we focused on the remaining classification schemes. To further prove the claim that the EWs separate galaxy types better than fluxes, we decided to create a simple diagnostic test: comparing the number of galaxies of one type that cross over to the classification of another type, using $\log([\text{OIII}]\lambda 5007/\text{H}\beta) = \log(\text{EW}([\text{OIII}]\lambda 5007/\text{H}\beta)) = 0.5$ as a simplified limit between AGN and SF and Composite galaxies, as classified by the BPT diagram. Overall, the comparison between EWs only have a total of 3 968 galaxies that cross over this line (or $\approx 3\%$ of the sample used), while the remaining diagrams have 5 893 galaxies that cross over ($\approx 4\%$ of the same sample). As we can see by Figure 5.4, the galaxies that cross over this line in the EWs have a distinct shape that the diagrams with fluxes only does not possess. Although the number of galaxies that cross over galaxy classification is similar in these three cases, we argue that the EWs could have an even smaller number of galaxies that cross over classification using a more complex separation function, something that cannot be done with the fluxes only.

Further comparison of the EWs only diagrams (top row of Figure 5.1) with the flux and EWs diagrams (middle row of Figure 5.1) is necessary to understand why one might be more useful than the other. Since the differences between these two diagrams come from the vertical axis, we have decided to compare these diagrams with a fixed horizontal axis, choosing the EW of Hα. Figure 5.5 shows the differences between these two diagrams, and we found that they are in the order of 0.1 dex for all galaxy types, or $\approx 26\%$. Noticeably, both diagrams retain the same shape for all the BPT diagram classifications. This means that, when we devise a function of separation, we can create one for the EWs only diagrams and raise it by 0.1 dex to have similar results in the EWs and flux diagrams and, in turn, we can use both of these classification schemes to separate galaxy types.

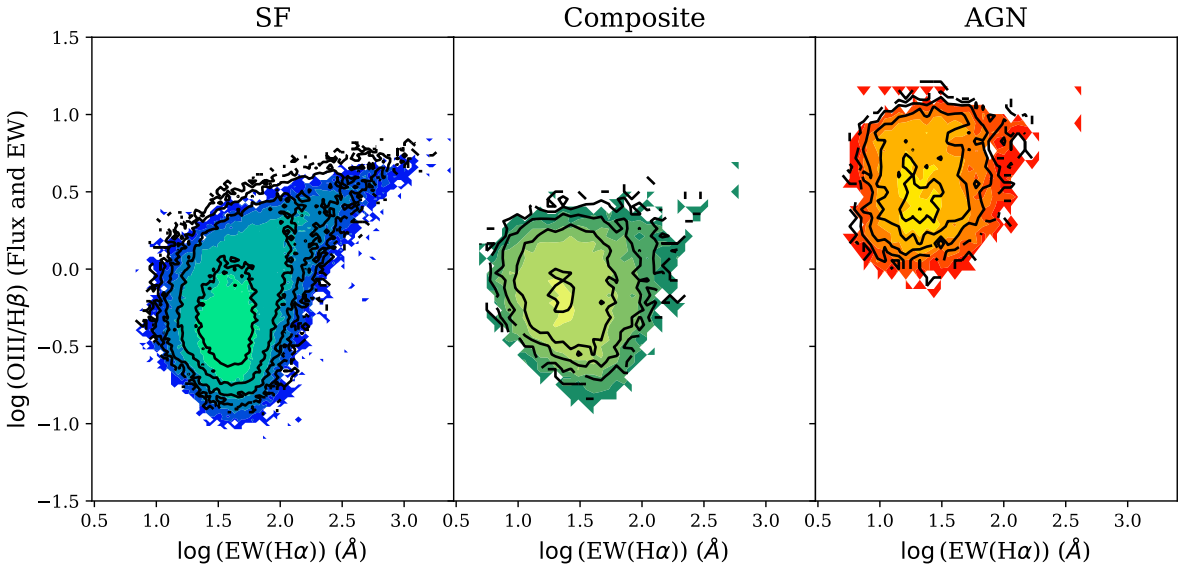

Figure 5.5: Comparison between the EWs only diagram (coloured contours) and the flux and EW diagram (black contours), for the EW of Hα. The differences between these diagrams are approximately 0.1 dex for each classification.

5.2 Functions of separation

The diagrams that compare EWs of Hα, Hβ and the sum of Hα and [NII]λ6584 with the ratio of the EWs of [OIII]λ5007 and Hβ (henceforth to be known as the EWA, EWB and EWAII diagrams, respectively) have a cleaner distinction of AGN and SF and Composite than their flux counterparts. It is important now to define this distinction, one for each of the three diagrams.

We decided to use the upper limit of the SF and Composite cluster for each of the three diagrams to delineate the frontier between galaxy classification, where we found an optimal solution with hyperboles. Equations 5.1, 5.2 and 5.3 are the solutions that minimise the amount of "contaminated" galaxies that exist in these diagrams, that is, galaxies that are classified by the BPT diagram as SF or Composite and cross over to the AGN side of the EW diagrams, and vice versa. The functions of separation can be seen in Figure 5.6.

$$\log\left(\frac{\mathrm{EW}([\mathrm{OIII}]\lambda 5007)}{\mathrm{EW}(\mathrm{H}\beta)}\right) = \frac{-3.03717}{\log(\mathrm{EW}(\mathrm{H}\alpha)) + 1.15584} + 1.50279 \tag{5.1}$$

$$\log\left(\frac{\mathrm{EW}([\mathrm{OIII}]\lambda 5007)}{\mathrm{EW}(\mathrm{H}\beta)}\right) = \frac{-3.09011}{\log(\mathrm{EW}(\mathrm{H}\beta)) + 1.83402} + 1.49031 \tag{5.2}$$

$$\log\left(\frac{\mathrm{EW}([\mathrm{OIII}]\lambda 5007)}{\mathrm{EW}(\mathrm{H}\beta)}\right) = \frac{-3.01187}{\log(\mathrm{EW}(\mathrm{H}\alpha) + \mathrm{EW}([\mathrm{NII}]\lambda 6584)) + 1.11286} + 1.47934 \tag{5.3}$$

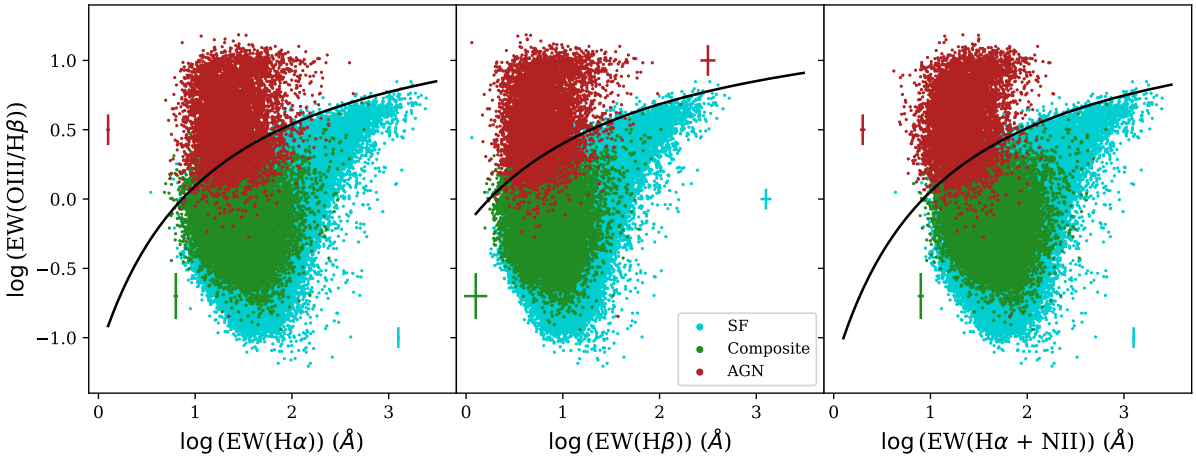

Figure 5.6: EW diagrams, and the functions that separate two clusters: one for AGNs and one for SF and Composite, according to the BPT diagram classification. On the left, we have the EWA diagram, where the black line represents equation 5.1. In the middle, we have the EWB diagram, where the black line represents equation 5.2. On the right, we have the EWAII diagram, where the black line represents equation 5.3. The error bars represent the median error for each classification.

As we can see from Table 5.1, for all the diagrams considered, these contaminated galaxies represent, at most, 1% of the 153 760 galaxies with common classification. The effect on SF and Composite galaxies is also very low (less than 1% for SF and a maximum of 2% for Composites), however, when it comes to AGN, the cut-off is much harsher (a maximum of 24%). This is a trade-off that comes with this diagram: sacrificing more AGNs to have the least amount of contamination possible. Despite this, we can see that the contaminated galaxies all lie near the borders of the BPT diagram (Figure 5.7), meaning that their placement in these EW diagrams translates over well to this emission line ratio diagram, as their positions near the border reveals some ambiguity to their classification as well. Overall,

although the contamination across all diagrams is similar, the EWB diagram presents the least amount of contaminated galaxies.

Table 5.1: Number of contaminated galaxies as defined in this work, followed by the how much (in percentage) do these galaxies represent in each classification, as defined by Table 3.9, i.e., the 18% in the EWA AGNs represent the percentage of AGNs that cross over the classification, in regards to the entire AGN population.

Criteria	EWA	%	EWB	%	EWAII	%
SF	195	< 1	199	< 1	276	<1
Composite	330	2	259	2	205	1
AGN	1 130	18	1 104	17	1 513	24
Total	1 655	1	1 562	1	1 994	1

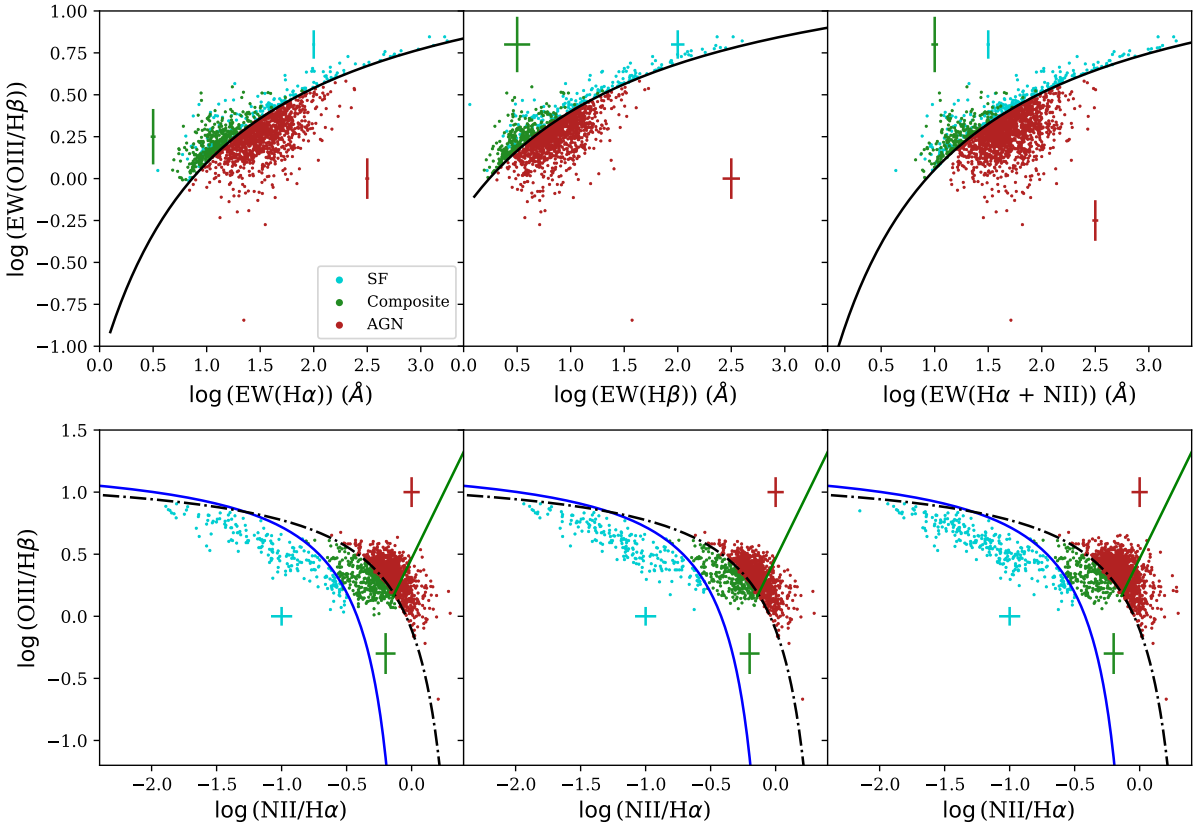

Figure 5.7: On the top row, we have the scatter plot of the EW of Hα, Hβ and the sum of Hα and [NII]λ6584 with the division of the EWs of [OIII]λ5007 and Hβ of the galaxies that we considered to be contaminated, for FADO (Table 5.1). On the bottom row, the BPT diagram of these galaxies is represented. The error bars represent the median errors .

It is also important to know, based on the BPT diagram classification, what is the chance that a galaxy could be misclassified according to these EW diagrams. Table 5.2 shows that the highest percentage chance of a galaxy being misclassified is 11%, specifically in the EWAII diagram. As expected, the EWB diagram shows the least chance of contamination for both of these categories. The Composites are not considered as a classification on its own since they are mixed with the SF galaxies and cannot be easily separated. This means that the EW diagrams performs well at classifying SF galaxies, but can have issues when classifying AGNs, according to the BPT classification.

Moving on to look at the comparison between fluxes and the EWs, we used the same functions of separation as defined in equations 5.1, 5.2 and 5.3, albeit with an addition of 0.1 dex, to compensate the differences seen in Figure 5.5. The result can be seen in Figure 5.8 and, from Table 5.3, these diagrams

Table 5.2: Number of galaxies in the EW diagrams that match with the BPT classifcation, as well as the percentage of how many galaxies are misclassified, i.e., the 8% in the EWA SF represent the percentage chance of a galaxy being classified as an AGN instead of SF when only considering the EWA diagram.

Criteria	EWA	Misclass %	EWB	Misclass %	EWAII	Misclass %
SF match	133 011	8	133 007	8	132 930	11
AGN match	5 247	10	5 273	9	4 864	10

have more contaminated galaxies than the purely EWs ones. Differences in the SF and Composites galaxies do not differ much from the EW diagrams ($< 1\%$ for SF and a maximum of 2% for Composites for both cases), with the highest differences belonging to the AGN classification, although by a small margin (maximum of 28%, compared with the 24% maximum in the EWAII diagram).

Table 5.3: Number of contaminated galaxies followed by the how much (in percentage) do these galaxies represent in each classification, as defined by Table 3.9, with the same reasoning as in Table 5.1.

Criteria	Hα	%	Hβ	%	Hα+NII	%
SF	283	< 1	288	< 1	409	<1
Composite	234	2	161	2	140	1
AGN	1 322	21	1 322	21	1 768	28
Total	1 839	1	1 771	1	2 317	2

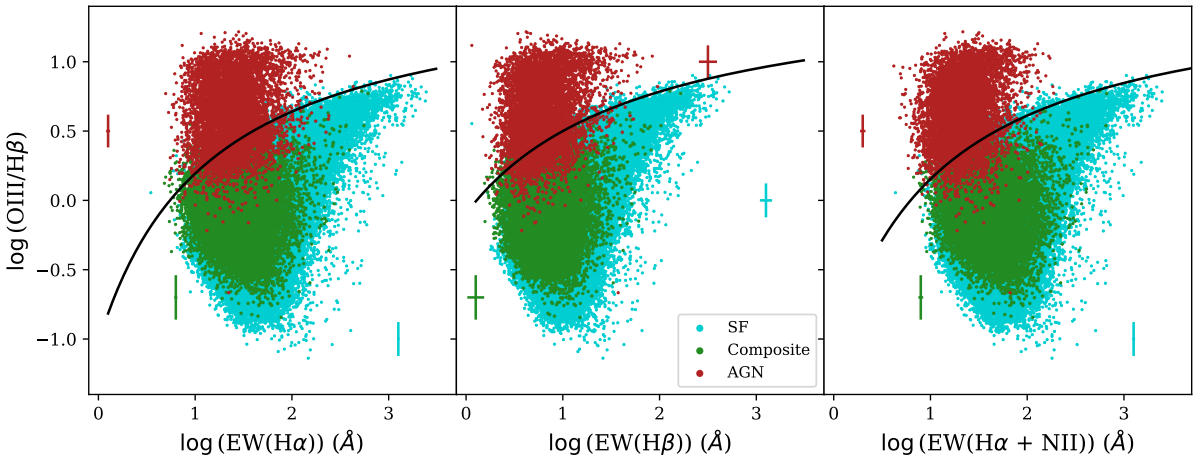

Figure 5.8: EW and flux diagrams, and the functions that separate two clusters: one for AGNs and one for SF and Composite, according to the BPT diagram classification. The black lines represent equations 5.1 (left diagram), 5.2 (middle diagram) and 5.3 (right diagram) with a 0.1 dex addition. The error bars represent the median error for each classification.

Although the differences between contaminated galaxies of the EW diagrams and the EW and flux diagrams are minimal, we should base ourselves on the ones that consider the least amount of contamination. This means that the EW diagrams are the best solution when it comes to the comparison between these emission lines, and the EW and flux diagrams are going to be disregarded from now on, even if they exhibit the same properties as the EW diagrams.

From these 3 EW diagrams, the best one at separating galaxies with the BPT diagram classifications (excluding Composites) is the EWB diagram. In the next section, we will be focusing on this diagram for the analysis of the Ambiguous and contaminated galaxies. It is of note that the differences between all these diagrams are minimal, meaning that each one could be used, depending on the context of the analysis of galaxies. Section 5.4 goes into more detail about these contexts and where to apply a given diagram.

5.3 Ambiguous and contaminated galaxies

To better understand this classification scheme, we compared the EWB diagram with the BPT, SII and OI diagrams, as it is of note to explore the relationship between contaminated and Ambiguous galaxies. Figure 5.9 and Table 5.4 shows the different relations between these samples. Overall, most contaminated galaxies are ambiguous, apart from contaminated AGN galaxies ($\approx 84\%$ match for SF galaxies, $\approx 92\%$ for Composites and $\approx 43\%$ for AGNs).

Table 5.4: Galaxies from the sample in Table 4.5 that are Ambiguous, contaminated and a common sample between Ambiguous and contaminated, as well as how much, in percentage, do they represent of each category.

Criteria	All	Ambiguous	%	Contaminated	%	Common	%
SF	122 812	11 762	10	173	< 1	145	< 1
Composite	12 147	3 473	29	206	2	190	2
AGN	5 799	605	10	979	17	422	7
Total	140 758	15 840	11	1 358	1	757	< 1

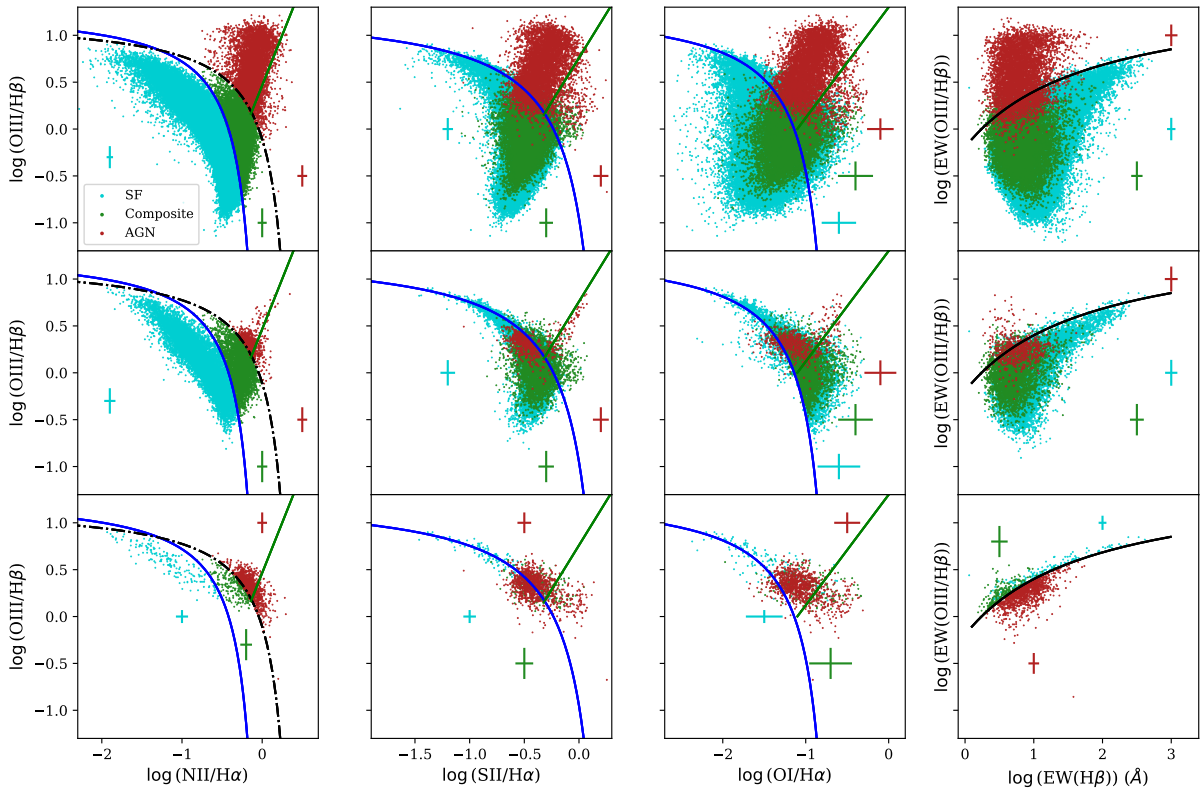

Figure 5.9: BPT (left), SII (middle left), OI (middle right) and EWB (right) diagrams of the sample defined in Table 5.4. The top row represents the full sample, the middle row the Ambiguous sample and the bottom row the contaminated sample. The error bars represent the median errors.

In the SII diagram, the contaminated galaxies lie mostly near the border between classifications, with $\approx 56\%$ belonging above the SF line. Of these galaxies, most of them are AGNs according to the BPT classification. In the OI diagram, $\approx 81\%$ of the contaminated galaxies are in the AGN region, with the BPT classified AGNs being again the galaxies that are mostly above the SF line. In the EWB diagram, the Ambiguous galaxies lie mostly ($\approx 97\%$) in the SF region, like the SII diagram.

We possess both the metallicity and stellar mass of the SF galaxies, so we can start by analysing them, as we did with the Ambiguous. Figure 5.10 shows that the Ambiguous galaxies in the EWB diagram

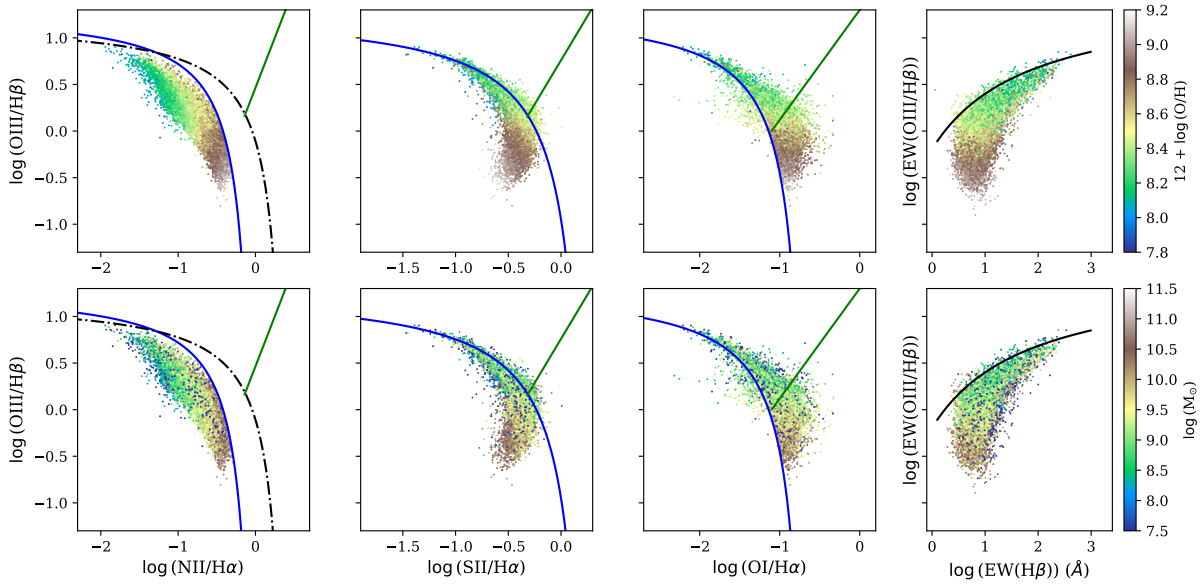

Figure 5.10: BPT (left), SII (middle left), OI (middle right) and EWB (right) diagrams of the SF Ambiguous galaxies, with the colormap representing the metallicity (top row) and stellar mass (bottom row). The remaining elements represent the same as in Figure 5.9.

have properties that are similar to the SII diagram, where the SF galaxies that have solar and super-solar metallicites or high stellar masses are "trapped" in the SF region, and galaxies with sub-solar metallicity and low stellar masses are closer to the border between classifications. Since the metallicity gradient is vertical and the function of separation is nearly horizontal, with high values of the ratio between the EWs of $[\mathrm{OIII}]\lambda 5007$ and $\mathrm{H}\beta$, then it is natural that the EWB diagram shows the same "trapping" properties as the SII diagram.

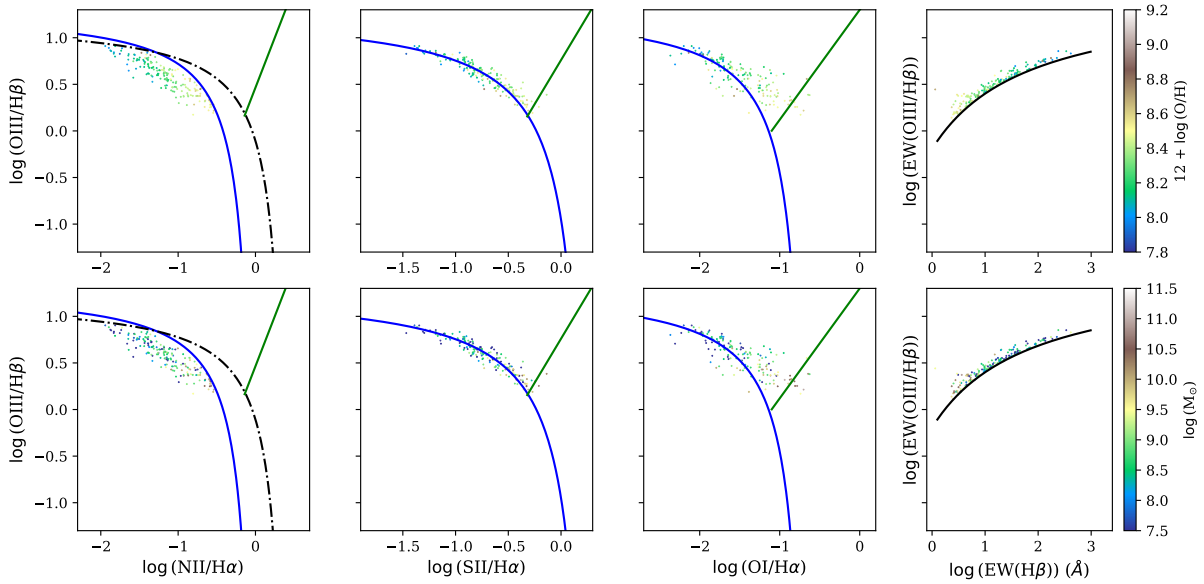

Figure 5.11: BPT (left), SII (middle left), OI (middle right) and EWB (right) of the SF contaminated galaxies, with the colormap representing the metallicity (top row) and stellar mass (bottom row). The remaining elements represent the same as in Figure 5.9.

This is translated well in the SF contaminated galaxies (Figure 5.11), where the galaxies that crossover to the AGN region in the both the SII and OI diagrams have sub-solar metallicity and low

stellar mass, essentially holding the same properties as the Ambiguous SF galaxies. This tracks as most of the contaminated SF galaxies are also Ambiguous and, as such, the contamination in the SF galaxies is a good indicator of ambiguity in classification, which has been thoroughly discussed in section 4.3.1. It is of note, however, that the contaminated galaxies are very close to the border and have a very low number count, so the match between contamination and ambiguity is not necessarily strict.

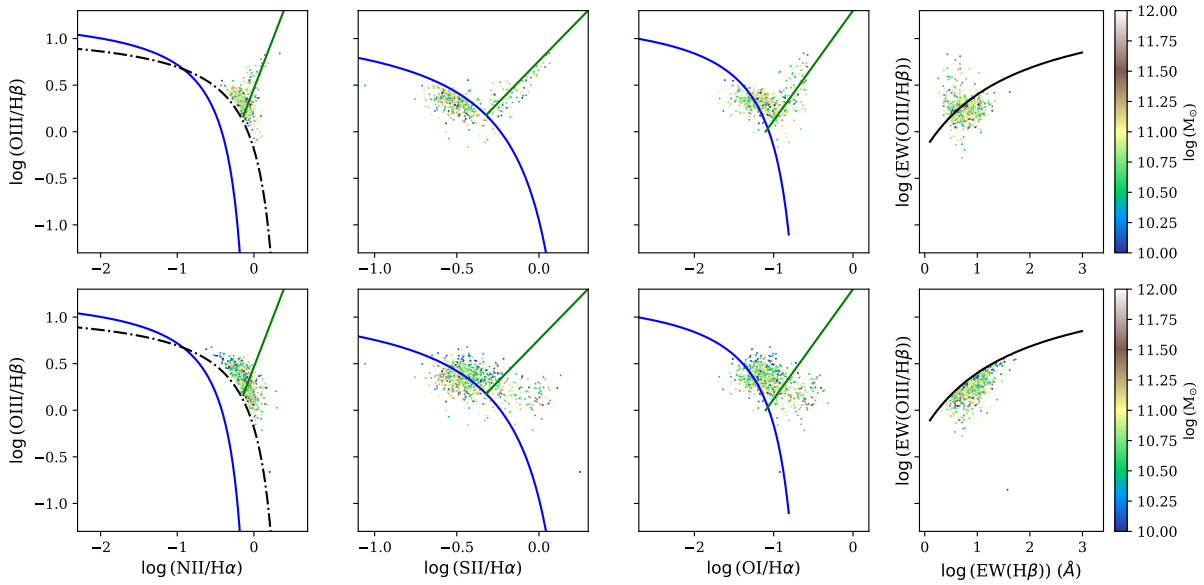

Figure 5.12: BPT (left), SII (middle left), OI (middle right) and EWB (right) of the AGN Ambiguous (top row) and contaminated (bottom row) galaxies, with the colormap representing the stellar mass (bottom row). The remaining elements represent the same as in Figure 5.9.

Composites that are contaminated are mostly Ambiguous as well, meaning they hold the same properties as previously discussed. Contaminated AGNs (Figure 5.12), however, have higher discrepancies in galaxy counts when compared to Ambiguous. This arises because of the way we classified a galaxy as Ambiguous: it is not only related to if they cross over to the SF region, but also if they are classified as Seyfert by one diagram and LINER by another, and vice versa. For simplicity, we have not made this distinction in the EWB diagram, causing these differences in galaxy counts. If we separate galaxies in Seyfert and LINER (according to the BPT diagram classification), $\approx 85\%$ of the Ambiguous LINERS are also contaminated and $\approx 66\%$ of the Ambiguous Seyferts are also contaminated. Figure 5.13 also shows that there is a better distinction between Seyfert and LINER galaxies in the emission line ratio diagrams rather than the EWB diagram, implying that the latter diagram is good at generally distinguishing AGNs from SF, but not different types of AGNs from one another: evidence in favour of this is due to the fact that half of the LINERs are placed in the SF region, with the remaining half mixed in the Seyfert population.

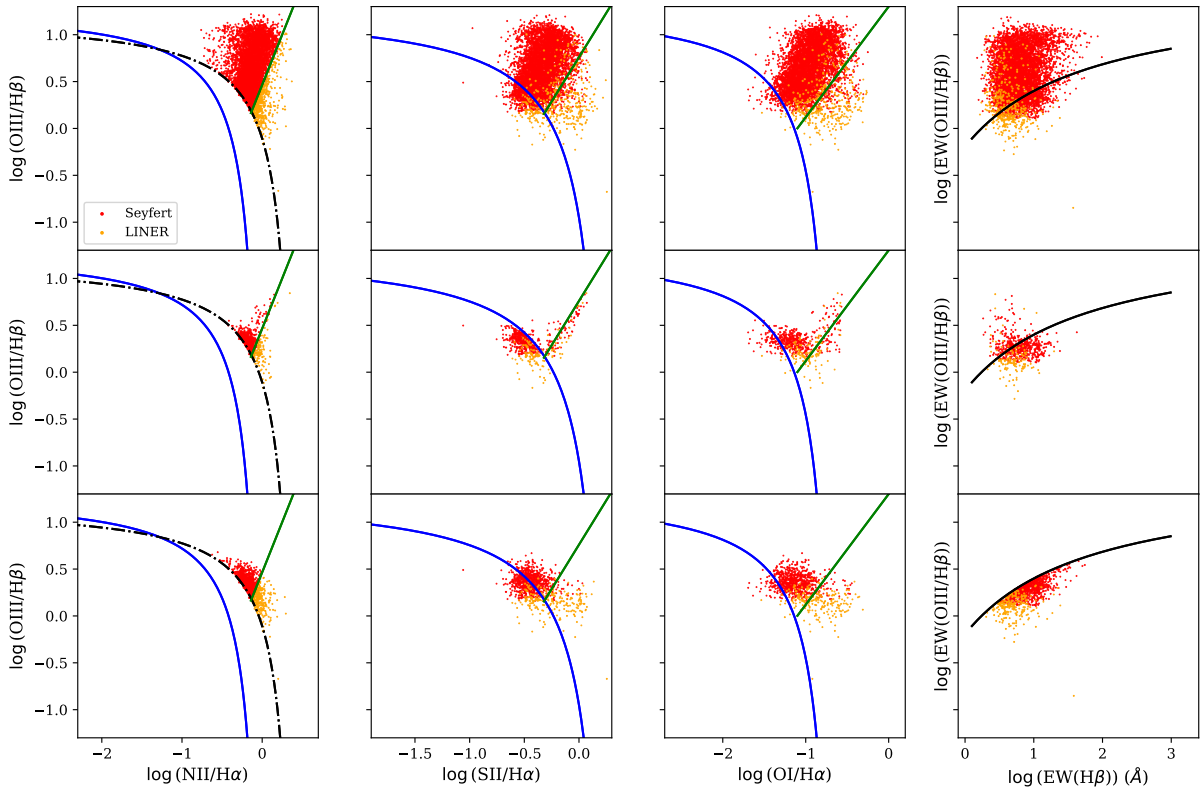

Figure 5.13: BPT (left), SII (middle left), OI (middle right) and EWB (right) of the AGN galaxies. The top row represents the full AGN sample, the middle row the Ambiguous AGNs and the bottom row the contaminated AGNs. The remaining elements represent the same as in Figure 5.9.

5.4 Uses for this classification scheme

As we move to higher redshifts, not even infrared spectroscopy can measure strong emission lines such as Hα and [NII]λ6584 (for example, in the future MOONRISE survey, Maiolino et al. 2020). However, in some galaxies, Hβ and [OIII]λ5007 are still able to be detected. This means that the BPT diagram is unable to be plotted, but the EWB diagram can be reproduced, making it useful for a preliminary diagnosis of the characteristics and properties of a galaxy. This diagram orders galaxies vertically according to metallicity and stellar mass (Figure 5.14 for the SF galaxies), and is clearly able to differentiate between the SF and AGN classifications, even if the Composite classification is unable to be parsed from the SF region. Adding to this, galaxies that cross the function of separation into other classifications have a very high chance of being Ambiguous as well, sharing its properties, despite their low number counts.

In the case we have a low resolution SED of a galaxy, the Hα and [NII]λ6584 emission lines may not be able to be resolved, meaning that the emission line ratio flux will not be able to be calculated and, consequently, the BPT diagram will not be able to plotted. If we consider the addition of the EWs of Hα and [NII]λ6584 as a lower limit for the EW of their blended line, then we can use the EWAII diagram to perform an analysis of these galaxies without having to resort to weaker emission lines such as [OI]λ6300 or the SII doublet (which may not be able to be de-blended as well, making us unable to plot the SII diagram). Another option is to use the EWB diagram instead, as we also need both the Hβ and [OIII]λ5007 emission lines to plot these diagrams.

Overall, although the EW diagrams do not have the same sensibility in the classification of galaxies as the BPT diagram (namely with the separation of the Composites from the SF and AGN), they provide us

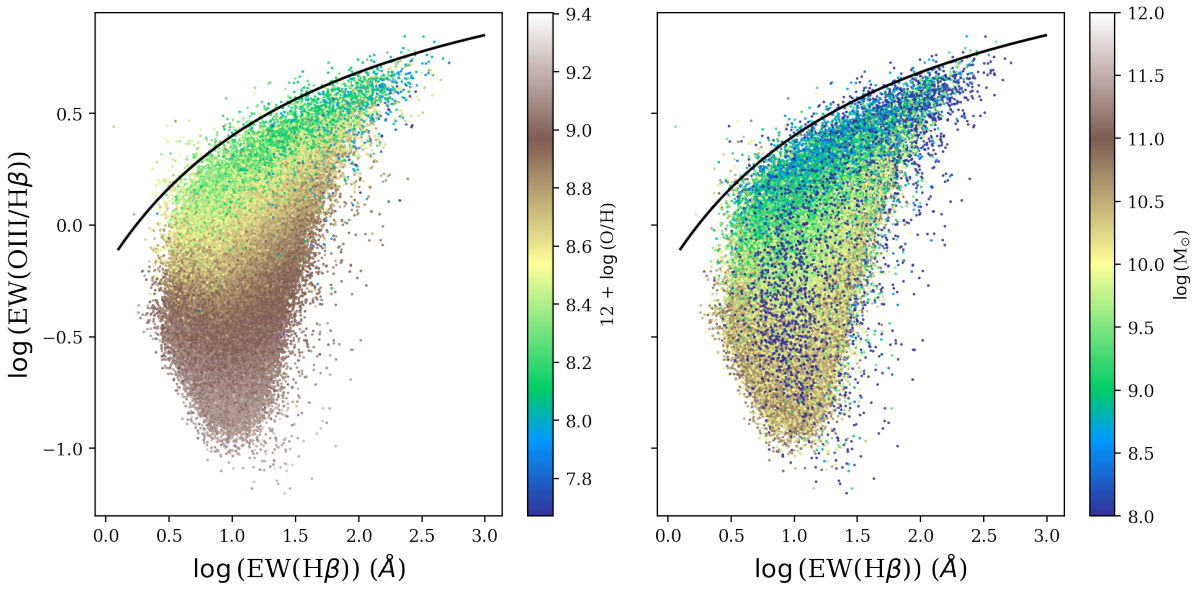

Figure 5.14: Comparison between the SF galaxies in the EWB diagram and the metallicity (left) and stellar mass (right).

useful information regarding galaxies that could have AGN activity and, more meaningfully, it does not require high resolution spectra or the 4 emission lines of the BPT diagram to be plotted and distinguish between galaxy types.

Chapter 6

Main sequence with FADO and MPA-JHU

In this chapter we will compare the main sequence of galaxies between FADO and MPA-JHU, to understand if the self-consistent inclusion of nebular emission affects the way these galaxies manifest their placement on this diagram. We will also analyse the Ambiguous galaxies in this diagram, to understand if their position in this diagram reveals us anything about their nature.

6.1 Estimation of the SFR and stellar mass

Firstly, we explain how we calculated the SFR for each galaxy and, afterwards, the same for the stellar mass, comparing the results between FADO and MPA-JHU along the way.

6.1.1 Star Formation Rate

The flux of the Hα emission line in purely SF galaxies is a proxy measurement of the SFR, due to the fact that this emission line originates from young and massive stars ionising the gas that surrounds them. If we can measure this property, we have a direct way to obtain the SFR from galaxies. As such, for the SF galaxies, we used the same approach as Miranda et al. (2023) to calculate the SFR, which uses the Kennicutt (1998) method, given by equation 6.1.

$$\mathrm{SFR} = \frac{L(\mathrm{H}\alpha)}{\eta(\mathrm{H}\alpha)} \tag{6.1}$$

where $L(\mathrm{H}\alpha)$ represents the Hα luminosity and $\eta(\mathrm{H}\alpha)$ represents the conversion factor between luminosity and SFR. This equation assumes a continuous star-forming period of 100 Myrs and metallicities equivalent with the solar metallicity. This model is very simple and, as such, it does not take into account many factors. However, since SF galaxies that lie outside these assumptions are so few and far between, it is a good general model to apply to them.

When it comes to the conversion factor, FADO and MPA-JHU used different values for $\eta(\mathrm{H}\alpha)$, due to the different IMFs chosen when fitting the model to the spectra (Chabrier 2003 and Kroupa 2001, respectively). These are $\eta_{\mathrm{Chabrier}}(\mathrm{H}\alpha) = 10^{-41.28}$ erg s^{-1} M$_\odot^{-1}$ yr, for FADO, and $\eta_{\mathrm{Kroupa}}(\mathrm{H}\alpha) = 10^{-41.31}$ erg s^{-1} M$_\odot^{-1}$ yr, for MPA-JHU (values adapted from Kennicutt 1998, according to Brinchmann et al. 2004).

$$\frac{\eta_{\mathrm{Chabrier}}(\mathrm{H}\alpha)}{\eta_{\mathrm{Kroupa}}(\mathrm{H}\alpha)} = \frac{10^{41.28}}{10^{41.31}} = 10^{-0.03} \tag{6.2}$$

This equations tells us the relation between these two conversion factors. To ensure that the comparison between the SFR of FADO and MPA-JHU is viable, we must convert the SFR of one dataset to match the SFR of the other, due to the different IMFs. As such, we have that:

$$\frac{\mathrm{SFR}_{\mathrm{Chabrier}}}{\mathrm{SFR}_{\mathrm{Kroupa}}} = 10^{-0.03} \Leftrightarrow \mathrm{SFR}_{\mathrm{Chabrier}} = 10^{-0.03}\,\mathrm{SFR}_{\mathrm{Kroupa}} \tag{6.3}$$

Since the SFR is usually given in logarithmic form, then we have the following:

$$\log(\mathrm{SFR}_{\mathrm{Chabrier}}) = \log(\mathrm{SFR}_{\mathrm{Kroupa}}) + \log(0.933) \tag{6.4}$$

To ensure that the SFRs are comparable, we must add $\log(0.933)$ to the logarithmic SFR of MPA-JHU.

To calculate the luminosity of Hα, we use equation 6.5:

$$L(\mathrm{H}\alpha) = F(\mathrm{H}\alpha)4\pi d_L^2 \tag{6.5}$$

where $F(\mathrm{H}\alpha)$ represents the emission line flux, corrected for dust absorption, of Hα and d_L the luminosity distance, for each galaxy. The luminosity distance is calculated through the redshift and uses a cosmology with $H_0 = 70$ km s^{-1} Mpc^{-1}, $\Omega_M = 0.3$ and $\Omega_\Lambda = 0.7$, with the same set of assumptions as described previously.

This SFR calculated is the one measured with the SDSS fibers which, due to the fiber size, might be unable to fully characterise a galaxy. The MPA-JHU dataset, however, empirically calculated the total SFR of the SDSS galaxies, and also provided the fiber SFR. As such, by dividing the fiber and total SFR we obtain a conversion factor for each galaxy that we will be able to use on the SFR calculated with equation 6.1 to obtain the total SFR of these SF galaxies. With this in mind, Figure 6.1 compares the fiber and total SFRs between FADO and MPA-JHU. Overall, the total SFR is, on average, three times higher than the fiber SFR, and the values obtained between FADO and MPA-JHU are essentially identical. That tracks, as the Hα emission line fluxes are also similar.

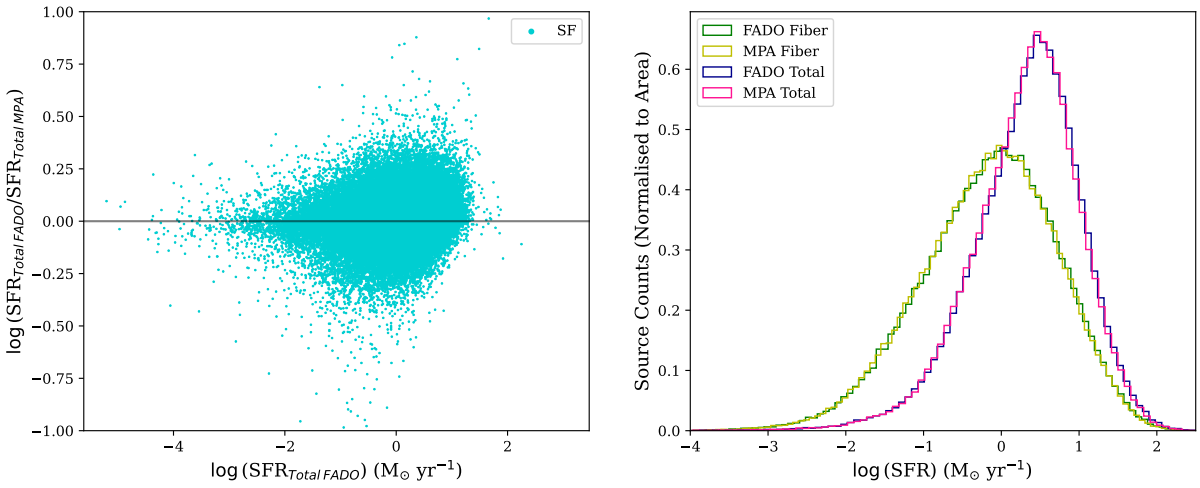

Figure 6.1: *Left:* Comparison of the total SFRs between FADO and MPA-JHU, in function of the total SFR of FADO, for the SF galaxies defined in Table 3.9. *Right:* Histograms that compare the fiber and total SFRs of FADO and MPA-JHU, for same galaxies on the left panel.

When it comes to Composites and AGNs, the emission line flux of Hα does not solely originate from SF regions, and can have contamination from the AGN in their cores. With this in mind, Figure 6.2 compares the SFR of these galaxies with the values obtained in equation 6.1 and the values in the

MPA-JHU dataset, which takes into account the relationship between the D4000 (which is a proxy for the age of a galaxy) and the specific SFR (the division between SFR and stellar mass of a galaxy) for every galaxy.

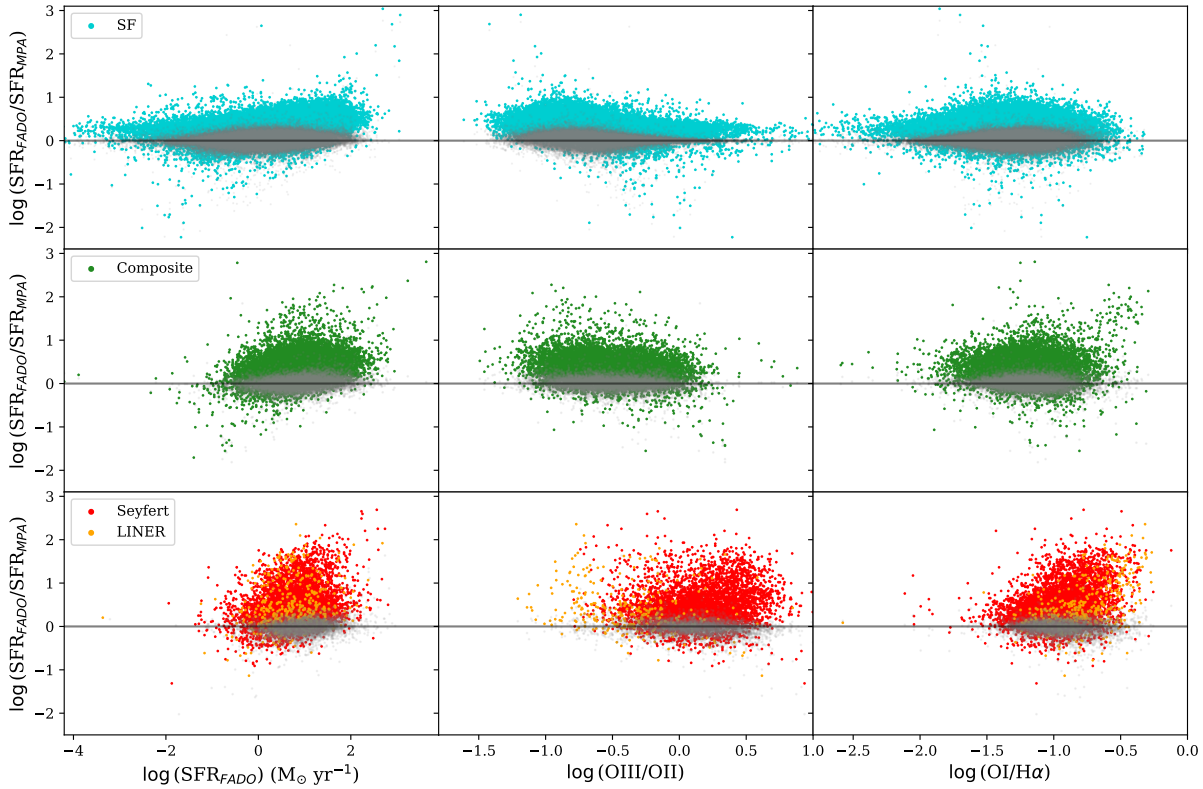

Figure 6.2: Comparison between the total SFR calculated with equation 6.1 and the total MPA-JHU SFR, for the SF, Composites and AGNs (divided in Seyfert and LINER according to the BPT diagram) defined in Table 3.9. The left column represents the comparison with the SFR_{FADO}, the middle column the comparison with $[OIII]\lambda5007/[OII]\lambda\lambda3726,3729$ and the right column the comparison between $[OI]\lambda6300/H\alpha$. The grey dots represent the comparison between FADO and MPA-JHU if we used the Hα emission line flux from both datasets to calculate the SFR.

The discrepancies between these estimates of the SFRs are much higher across all categories, with the Composites and AGNs reaching higher values than the SF galaxies. This is expected, since the Hα emission line fluxes of Composites and AGNs are contaminated by the presence of an AGN, something that isn't present in the SF galaxies. To keep consistency with previous studies of FADO (Miranda et al., 2023), we are going to use the SFR calculated from equation 6.1 for the SF galaxies, instead of the MPA-JHU dataset SFR.

When it comes to Composites and AGNs, through the comparison between $[OIII]\lambda5007/[OII]\lambda\lambda3726,3729$ - a proxy for the ionisation parameter of the gas inside a galaxy - and $[OI]\lambda6300/H\alpha$ - a proxy for the hardness of the ionisation field - it is clear to see from Figure 6.2 that AGNs have higher ionising parameters and harder ionising fields than Composites, and are on par with some SF galaxies. However, it is of note that the differences between SFRs is higher for AGNs for higher ionising fields (namely in the Seyfert-type galaxies) and that when we use only the Hα emission line flux to calculate the SFR, these discrepancies do not appear in either the Composites or the AGNs. It is also of note that, despite Composites having ionising fields on par with the SF, the SFR discrepancies are higher than the SF galaxies.

Furthermore, when comparing the D4000 and the Light-weighted age of these galaxies - an estimation of the age of a galaxy taking into account their luminosity - Figure 6.3 shows that higher D4000

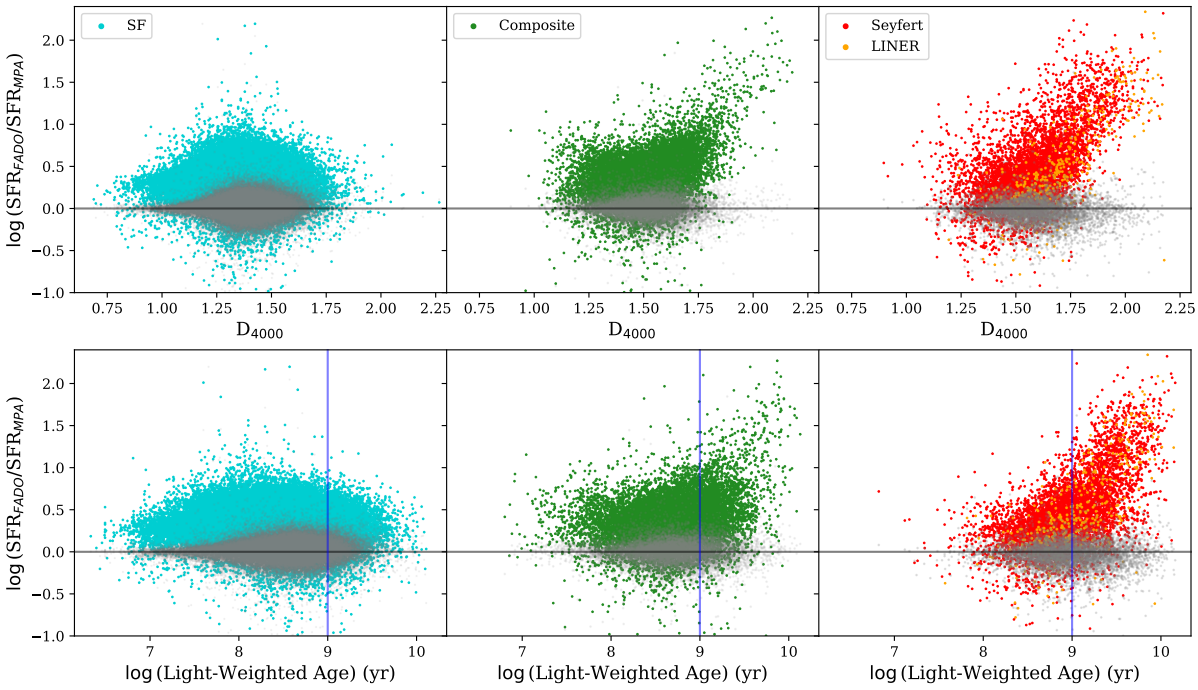

Figure 6.3: Top row represents the comparison between the total SFR in FADO according to equation 6.1 and the total SFR from the MPA-JHU dataset, compared to the D4000 of each galaxy. Bottom row represents the same, but compared with the light-weighted age of each galaxy, with the blue line representing galaxies with 1 Gyr. The gray dots represent the same comparison but with the SFR estimation from 6.1 with the MPA-JHU data.

and older Composites and AGNs (namely, after the 1 Gyr mark) have higher discrepancies when using the MPA-JHU dataset SFR. To compound even more evidence in favour of the MPA-JHU dataset SFR for Composites and AGNs, when using only the Hα estimated SFR, the discrepancies are not present, as when we compared the ionising fields.

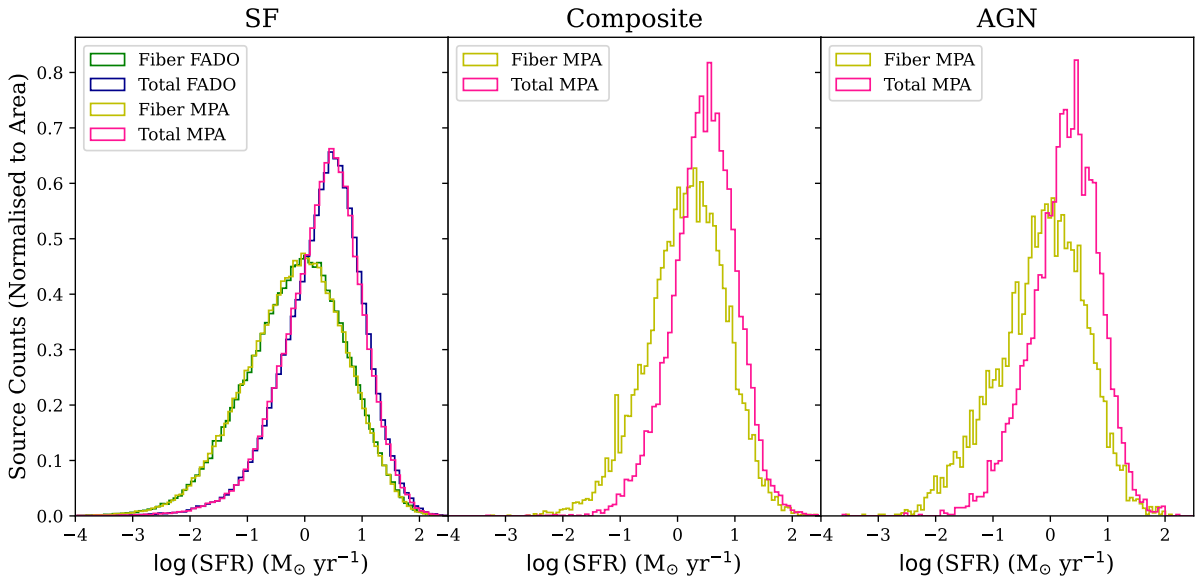

Figure 6.4: Histograms that compare the fiber and total SFR of the SF (left), Composites (middle) and AGN (right) galaxies.

This explains why solely using the emission line flux of Hα is insufficient to characterise the SFR of a galaxy that has an AGN present: the ionisation of the gas does not only come from active star-

forming regions in galaxies, and older galaxies have significantly less SF activity than younger galaxies. Therefore, for Composites and AGNs, we will use the MPA-JHU estimations of the SFR.

The comparison across all the classification of galaxies between the fiber and total SFR can be seen in Figure 6.4. Although the trend of the total SFR being higher than the fiber SFR remains the same, as expected, they are only 2.5 times higher for Composites and two times higher for AGNs. This tracks, since the differences between the fiber and total SFR in these galaxies should be lower as they form less stars than the SF galaxies.

6.1.2 Stellar Mass

FADO automatically estimates the stellar mass of galaxies through the fitting process of the SED. MPA-JHU follows a process similar to Kauffmann et al. (2003) and Salim et al. (2007), where the stellar mass is estimated from the photometry, rather than the spectroscopy. The MPA-JHU website discusses in more detail the differences between using spectral indices (such as the D4000 and the Hδ emission line absorption) and photometric bands (such as the SDSS u, g, r, i and z bands), but the conclusion is that, overall, the differences between these methods are very small (median of 2%).

Before we compare the stellar mass of these two datasets, we must once again take into account that they processed the data using different IMFs. From Speagle et al. (2014), we know that:

$$M_{\text{Chabrier}} = \frac{1}{1.06} M_{\text{Kroupa}} \tag{6.6}$$

where M_{Kroupa} is the stellar mass estimated from MPA-JHU and M_{Chabrier} the stellar mass estimated from FADO. Since the stellar masses are usually in logarithmic scale, then:

$$\log(M_{\text{Chabrier}}) = \log(M_{\text{Kroupa}}) + \log(0.943) \tag{6.7}$$

To ensure that the stellar masses are comparable, we must add $\log(0.943)$ to the logarithmic stellar mass of MPA-JHU. Furthermore, just as the SFR, we used the ratio between total and fiber stellar mass of the MPA-JHU catalogue to define the conversion factor between the fiber and total stellar masses in FADO.

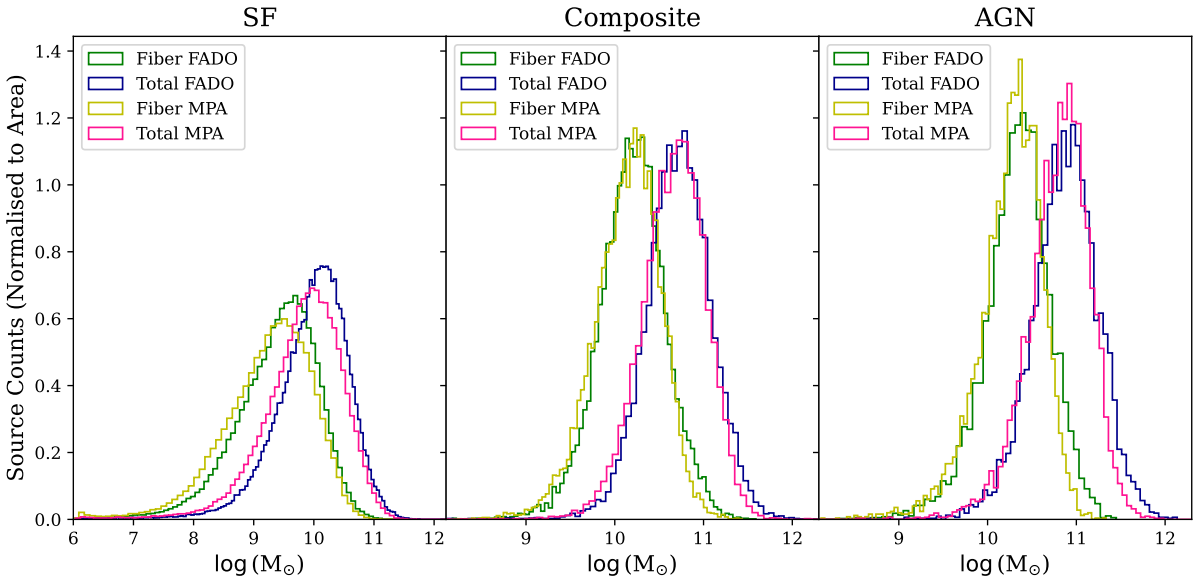

Figure 6.5: Histograms that compare the fiber and total stellar mass of the SF (left), Composites (middle) and AGN (right) galaxies.

Figure 6.5 shows that for the SF galaxies, the differences between the fiber stellar masses for both FADO and MPA-JHU are $\approx 42\%$, which is maintained for the total stellar masses. These differences are less notable for the Composites and AGNs, having $\approx 13\%$ and $\approx 10\%$ differences, respectively, for both the fiber and total stellar masses. Overall, FADO estimates higher stellar masses than MPA-JHU. The differences in the SF galaxies can be explained due to the uncertainties of each dataset (Table 6.2), which is in accordance with the results of Miranda et al. (2023). Since the information FADO provide about the uncertainties of the stellar mass are the formal errors from the minimisation algorithm and not the physical ones, we decided to use 0.2 dex as a conservative estimate for this value.

6.2 The main sequence

We plotted the main sequence in three samples: the BPT sample without mismatched classifications as defined in Chapter 3, the sample that considers the same but for all diagrams in Chapter 4, and the FADO Ambiguous sample from the same chapter. Figure 6.6 shows all of these, Table 6.1 shows the differences between FADO and MPA-JHU for both the BPT and all diagrams sample and Table 6.2 its associated uncertainties.

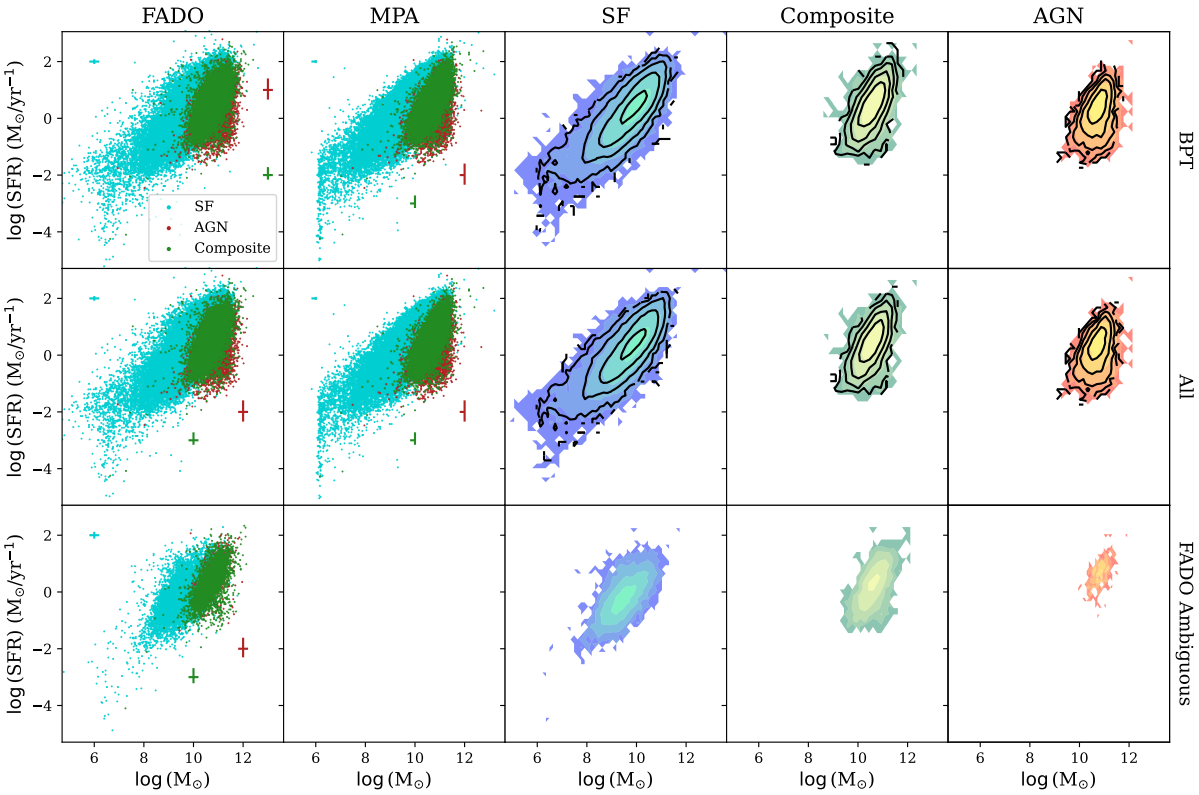

Figure 6.6: *Top row:* Main sequence for the galaxies with common classification defined in Table 3.9, for FADO and MPA-JHU. In the contours, FADO represents the colours and MPA-JHU the black lines. *Middle row:* The same as previously, only with the galaxies with common classification as defined in Table 4.5. *Bottom row:* FADO Ambiguous galaxies as defined in Table 4.7.

Overall, the differences between SFRs are extremely low across both samples, with a maximum difference of $\approx 4\%$, namely in the SF galaxies. This makes sense, as we've seen from Figure 6.4, the SFRs for all galaxies are essentially identical, with the Composites and AGNs having the same SFR for both datasets, as we chose to do. When it comes to the stellar mass, however, the differences become more pronounced, although they remain consistent across classifications. SF galaxies have the highest

Table 6.1: Median differences between FADO and MPA-JHU, in dex, for both the galaxies with common classification defined in Table 3.9 (here represented by the BPT rows) and the galaxies with common classification as defined in Table 4.5 (here represented by the All rows).

		FADO−MPA-JHU		
Sample	Criteria	SF	Composite	AGN
BPT	SFR	0.015	−0.004	0
	$M_\odot$	0.151	0.042	0.053
All	SFR	0.017	0	0
	$M_\odot$	0.155	0.043	0.051

Table 6.2: Uncertainties associated with FADO and MPA-JHU in each classification, for the SFR and the stellar mass, in dex. Each sample is also described in the chapter is also represented, as in Table 6.1.

		FADO			MPA-JHU		
Sample	Criteria	SF	Composite	AGN	SF	Composite	AGN
BPT	SFR	0.084	$^{+0.281}_{-0.160}$	$^{+0.406}_{-0.340}$	0.054	$^{+0.281}_{-0.160}$	$^{+0.406}_{-0.340}$
	$M_\odot$	0.2	0.2	0.2	$^{+0.17}_{-0.175}$	$^{+0.014}_{-0.196}$	$^{+0.013}_{-0.198}$
All	SFR	0.081	$^{+0.280}_{-0.157}$	$^{+0.407}_{-0.341}$	0.053	$^{+0.280}_{-0.157}$	$^{+0.407}_{-0.341}$
	$M_\odot$	0.2	0.2	0.2	$^{+0.017}_{-0.175}$	$^{+0.014}_{-0.196}$	$^{+0.013}_{-0.198}$

differences when it comes to stellar masses ($\approx 42\%$), which is in line with the histograms in Figure 6.5. AGNs have the second lowest difference, ($\approx 13\%$), although they are similar to the Composites ($\approx 10\%$), much like the previously mentioned histograms. For all classifications, the uncertainties compensate for these differences, meaning that FADO and MPA-JHU's stellar masses overlap.

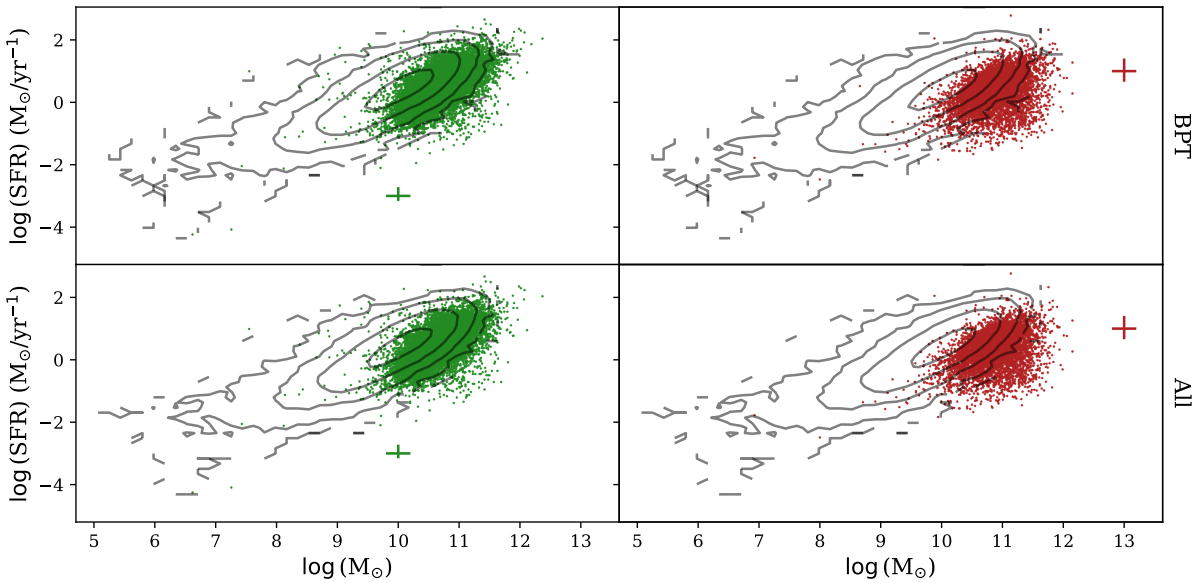

Figure 6.7: Main sequence of both the BPT sample (top row) and the sample with all the diagrams (bottom row), as explained in Figure 6.6. The black contours represent the SF sample, the left column the Composites and the right column the AGNs.

It is of note that we do not see any bimodality in the main sequence when it comes to the different classifications. This can be attributed to the S/N cut that we performed in the selection of our sample, as the bimodality in the flux was also cut. This translated into working with galaxies that have AGNs and high SFRs simultaneously, even with the MPA-JHU considerations of the D4000 and the age of the galaxies. However, there are some AGNs and Composites that do not reside in the SF area, as we can

see from Figure 6.7 (using FADO only, as a demonstrative point). Although it is not a strong separation, there are still some notable differences, namely in the AGNs, due to their higher stellar masses.

In conclusion, the addition of self-consistent nebular emission to a spectral fitting model does not significantly change the position of galaxies inside the main sequence, when compared to purely stellar models, which is a trend that we have come to see across all classification schemes we have compared. Specifically, in the star-forming main sequence, our results are in agreement with a previous study by Miranda et al. (2023), where there was a thorough investigation of the SF sample.

6.2.1 Ambiguous galaxies

A look at the Ambiguous galaxies is interesting, as comparing their position in the main sequence versus where they "should" be could tell us more about their properties. However, because all of the classifications are quite close together, and specifically because the uncertainties in the stellar mass are large, we cannot certainly say that, even if they are outside the locations where they "should" be, it does not necessarily mean that it is an incorrect classification.

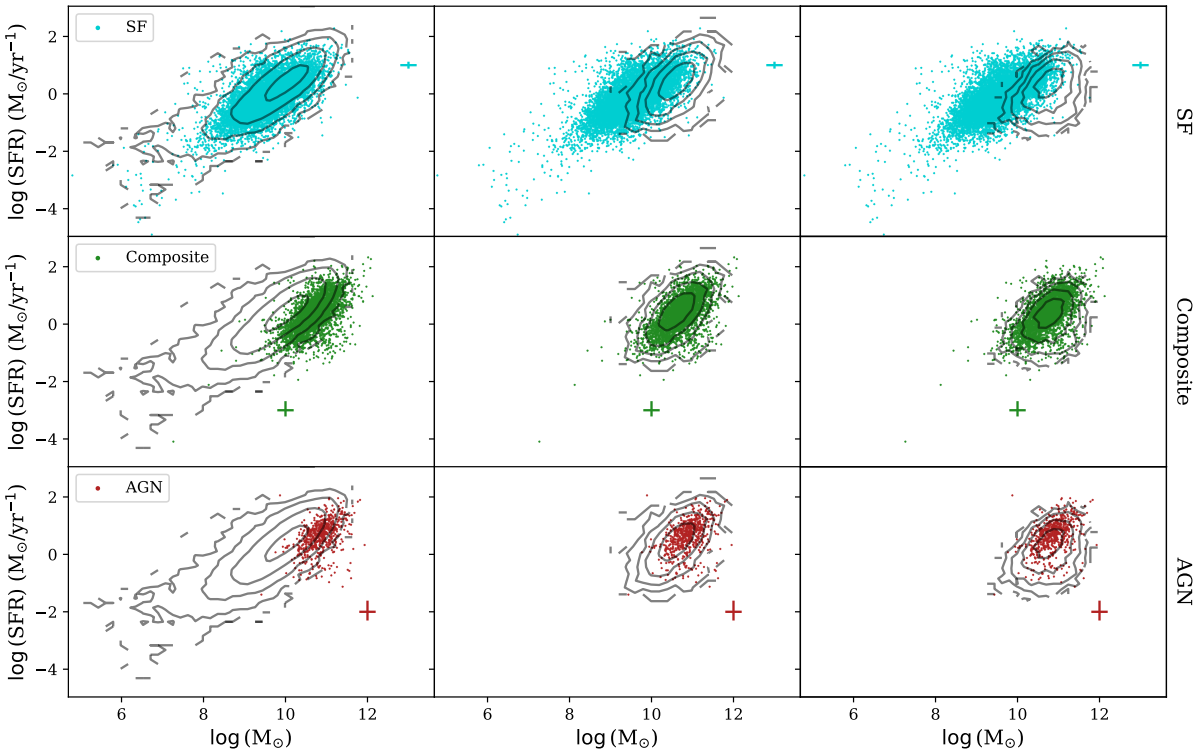

Figure 6.8: Main sequence of both the FADO sample with all the diagrams (black contours) and the Ambiguous sample (coloured dots). The left column represents the SF contours, the middle column the Composite contours and the right column the AGN contours, with each Ambiguous classification represented in each row.

By the nature of their definition, Ambiguous galaxies are misclassified in one or more diagrams (see Chapter 4). Figure 6.8 shows the position of these galaxies when compared with the contours of the full sample of galaxies. When it comes to SF, all galaxies fit well into their designated area, delineated by the contours. Composites and AGNs behave similarly, with only 2 Composites and 1 AGN escaping beyond the uncertainties into the SF area of the diagram. Overall, Ambiguous galaxies behave similarly to their classification in the main sequence as the non-Ambiguous ones, meaning we cannot extract many conclusions when looking at this diagram only, and we need the BPT, SII, OI or the EW diagrams to understand the true nature of these galaxies.

Chapter 7

Conclusion

Over the course of this work, we compared FADO, an SPS model that considers the stellar and nebular contributions of galaxies, with MPA-JHU, a model of the same type that only considers the stellar contribution. These models were applied to SDSS DR7, to understand if the self-consistent addition of the nebular component affects galaxy classification through emission line ratio diagrams. Furthermore, we found a new classification diagram and analysed it. Finally, we briefly looked at how these galaxies were placed in the main sequence.

We found that, regarding the BPT diagram emission lines, when comparing FADO with MPA-JHU, the former dataset estimates higher fluxes for Hβ (55%) and [OIII]λ5007 (58%), than for Hα (19%) and [NII]λ6584 (22%). These differences can be justified due to the different extinction laws that each dataset used. It is also of note that the differences in flux between datasets are compensated by the uncertainties.

The EW of these emission lines also suffer from this effect. FADO estimates higher EWs than MPA-JHU, again in the Hβ (75%) and [OIII]λ5007 (38%) emission lines, rather than the Hα (29%) and [NII]λ6584 (90%) emission lines, for the same reason as mentioned above. It is of note that the differences are more pronounced in the Balmer lines than the forbidden lines, when compared with the fluxes. This is due to the fact that FADO measures the continuum from the bottom of the absorption line, if it exists. Since the Balmer lines have them and the forbidden lines do not, then this explains why this happens. Nevertheless, the uncertainties compensate for the differences between datasets.

After correcting emission lines for line-of-sight extinction and ensuring that both dataset's galaxies followed the classification dictated by equations 3.9, 3.10 and 3.11, we found that 98% of the galaxies between FADO and MPA-JHU had the same classification, and 2% did not. The latter galaxies are all placed near the border and their uncertainties can place them in the "correct" region of the BPT diagram, so we discarded them from the analysis. The other galaxies have minute differences between datasets (a maximum of 10%), to which the uncertainties more than compensate for. In other words, the addition of self-consistent nebular emission does not significantly change the classification of galaxies through the BPT diagram.

When comparing the BPT diagram classification with the SII and OI diagrams classifications, we found that, for FADO, 11% of the galaxies are considered Ambiguous. Of these, we found that the Ambiguous SF galaxies were 95% more metal-poor and 223% less massive than the full SF sample, meaning that galaxies that have these characteristics are more likely to be misclassified by the BPT diagram. We compared these galaxies with WISE data, and found that the BPT diagram and WISE classifications were in agreement. This means that there might be a bias in these classification schemes, where galaxies with low mass and sub-solar metallicity that have a weak AGN in their core are consistently misclassified by these diagrams. We also found that Ambiguous Composites and Ambiguous AGNs are misclassified due

to the different metallicity sensitivity between the three diagrams.

The differences between non-Ambiguous galaxies in the BPT and SII diagrams are relatively small (a maximum of 4%), with bigger differences in the OI diagram (maximum of 22%). These differences in percentages are expected since we did not restrict the S/N of the [OI]λ6300 emission line. Furthermore, the uncertainties compensate for these values. As previously, we see that self-consistent addition of nebular contributions to the SED fitting does not affect galaxy classification through emission line ratio diagrams.

Comparing the EWs of the Balmer lines (Hα and Hβ) with the ratio of the EWs of [OIII]λ5007 and Hβ, galaxies naturally separated themselves into two types. According to the BPT diagram classification, the SF and Composite galaxies are mixed in one cluster, while the AGNs are in another. We defined functions of separations for these diagrams and found that the EWB diagram was the best at separating galaxies. We compared this diagram with the BPT, SII and OI diagrams and found that contaminated galaxies (or galaxies that crossed over the function of separation in the EWB diagram) and Ambiguous galaxies were correlated: 84% of contaminated SF galaxies and 92% of contaminated Composite galaxies matched with Ambiguous SF and Composites. Contaminated AGNs were the exception, with only a 43% match with Ambiguous AGNs, but this is due to the way we defined these galaxies as Ambiguous. This means that contamination, overall, is a good indicator for ambiguity.

These EW diagrams can have several uses. For example: we do not need four emission lines to plot a classification diagram, as the EWB diagram only needs two of them; since contamination is a good indicator of ambiguity, we do not need five emission lines to understand if certain galaxies are Ambiguous or not; and these diagrams can also be useful to classify galaxies in the case we cannot access neither the Hα nor [NII]λ6584 emission lines (for example, the MOONRISE survey at high redshifts).

The SFR when based solely on Hα luminosity and when based on it and the D4000 index were different. We found that the latter was a better fit for Composites and AGNs and the former better for SF galaxies. We also found that stellar mass estimations were different between datasets (42% difference for SF galaxies, 13% for Composites and 10% for AGNs), but the uncertainties compensated for their different values, which is in line with previous results.

Comparing the Main Sequence of galaxies between FADO and MPA-JHU yielded the same results, even with the high differences in stellar mass, due to the uncertainties present. Again, FADO does not have any meaningful differences in all classifications of galaxies when compared to MPA-JHU, much like the emission line ratio diagrams. Of note is that we also compared the Ambiguous galaxies and found that the BPT diagram classification matches with their main sequence placement, so the main sequence cannot be used to evaluate ambiguity.

In the future, we plan on further comparing the sub-solar metallicity and low mass galaxy population found in the Ambiguous SF sample to better understand their nature, i.e., to understand if it is an AGN that drives the differences between emission line ratio diagrams or, if it is something else, what it is and how we can identify it.

Bibliography

Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., & et al. (2009). The Seventh Data Release of the Sloan Digital Sky Survey. ApJS, *182*(2), 543–558. https://doi.org/10.1088/0067-0049/182/2/543

Abdurro'uf, Accetta, K., Aerts, C., & et al. (2022). The Seventeenth Data Release of the Sloan Digital Sky Surveys: Complete Release of MaNGA, MaStar, and APOGEE-2 Data. ApJS, *259*(2), Article 35, 35. https://doi.org/10.3847/1538-4365/ac4414

Amorín, R., Pérez-Montero, E., Vílchez, J. M., & Papaderos, P. (2012). The Star Formation History and Metal Content of the Green Peas. New Detailed GTC-OSIRIS Spectrophotometry of Three Galaxies. ApJ, *749*(2), Article 185, 185. https://doi.org/10.1088/0004-637X/749/2/185

Baldwin, J. A., Phillips, M. M., & Terlevich, R. (1981). Classification parameters for the emission-line spectra of extragalactic objects. PASP, *93*, 5–19. https://doi.org/10.1086/130766

Beckmann, V., & Shrader, C. R. (2012). *Active Galactic Nuclei*.

Belfiore, F., Maiolino, R., Maraston, C., & et al. (2016). SDSS IV MaNGA - spatially resolved diagnostic diagrams: a proof that many galaxies are LIERs. MNRAS, *461*(3), 3111–3134. https://doi.org/10.1093/mnras/stw1234

Blanton, M. R., Bershady, M. A., Abolfathi, B., & et al. (2017). Sloan Digital Sky Survey IV: Mapping the Milky Way, Nearby Galaxies, and the Distant Universe. AJ, *154*(1), Article 28, 28. https://doi.org/10.3847/1538-3881/aa7567

Bond, N. A., Benford, D. J., Gardner, J. P., & et al. (2012). The Infrared Properties of Sources Matched in the WISE All-sky and Herschel ATLAS Surveys. ApJ, *750*(1), Article L18, L18. https://doi.org/10.1088/2041-8205/750/1/L18

Bressan, A., Marigo, P., Girardi, L., & et al. (2012). PARSEC: stellar tracks and isochrones with the PAdova and TRieste Stellar Evolution Code. MNRAS, *427*(1), 127–145. https://doi.org/10.1111/j.1365-2966.2012.21948.x

Brinchmann, J., Charlot, S., White, S. D. M., & et al. (2004). The physical properties of star-forming galaxies in the low-redshift Universe. MNRAS, *351*(4), 1151–1179. https://doi.org/10.1111/j.1365-2966.2004.07881.x

Calzetti, D., Armus, L., Bohlin, R. C., & et al. (2000). The Dust Content and Opacity of Actively Star-forming Galaxies. ApJ, *533*(2), 682–695. https://doi.org/10.1086/308692

Cardoso, L. S. M., Gomes, J. M., Papaderos, P., Pappalardo, C., Miranda, H., Paulino-Afonso, A., Afonso, J., & Lagos, P. (2022). Revisiting stellar properties of star-forming galaxies with stellar and nebular spectral modelling. *arXiv e-prints*, Article arXiv:2208.14036, arXiv:2208.14036.

Chabrier, G. (2003). Galactic Stellar and Substellar Initial Mass Function. PASP, *115*(809), 763–795. https://doi.org/10.1086/376392

Charlot, S., & Fall, S. M. (2000). A Simple Model for the Absorption of Starlight by Dust in Galaxies. ApJ, *539*(2), 718–731. https://doi.org/10.1086/309250

Chary, R., & Elbaz, D. (2001). Interpreting the Cosmic Infrared Background: Constraints on the Evolution of the Dust-enshrouded Star Formation Rate. ApJ, *556*(2), 562–581. https://doi.org/10.1086/321609

Chiu, H.-Y. (1964). Gravitational Collapse. *Physics Today*, *17*(5), 21. https://doi.org/10.1063/1.3051610

Cid Fernandes, R., Stasińska, G., Mateus, A., & Vale Asari, N. (2011). A comprehensive classification of galaxies in the Sloan Digital Sky Survey: how to tell true from fake AGN? MNRAS, *413*(3), 1687–1699. https://doi.org/10.1111/j.1365-2966.2011.18244.x

Conroy, C. (2013). Modeling the Panchromatic Spectral Energy Distributions of Galaxies. ARA&A, *51*(1), 393–455. https://doi.org/10.1146/annurev-astro-082812-141017

Cutri, R. M., Wright, E. L., Conrow, T., Fowler, J. W., & et al. (2021). VizieR Online Data Catalog: AllWISE Data Release (Cutri+ 2013). *VizieR Online Data Catalog*, Article II/328, II/328.

da Cunha, E., Charlot, S., & Elbaz, D. (2008). A simple model to interpret the ultraviolet, optical and infrared emission from galaxies. MNRAS, *388*(4), 1595–1617. https://doi.org/10.1111/j.1365-2966.2008.13535.x

De Robertis, M. M., Dufour, R. J., & Hunt, R. W. (1987). A five-level program for ions of astrophysical interest. JRASC, *81*, 195–220.

Eales, S., de Vis, P., Smith, M. W. L., & et al. (2017). The Galaxy End Sequence. MNRAS, *465*(3), 3125–3133. https://doi.org/10.1093/mnras/stw2875

Ferland, G. J., Korista, K. T., Verner, D. A., & et al. (1998). CLOUDY 90: Numerical Simulation of Plasmas and Their Spectra. PASP, *110*(749), 761–778. https://doi.org/10.1086/316190

Fernández, V., Amorín, R., Pérez-Montero, E., & et al. (2022). New insights on the nebular emission, ionizing radiation, and low metallicity of Green Peas from advanced modelling. MNRAS, *511*(2), 2515–2534. https://doi.org/10.1093/mnras/stab3150

Fioc, M., & Rocca-Volmerange, B. (1997). PEGASE: a UV to NIR spectral evolution model of galaxies. Application to the calibration of bright galaxy counts. A&A, *326*, 950–962.

Fitzpatrick, E. L. (2004). Interstellar Extinction in the Milky Way Galaxy. In A. N. Witt, G. C. Clayton, & B. T. Draine (Eds.), *Astrophysics of dust* (p. 33).

Gomes, J. M., & Papaderos, P. (2017). Fitting Analysis using Differential evolution Optimization (FADO):. Spectral population synthesis through genetic optimization under self-consistency boundary conditions. A&A, *603*, Article A63, A63. https://doi.org/10.1051/0004-6361/201628986

Groves, B. A., Dopita, M. A., & Sutherland, R. S. (2004). Dusty, Radiation Pressure-Dominated Photoionization. I. Model Description, Structure, and Grids. ApJS, *153*(1), 9–73. https://doi.org/10.1086/421113

Heckman, T. M. (1980). An Optical and Radio Survey of the Nuclei of Bright Galaxies - Activity in the Normal Galactic Nuclei. A&A, *87*, 152.

Husser, T. .-O., Kamann, S., Dreizler, S., & Hauschildt, P. H. (2012). A new extensive library of synthetic stellar spectra from PHOENIX atmospheres and its application to fitting VLT MUSE spectra. In P. Prugniel & H. P. Singh (Eds.), *Astronomical society of india conference series* (p. 71).

Kauffmann, G., Heckman, T. M., Tremonti, C., & et al. (2003). The host galaxies of active galactic nuclei. MNRAS, *346*(4), 1055–1077. https://doi.org/10.1111/j.1365-2966.2003.07154.x

Kennicutt, J., Robert C. (1998). Star Formation in Galaxies Along the Hubble Sequence. ARA&A, *36*, 189–232. https://doi.org/10.1146/annurev.astro.36.1.189

Kewley, L. J., Dopita, M. A., Sutherland, R. S., & et al. (2001). Theoretical Modeling of Starburst Galaxies. ApJ, *556*(1), 121–140. https://doi.org/10.1086/321545

Kewley, L. J., Groves, B., Kauffmann, G., & et al. (2006). The host galaxies and classification of active galactic nuclei. MNRAS, *372*(3), 961–976. https://doi.org/10.1111/j.1365-2966.2006.10859.x

Kroupa, P. (2001). On the variation of the initial mass function. MNRAS, *322*(2), 231–246. https://doi.org/10.1046/j.1365-8711.2001.04022.x

Kroupa, P., Weidner, C., Pflamm-Altenburg, J., & et al. (2013). The stellar and sub-stellar initial mass function of simple and composite populations. *Planets, stars and stellar systems* (pp. 115–242). Springer Netherlands. https://doi.org/10.1007/978-94-007-5612-0_4

Leitherer, C., Schaerer, D., Goldader, J. D., & et al. (1999). Starburst99: Synthesis Models for Galaxies with Active Star Formation. ApJS, *123*(1), 3–40. https://doi.org/10.1086/313233

Lovell, C. C., Acquaviva, V., Thomas, P. A., & et al. (2019). Learning the relationship between galaxies spectra and their star formation histories using convolutional neural networks and cosmological simulations. MNRAS, *490*(4), 5503–5520. https://doi.org/10.1093/mnras/stz2851

Maiolino, R., Cirasuolo, M., Afonso, J., & et al. (2020). MOONRISE: The Main MOONS GTO Extragalactic Survey. *The Messenger*, *180*, 24–29. https://doi.org/10.18727/0722-6691/5197

Marigo, P., Girardi, L., Bressan, A., Groenewegen, M. A. T., Silva, L., & Granato, G. L. (2008). Evolution of asymptotic giant branch stars. II. Optical to far-infrared isochrones with improved TP-AGB models. A&A, *482*(3), 883–905. https://doi.org/10.1051/0004-6361:20078467

Marigo, P., Bressan, A., Nanni, A., & et al. (2013). Evolution of thermally pulsing asymptotic giant branch stars - I. The COLIBRI code. MNRAS, *434*(1), 488–526. https://doi.org/10.1093/mnras/stt1034

Marshall, F. E., Mushotzky, R. F., Boldt, E. A., Holt, S. S., Rothschild, R. E., & Serlemitsos, P. J. (1978). N Galaxies-a new class of X-ray sources. Nature, *275*(5681), 624–625. https://doi.org/10.1038/275624a0

Miranda, H., Pappalardo, C., Papaderos, P., Afonso, J., Matute, I., Lobo, C., Paulino-Afonso, A., Carvajal, R., Lorenzoni, S., & Santos, D. (2023). An investigation of the star-forming main sequence considering the nebular continuum emission at low-z. A&A, *669*, Article A16, A16. https://doi.org/10.1051/0004-6361/202244390

Noeske, K. G., Weiner, B. J., Faber, S. M., & et al. (2007). Star Formation in AEGIS Field Galaxies since z=1.1: The Dominance of Gradually Declining Star Formation, and the Main Sequence of Star-forming Galaxies. ApJ, *660*(1), L43–L46. https://doi.org/10.1086/517926

Offner, S. S. R., Clark, P. C., Hennebelle, P., & et al. (2014). The Origin and Universality of the Stellar Initial Mass Function. In H. Beuther, R. S. Klessen, C. P. Dullemond, & T. Henning (Eds.), *Protostars and planets vi* (p. 53). https://doi.org/10.2458/azu_uapress_9780816531240-ch003

Osterbrock, D. E., & Ferland, G. J. (2006). *Astrophysics of gaseous nebulae and active galactic nuclei*.

Rodighiero, G., Daddi, E., Baronchelli, I., & et al. (2011). The Lesser Role of Starbursts in Star Formation at z = 2. ApJ, *739*(2), Article L40, L40. https://doi.org/10.1088/2041-8205/739/2/L40

Salim, S., Rich, R. M., Charlot, S., & et al. (2007). UV Star Formation Rates in the Local Universe. ApJS, *173*(2), 267–292. https://doi.org/10.1086/519218

Salpeter, E. E. (1955). The Luminosity Function and Stellar Evolution. ApJ, *121*, 161. https://doi.org/10.1086/145971

Sánchez-Blázquez, P., Peletier, R. F., Jiménez-Vicente, J., & et al. (2006). Medium-resolution Isaac Newton Telescope library of empirical spectra. MNRAS, *371*(2), 703–718. https://doi.org/10.1111/j.1365-2966.2006.10699.x

Seyfert, C. K. (1943). Nuclear Emission in Spiral Nebulae. ApJ, *97*, 28. https://doi.org/10.1086/144488

Shaw, R. A., & Dufour, R. J. (1994). The FIVEL Nebular Modelling Package in STSDAS. In D. R. Crabtree, R. J. Hanisch, & J. Barnes (Eds.), *Astronomical data analysis software and systems iii* (p. 327).

Shaw, R. A., & Dufour, R. J. (1995). Software for the Analysis of Emission Line Nebulae. PASP, *107*, 896. https://doi.org/10.1086/133637

Speagle, J. S., Steinhardt, C. L., Capak, P. L., & Silverman, J. D. (2014). A Highly Consistent Framework for the Evolution of the Star-Forming "Main Sequence" from z ∼0-6. ApJS, *214*(2), Article 15, 15. https://doi.org/10.1088/0067-0049/214/2/15

Teimoorinia, H., Archinuk, F., Woo, J., & et al. (2022). Mapping the Diversity of Galaxy Spectra with Deep Unsupervised Machine Learning. AJ, *163*(2), Article 71, 71. https://doi.org/10.3847/1538-3881/ac4039

Tremonti, C. A., Heckman, T. M., Kauffmann, G., & et al. (2004). The Origin of the Mass-Metallicity Relation: Insights from 53,000 Star-forming Galaxies in the Sloan Digital Sky Survey. ApJ, *613*(2), 898–913. https://doi.org/10.1086/423264

Veilleux, S., & Osterbrock, D. E. (1987). Spectral Classification of Emission-Line Galaxies. ApJS, *63*, 295. https://doi.org/10.1086/191166

Walcher, J., Groves, B., Budavári, T., & et al. (2011). Fitting the integrated spectral energy distributions of galaxies. Ap&SS, *331*, 1–52. https://doi.org/10.1007/s10509-010-0458-z

Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., & et al. (2010). The Wide-field Infrared Survey Explorer (WISE): Mission Description and Initial On-orbit Performance. AJ, *140*(6), 1868–1881. https://doi.org/10.1088/0004-6256/140/6/1868