Multicomponent models of the interstellar medium in low-metallicity dwarf galaxies

Lise Ramambason

Université Paris Cité
École Doctorale d'Astronomie et Astrophysique d'Île-de-France (ED 127)
CEA, CNRS, AIM, F-91191, Gif-sur-Yvette, France

MULTICOMPONENT MODELS OF THE INTERSTELLAR MEDIUM IN LOW-METALLICITY DWARF GALAXIES

PAR LISE RAMAMBASON

THÈSE DE DOCTORAT EN ASTRONOMIE & ASTROPHYSIQUE

DIRIGÉE PAR VIANNEY LEBOUTEILLER

Présentée et soutenue publiquement le 1er Décembre 2022

DEVANT UN JURY COMPOSÉ DE:

VIANNEY LEBOUTEILLER	Directeur de thèse	Chercheur AIM, CEA Saclay
DENIS BURGARELLA	Rapporteur	Astronome Laboratoire d'Astrophysique de Marseille
GÖRAN ÖSTLIN	Rapporteur	Professeur Stockholm University
PIERRE GUILLARD	Examinateur	Maître de conférences Institut d'Astrophysique de Paris
AMÉLIE SAINTONGE	Examinatrice	Professeure University College London
ANNE VERHAMME	Examinatrice	Professeure assistante Observatoire de Genève
FRANCK LE PETIT	Président du jury	Astronome Observatoire de Paris

Acknowledgements / Remerciements

Thanks to all the jury members for their careful reading of the manuscript and their interesting questions during the defense, which open the way for more discussions and further developments of this thesis work. A special thanks to Göran Östlin for his detailed comments, which have contributed to improve the current version of the manuscript.

Thanks to Vianney for his guidance and help throughout the thesis. Thanks for the huge amount of time allocated to work with me in a very complementary and efficient way. I keep in mind our many discussions, which were precious keys to overcome obstacles regarding technical issues or physical interpretations. Thanks for always pushing me to develop my ideas and refine my interpretations a bit further, while also reminding me that interesting physics can still be done with models that we know are not perfect, and that it is ok to progress step by step. Discussing with you was always an excellent reminder of why we do what we do, and why we do it that way. Thanks for your kindness and support, including for non-work-related aspects.

Thanks to the NUAGES team in its extended version. This thesis work builds on excellent previous works carried out in the team, which have saved me a lot of time; in particular I want to acknowledge the work of Diane Cormier, Aurélie Rémy-Ruyer, Mélanie Chevance and Fiorella Polles, which were very useful to me. It also gives me hope that, maybe, this manuscript will be useful to someone, as their manuscripts and papers were to me.

Of course, I also want to thank the recent and current members of the group, with whom I had great interactions during our cookies-group-meetings, lunches, and conferences: thanks Sue, for great advices on scientific projects and life, and for taking care of everyone's well-being in the group, thanks Fred, for being an inspiration in the way you carry out meticulous and original projects and for your dedication to foster meaningful interactions between scientists through the organization of schools, seminars, and workshops. Thanks to all the other NUAGES – and affiliated – people for the shared moments: Dangning, Lara, Antigone, Maud, André, Fiorella, Pierre, Marc-Antoine, Gerardo, Frances, Axel, Jack, Corentin, Jean...

Thanks to my officemates over various timescales and teams: Maxime, Valentin, André, Lara and Shiying. A huge and special thanks to Boris, Michael, Fadi, and Sylvain for dragging me out of my office for breaks and pushing me to meet people. It has changed my thesis experience and I am infinitely grateful for that. Thanks for all the shared lunches, coffees, beers, and Master Chef dinners, but also for the swimming sessions in Orsay, two great trips in Marseille, including a diving baptism, an epic kayaking experience with the LCEG team, an everlasting D&D quest, occasional runs on the plateau, and bouldering sessions in Arkose. All of these featuring the awesome CEA crowd: Boris, Nacho, Michael, Sunayana, Rose, Vicky, Christian, Benjamin(s), Mengyuan, Carlos, Lucas, Lara (et al. !).

Thanks to the Grenade team (Noé, Frances, Adnan, and Tine) for an unforgettable week during my first post-covid winter school. Thanks to the friends I keep from this school and other meetings, I hope our path cross again in the future ! Thanks also to my scientific collaborators from abroad for the great projects and fruitful discussions.

Thanks to the teachers and scientists that have led me to doing research and studying the interstellar medium of galaxies: Alain Lumediluna, Mathieu Langer, Alain Albergel, Laurent Verstraete, Mika Juvela and Daniel Schaerer.

Côté français, merci aux marcheurs et marcheuses du CEA, mené.e.s par Marie, de m'avoir fait re-découvrir le plateau et la vallée et pour toutes ces belles marches partagées. Merci en particulier à Laura et Valérie pour leur aide précieuse dans l'organisation des missions, ainsi qu'à Marie et Hugues pour toutes les questions informatiques.

Merci à la Maison d'Initiation et de Sensibilisation aux Sciences (MISS) pour deux belles années de médiation scientifique. Merci à Valérie, Élisabeth, Sylvie, et à tou.te.s mes co-animateurs et co-animatrices pour les échanges bienveillants et enrichissants. Merci d'avoir élargi mon horizon et piqué ma curiosité sur des sujets éloignés de l'astrophysique !

Merci à toute l'équipe de la Guyane vers les Étoiles pour trois semaines inoubliables d'animations et de conférences sur la côte guyanaise autour du lancement du télescope Webb. Contente d'avoir vécu ce moment si particulier à vos côtés et d'avoir retrouvé l'équipe, presque au complet, à l'occasion de ma soutenance! Merci aussi à Miguel pour l'invitation à présenter mon travail au grand public lors du festival De la Plage aux Étoiles, à Collioure, qui a marqué une pause bienvenue et très appréciée dans la rédaction de ce manuscrit.

Merci à tous mes amis de Paris, d'Angers, du reste de la France, et de contrées lointaines (US, Vietnam, Suède...) que je garde au plus près de mon cœur. Merci en particulier à celles et ceux qui se sont déplacé.e.s jusqu'à Gif-sur-Yvette pour ma soutenance : Pascale & Philippe, Bruno, Samy, Maël, Maëlle, Simon, Sofiane, Morgane, Esther, Alain, Raphaël, Daniel, Lester, Mathilde, Achrène, Miguel et Marion.

Merci à Sélim, mon compagnon de route avec qui j'espère partager un encore un long bout de chemin et de nouvelles aventures franco-allemandes. Merci pour le soutien logistique et moral, pour 1001 discussions intéressantes, et pour avoir saupoudré, l'air de rien, de très bons arguments pour me pousser à m'intéresser aux applications du machine learning en astro.

Merci, enfin, à ma famille; à celles et ceux qui me donnent de la force au quotidien, qu'ils soient proches ou loin. En particulier, merci à celles qui ont pu être présentes pour ma soutenance : Marianne, Jihane, Camille et Jeanne.

Merci aux femmes inspirantes de ma famille qui m'ont infusée de leur goût pour la recherche et la transmission des savoirs. À ma grand-mère Micheline Baulant, chercheuse CNRS en histoire médiévale, à ma maman, Camille Baulant, enseignante-chercheuse en économie internationale et risque pays, à ma petite sœur, Jeanne Ramambason, et à ses futurs succès quelle que soit la voie qu'il lui conviendra de suivre.

CONTENTS

Table of content vi

List of figures viii

List of tables ix

Abstract 1

Résumé 3

Résumé étendu en français 11

Introduction 12

I Observational context & big questions 15

1 The multiphase ISM 16
- 1.1 Dynamical picture . . . 16
  - 1.1.1 From diffuse gas to star formation . . . 16
  - 1.1.2 Metal and dust enrichment . . . 18
- 1.2 ISM at thermal equilibrium . . . 19
  - 1.2.1 ISM phases . . . 20
  - 1.2.2 Thermal processes . . . 22
  - 1.2.3 Impact of metallicity on thermal equilibrium . . . 26

2 Low-metallicity dwarf galaxies near and far 31
- 2.1 Primordial dwarf galaxies . . . 31
  - 2.1.1 Galaxy evolution . . . 31
  - 2.1.2 Star formation in pristine gas . . . 32
  - 2.1.3 Primordial galaxies and the reionization of the universe . . . 33
- 2.2 Local dwarf galaxies . . . 34
  - 2.2.1 Definition . . . 35
  - 2.2.2 Morphological classification . . . 35
  - 2.2.3 Primitive chemistry . . . 36
  - 2.2.4 Impact of low-metallicity on the gas reservoirs . . . 38
  - 2.2.5 Local dwarf as analogs to primordial galaxies? . . . 40
- 2.3 Relevant samples of dwarf galaxies . . . 41
  - 2.3.1 Star-forming dwarf galaxies in the local universe . . . 41
  - 2.3.2 BCDs and Green Peas at intermediate redshift ($z \lesssim 1$) . . . 45
- 2.4 Focus of the thesis . . . 45

II Modeling the interstellar medium 47

3 State-of-the-art ISM models 48
- 3.1 Modeling tools . . . 48
  - 3.1.1 Stationary chemical models . . . 48
  - 3.1.2 1D RT codes . . . 49

3.1.3 3D codes . . . 50
3.1.4 Simulations . . . 52
3.2 Topological models for unresolved galaxies . . . 52
3.2.1 Iterative improvements of topological models . . . 54
3.2.2 Limitations of the χ^2 method . . . 56

4 Defining flexible topological models **58**
4.1 Configurations . . . 58
4.2 Single cluster, multisector . . . 59
4.3 Multicluster, multisector . . . 60
4.4 Distributions of parameters . . . 60

5 Grid of models SFGX **62**
5.1 Overview . . . 62
5.2 Primary parameters . . . 64
5.2.1 Incident radiation fields . . . 66
5.2.2 Metals and dust . . . 69
5.2.3 Geometry of H II regions . . . 74
5.3 Secondary parameters . . . 79
5.3.1 Extensive parameters vs. intensive parameters . . . 79
5.3.2 Important secondary parameters . . . 80

III Tools **84**

6 Bayesian statistics and ISM studies **85**
6.1 Bayesian framework . . . 85
6.2 Interest of sampling methods . . . 87

7 `MULTIGRIS`: a new Bayesian multidimensional grid search **90**
7.1 General structure . . . 90
7.2 Model definition . . . 92
7.2.1 Lines and upper limits . . . 92
7.2.2 Scaling factor . . . 92
7.2.3 Extinction law . . . 93
7.2.4 Priors . . . 94
7.2.5 Systematic uncertainties . . . 95
7.3 MCMC sampling . . . 96
7.3.1 Choice of the sampler . . . 96
7.3.2 Sequential Monte Carlo sampler (SMC) . . . 98
7.4 Processing the inference results . . . 100
7.4.1 Component identification . . . 100
7.4.2 A posteriori checks . . . 101
7.4.3 Visualizing PDFs for individual objects . . . 102
7.4.4 Visualizing PDFs for a sample of objects . . . 103
7.5 Model comparison . . . 111
7.5.1 Predictive power . . . 111
7.5.2 Inferring the optimal number of components . . . 113

IV Results **114**

8 Application to the ISM **115**
8.1 Benchmarks on mock observations . . . 115
8.1.1 Recovering the right configuration . . . 115
8.1.2 Recovering parameter values . . . 117
8.1.3 Discussion on the optimal number of constraints . . . 117
8.2 Comparison with empirical constraints . . . 119
8.2.1 Metallicity . . . 119
8.2.2 Star formation rates . . . 123
8.2.3 Ionization parameter . . . 125

9 Estimating the escape fractions of ionizing photons **130**
9.1 Context . . . 130
9.2 Preliminary analysis . . . 133
9.2.1 Defining the cuts . . . 133
9.2.2 Defining a proxy for the escape fraction . . . 136
9.2.3 Summary and definition of $f_{\rm esc,HII}$ . . . 140
9.2.4 Variation of $f_{\rm esc,HII}$ in the grid . . . 141
9.3 Inferring $f_{\rm esc,HII}$ in the Dwarf Galaxy Survey . . . 143
9.4 Developments . . . 182
9.4.1 $f_{\rm esc}$ in different energy ranges . . . 182
9.4.2 Power-law models . . . 186
9.5 Summary . . . 187

10 Estimating CO-dark H_2 masses **189**
10.1 Context . . . 189
10.1.1 Molecular gas traced by H_2 and CO lines . . . 189
10.1.2 Deriving CO-to-H_2 conversion factors . . . 190
10.1.3 Calibrating $\alpha_{\rm CO}$ . . . 192
10.2 Evidence for a CO-dark gas reservoir in the DGS . . . 193
10.2.1 Reference study . . . 193
10.2.2 Comparison with Madden et al. (2020) . . . 195
10.3 Reproducing CO(1-0) observations with topological models . . . 196
10.3.1 CO measurements and unit conversions . . . 196
10.3.2 Tension between H_2, CO, and [C II] lines . . . 198
10.3.3 Single-sector models . . . 200
10.3.4 Multisector models . . . 202
10.3.5 Broken power-law distribution . . . 205
10.4 Geometries of the models . . . 206
10.4.1 A_V dependencies in single and multisector models . . . 206
10.4.2 PDFs of the power law parameters . . . 208
10.5 Total H_2 mass, CO-dark gas fractions, and $\alpha_{\rm CO}$ estimates . . . 210
10.5.1 Total H_2 mass . . . 210
10.5.2 CO-dark gas fractions . . . 210
10.5.3 $\alpha_{\rm CO}$ conversion factor . . . 212
10.6 Summary . . . 213

V Prospectives and conclusion **217**

11 Prospective and conclusion **218**
11.1 Technical improvements 218
11.1.1 Improving `Cloudy` models 218
11.1.2 Improving `MULTIGRIS` 220
11.2 Other applications of `MULTIGRIS` 222
11.2.1 Application on other samples of galaxies 222
11.2.2 Application to other data 224
11.3 Conclusion 225

VI Annexes **228**

Glossary **232**

Acronyms **234**

Bibliography **250**

List of Figures

1 Aperçu de la stratégie de modélisation 7

1.1 Evolution of the gas surface density and stars for a simulated GMC 17
1.2 Nucleosynthesis origin of the main elements 18
1.3 Main ISM phases 20
1.4 Schematic view of the ISM phases 21
1.5 Thermal processes in 1D photoionization and photodissociation codes 23
1.6 Cooling function of the ISM at solar-metallicity 25
1.7 Density and temperature profiles for two `Cloudy` models 26
1.8 Heating and cooling at solar-metallicity 29
1.9 Heating and cooling at low-metallicity with an X-ray source 30

2.1 Galaxy formation and evolution timeline 32
2.2 Illustration of the Dwarf Galaxy Survey 42
2.3 Images of a few well-known galaxies in the DGS 43

3.1 Open vs. close geometry in `Cloudy` 49
3.2 Pseudo-3D representation 50
3.3 3D model of a bipolar H II region 51
3.4 3D model of a fractal H II region 51
3.5 Schematic of the ISM and representative topology 54
3.6 Topological configurations considered in previous studies 57

4.1 Single cluster, multisector representative topology 59
4.2 Multicluster, multisector representative topology 60

5.1 Structure of the datasets in each HDF5 file associated to a `Cloudy` model 63
5.2 Incident SEDs 66
5.3 Time evolution of the ionizing photon production rate from Xiao et al. (2018) 67
5.4 Abundance profiles 69
5.5 Comparison of various abundance profiles 70
5.6 Z_{dust} and q_{AF} vs. 12+log(O/H) from Galliano et al. (2021) 72
5.7 Density profile from Cormier et al. (2019) 75
5.8 Hα luminosity function from Bradley et al. (2006) 77
5.9 Schematic of the cuts 78

6.1 Illustration of the Bayes theorem 86
6.2 Illustration of the MCMC principle 88

7.1 Workflow within `MULTIGRIS`. 91
7.2 Comparison SMC/NUTS samplers 97
7.3 Illustration of SMC principle 99
7.4 Illustration of the component identification step 100
7.5 Diagnostic plots for primary parameters in a 3-sector model 104
7.6 Violin plots for primary parameters for a 3-sector model 105
7.7 Corner plots for primary parameters in a 3-sector model 105
7.8 Line agreement diagnostic for a 3-sector model 106
7.9 Match plot and correlation diagnostic for a 3-sector model 107
7.10 Posterior PDF 108
7.11 Posterior PDF 109

7.12 Skewed Uncertainty Ellipse 110

8.1 Test on mock observations with varying number of sectors 116
8.2 Benchmark to estimate the number of sectors required in radiation-bounded models . 118
8.3 Evolution of the solution with the number of lines for radiation-bounded models . . . 120
8.4 Evolution of the solution with the number of lines for density-bounded models . . . 121
8.5 Predicted metallicity as compared to metallicity estimate from the R23-P and T_e-direct methods 122
8.6 SFR proxies 123
8.7 SFR UV proxies 124
8.8 $<U>$ vs. $U_{\rm in}$ and $U_{\rm out}$ color coded by L_X/L_* and L_* 126
8.9 $<U>$ vs. $U_{\rm in}$ and $U_{\rm out}$ color coded by density, metallicity, and thickness 127
8.10 Dependencies of $<U>$ with various galactic parameters in the DGS 128

9.1 Cuts vs. A_V 134
9.2 Cuts vs. physical depth 135
9.3 Proxies to estimate $f_{\rm esc}$ 137
9.4 Cuts vs. $f_{\rm esc,HII}$ for models without an X-ray source 141
9.5 Cuts vs. $f_{\rm esc,HII}$ for models with an X-ray source 143
9.6 Cuts vs. $f_{\rm esc}$ by energy range 182
9.7 Cuts vs. Q by energy range 183
9.8 $f_{\rm esc}$by E range 184
9.9 $f_{\rm esc}$(1-∞) and $f_{\rm esc}$(20-∞) distributions in the DGS 185
9.10 $f_{\rm esc,HII}$ and f[C II] vs. metallicity for power-law models 186
9.11 $f_{\rm esc,HII}$ dependencies for power-law models 186

10.1 Schematics of the CO-dark gas reservoir at low- and solar-metallicity 194
10.2 Comparison of predictions for emission lines from single-sector models including H_2 as constraint 198
10.3 Comparison of predictions for single-sector models using either H_2 or CO as constraint 199
10.4 Predicted vs. observed CO, [C II], and [O I]lines 201
10.5 Representation and properties of the 4-sector model for IZw18 202
10.6 $\alpha_{\rm CO}$ estimates for I Zw 18 203
10.7 Best solutions among multisector models for the DGS galaxies. 204
10.8 A_V dependencies for single-sector and multisector models 207
10.9 Posterior PDFs of the power-law parameters 208
10.10 Values of the power law parameters controlling the density and ionization parameters 209
10.11 Values of the broken power law parameters controlling the cut 210
10.12 H_2 masses, CO dark fractions, and $\alpha_{\rm CO}$ estimates 211
10.13 Comparison of the H II, H I, and H_2 masses for different topologies. 213
10.14 Comparison of the H_2 masses associated with C^+, C^0, and CO for different topologies. 214
10.15 Comparison of the CO-dark to CO-bright H_2 mass ratios for different topologies. . . 215

List of Tables

4.1 Example of pre-defined configurations in `MULTIGRIS` 58

5.1 Input parameters of Cloudy models. 65
5.2 Abundances from Cloudy ISM table. 70
5.3 Tabulated $Q(H^0)$ for BPASS SED 76
5.4 Tabulated $Q(H^0)$ for the multicolor blackbodies. 76

10.1 Table of CO measurements in the DGS. 196
10.2 Distribution of parameters adopted in the broken power-law configuration. 205

A1 List of the lines saved in the SFGX grid 231

Abstract

In this thesis, I will present a new modeling framework adapted to study the multiphase interstellar medium (ISM), including and especially within low-metallicity galaxies. In such chemically unevolved environments, important changes of the gas heating and cooling mechanisms are expected, with an impact on the resulting spectral signatures. We intend to use those spectral signatures as probes of the physical conditions in low-metallicity dwarf galaxies, which may resemble the conditions expected in the early universe.

To analyze the complex spectral signatures arising from the interstellar gas heated by stars and X-ray sources, we selected a photoionization and photodissociation code (`Cloudy`). While the geometry of such 1D models is simplistic, they can be improved using topological models which reproduce the emission by combining several models as components drawn from a large pre-computed grid. To explore this N-dimensional parameter space, we exploited the strength of Bayesian techniques and developed a new code, `MULTIGRIS`, which uses suites of emission lines to constrain the posterior probability distribution functions of various parameters.

After the technical description of the code and its inner workings, I will present a first application to a sample of star-forming dwarf galaxies in the local universe, using multisector topologies to account for the contribution of density-bounded H II regions from where ionizing photons might escape. We find that the escape fraction is globally larger at low metallicity, especially for high specific star formation rates, though with a wide scatter which suggests the influence of secondary parameters.

In a second application, we intend to determine the gas mass distribution in the different phases of the ISM, with a specific focus on the molecular gas phase and its ubiquitous "CO-dark" component, unseen with classical tracers like CO. We find that the molecular gas is dominated by this CO-dark component at low metallicity, and discuss the impact of different clouds distributions on the CO-to-H_2 conversion factors we derive.

Tools like `MULTIGRIS`, applied to mock observations or observations of nearby galaxies, may help gain insights into the complex mechanisms at play in the first galaxies that formed within the Reionization epoch. This approach opens promising opportunities for future applications, including for high-redshift galaxies, with the already available and future detections from ALMA and JWST.

Keywords: `interstellar medium; star formation; H II regions; molecular gas; dwarf galaxies; spectroscopy; infrared; Bayesian statistics; numerical methods.`

Dans cette thèse, je présente un nouveau formalisme mathématique adapté à l'étude des différentes phases du milieu interstellaire (MIS), y compris et en particulier dans les galaxies de faible métallicité. Dans de tels environnements peu évolués chimiquement, les mécanismes de chauffage et de refroidissement sont bouleversés, ce qui a un impact marqué sur les signatures spectrales qui en résultent. Nous souhaitons utiliser ces signatures spectrales caractéristiques pour sonder les conditions physiques dans les galaxies naines de faible métallicité. Ces conditions pourraient ressembler celles attendues dans l'univers primordial.

Pour analyser les signatures spectrales complexes associées au chauffage du gaz interstellaire par des étoiles et des sources de rayonnement X, nous avons sélectionné un code de photo-ionisation et photodissociation (`Cloudy`). Bien que la géométrie de ces modèles 1D soit simpliste, ils peuvent être aisément adaptés avec des modèles topologiques qui reproduisent l'émission observée en combinant plusieurs modèles issus d'une grille pré-calculée. Pour explorer cet espace de paramètres à N dimensions, nous avons utilisé des méthodes Bayésiennes et développé un nouveau code, `MULTIGRIS` qui se base sur des suites de raies d'émission pour contraindre les distributions postérieures de probabilité de différents paramètres.

Après la description technique du code et de son fonctionnement, je présenterai une première application à un échantillon de galaxies naines formant des étoiles dans l'univers proche, qui utilise des modèles topologiques multisecteurs pour prendre en compte la contribution de régions H II bornées par la densité, par lesquelles des photons ionisants peuvent s'échapper. Nous trouvons des fractions d'échappement globalement plus élevées à faible métallicité, particulièrement pour les taux spécifiques de formation d'étoiles les plus hauts, avec cependant une large dispersion qui suggère l'influence de paramètres secondaires.

Dans une seconde application, nous souhaitons déterminer la distribution en masse du gaz dans les différentes phases du MIS, en se focalisant particulièrement sur l'étude du gaz moléculaire et de sa composante sombre, invisible avec les traceurs habituels comme CO. Nous trouvons que le gaz moléculaire est dominé par cette composante CO-dark à faible métallicité et discutons l'impact de différentes distributions de nuages sur les facteurs de conversion de CO à H_2 que nous dérivons.

L'application d'outils comme `MULTIGRIS`, à la fois à des galaxies simulées et à des observations de galaxies proches, pourrait permettre d'améliorer notre compréhension des mécanismes en jeu au sein des premières galaxies à s'être créées, durant l'époque de la Réionisation. Cette approche ouvre la voie à de futures applications, y compris pour des galaxies lointaines, pour exploiter les observations actuelles et à venir avec des télescopes tels qu'ALMA ou le JWST.

Mots-clés: milieu interstellaire; formation d'étoiles; régions H II; gaz moléculaire; galaxies naines; spectroscopie; infrarouge; statistiques bayésiennes; méthodes numériques.

Résumé étendu en francais

Introduction

Les spectres observés à l'aide de télescopes au sol (par exemple, ALMA, VLT) et de télescopes spatiaux (par exemple, *Herschel*, *Spitzer*, et JWST) comprennent des signatures uniques du milieu interstellaire (MIS), dans lesquelles est intégré un grand nombre d'informations physiques et chimiques sur les différentes phases du MIS: le gaz ionisé, les régions de photodissociation, et le gaz moléculaire. Des études détaillées de cette structure multiphase sont cruciales pour comprendre les mécanismes physiques qui agissent dans le MIS des galaxies primordiales, pour lesquelles peu d'informations sont généralement accessibles.

Un élément clé de l'étude des environnements peu évolués chimiquement est l'impact du contenu en métaux et en poussières sur les réservoirs de gaz nécessaires à la formation des étoiles. En particulier, nous souhaitons comprendre comment les processus physiques associés à la formation d'étoiles (par exemple, les processus de chauffage et de refroidissement, les rétroactions radiatives et mécaniques, la formation et la fragmentation des nuages) diffèrent pour des métallicités extrêmement faibles (voir Chapitre 1). Nos principales questions scientifiques sont résumées dans le Chapitre 2 : le gaz atomique neutre participe-t-il au processus de formation d'étoiles ? Comment le gaz moléculaire est-il distribué ? Les galaxies naines pauvres en métaux et formant des étoiles sont-elles potentiellement responsables de la réionisation de l'univers et quelles conditions physiques déterminent la fraction d'échappement des photons ionisants ?

Pour apporter des éléments de réponse à ces questions, différents types d'outils peuvent nous permettre de modéliser le MIS (Chapitre 3). Nous avons choisi de développer de nouveaux modèles à plusieurs composantes ou modèles "topologiques" (Chapitre 4) qui peuvent être contraints par les signatures spectrales disponibles pour chaque galaxie. Chaque composante est issue d'une large grille de modèles de photo-ionisation et de photodissociation, adaptée à l'application aux galaxies formant des étoiles, avec une large gamme de métallicités (Chapitre 5). Afin de trouver la meilleure combinaison de modèles pour reproduire un ensemble donné de raies d'émission, nous avons codéveloppé un nouveau code, `MULTIGRIS` (Lebouteiller & Ramambason 2022a, Chapitre 7), qui utilise des méthodes de statistique bayésiennes (Chapitre 6) pour déduire les paramètres physiques à partir des mesures de raies d'émission.

Ce nouveau code a d'abord été testé (Chapitre 8) pour reproduire et étendre les résultats d'études précédentes concernant la galaxie prototypique de faible métallicité I Zw 18 (Lebouteiller & Ramambason 2022b). L'application principale de ma thèse (Ramambason et al. 2022) a ensuite consisté à appliquer notre code à un échantillon de 39 galaxies issues du Dwarf Galaxy Survey (DGS; Madden et al. 2013) pour réaliser une étude détaillée de leur porosité aux photons ionisants à partir de leurs raies d'émission dans l'infrarouge moyen à lointain (Chapitre 9). Un second projet s'est concentré sur l'estimation de la distribution des masses de gaz neutre et moléculaire dans le DGS, y compris la contribution relative de la composante de gaz moléculaire "CO-dark" dont on s'attend à ce qu'elle augmente à faible métallicité (Chapitre 10).

Le nouveau cadre statistique présenté dans cette thèse représente une étape majeure vers une

modélisation plus réaliste du MIS mutliphase des galaxies et ouvre des opportunités prometteuses pour des applications futures, décrites dans le chapitre 11.

Contexte observationnel et grandes questions

Le milieu interstellaire multiphase

Le milieu interstellaire désigne l'ensemble du gaz et de la poussière se trouvant entre les étoiles. Sa dynamique et sa structure sont régies par l'équilibre entre les différents processus énergétiques qui assurent son chauffage et son refroidissement. Ces mécanismes sont particulièrement sensibles au contenu en métaux et en poussière des galaxies, avec des différences substantielles attendues dans les environnements de faible métallicité, en particulier dans l'univers primordial.

Les étoiles se forment dans les régions les plus denses du MIS, appelées nuages moléculaires, ayant des densités de 100 à 100000 cm^{-3}. Ces nuages sont également massifs ($10^4 - 10^7\ M_\odot$) et froids ($10 - 40$ K). Les étoiles s'y forment par effondrement gravitationnel lorsque les mécanismes de pressions thermiques ne supportent plus la gravité propre du nuage. Cette vision simpliste d'un équilibre entre pression thermique et gravité permet de décrire simplement la formation d'étoile à l'aide de grandeurs caractéristiques telles que la masse de Jeans au-dessus de laquelle un nuage s'effondre sous l'effet de sa gravité propre :

$$M_J = \frac{4\pi}{3}\rho\left(\frac{\lambda_J}{2}\right)^3, \tag{1}$$

avec λ_J la longueur caractéristique de Jeans :

$$\lambda_J = \left(\frac{\pi c^2}{G\rho}\right) \sim 2.2\,\mathrm{pc}\left(\frac{c_S}{0.2\,\mathrm{km\,s^{-1}}}\right)\left(\frac{n}{10^2\,\mathrm{cm^{-3}}}\right)^{-1/2}, \tag{2}$$

où c_s est la vitesse de la lumière, G la constante gravitationnelle, ρ la densité volumique du gaz et n la densité du gaz.

En réalité, la formation d'étoiles est aussi régulée par des effets de turbulence et des phénomènes de rétroaction locaux (passages de bras spiraux, explosion de supernovae (SNe), etc). Après une période de ~3 à 20 millions d'années, les étoiles finissent par disperser leurs nuages moléculaires. Cette dispersion peut être accélérée par des éjections protostellaires, des flux de gaz ionisé et des SNe.

Les étoiles sont étroitement liées au milieu dans lequel elles se forment, évoluent et meurent. En particulier, la rétroaction stellaire dépose de l'énergie, de la quantité de mouvement, de la masse et des métaux dans le MIS environnant via les phénomènes de photo-ionisation, les vents stellaires et les SNe. Par conséquent, l'histoire de la formation d'étoiles dans les galaxies joue un rôle majeur dans l'évolution du MIS. En particulier, les vieilles étoiles jouent un rôle important, car elles libèrent de la poussière et des métaux dans le MIS à la fin de leur vie.

À l'équilibre thermodynamique, le MIS irradié par des sources ionisantes se stabilise et développe une structure complexe, composée de couches superposées, "en oignon". Le gaz diffus (c'est-à-dire le

gaz très chaud ionisé, le gaz chaud ionisé, le milieu neutre chaud, et le milieu neutre froid) constitue la majeure partie du volume du MIS, mais ne représente qu'une petite fraction de la masse et de la luminosité.

Au contraire, l'émission observée, par exemple, dans les domaines optique et infrarouge est dominée par des régions relativement plus denses et irradiées, proches des amas stellaires. Nous nous concentrons donc sur la structure du MIS autour des amas stellaires. Ces régions sont, au premier ordre, entièrement responsables de l'émission observée dans les galaxies naines formant des étoiles que nous examinons dans cette thèse. Nous supposons donc que l'émission intégrée provenant d'une galaxie peut être exprimée comme la somme des régions les plus irradiées, près des étoiles.

Ces régions sont structurées comme suit : autour des étoiles, une sphère de gaz ionisé appelé région H II se développe. Les régions H II peuvent être bornées par la radiation, lorsque leur rayon externe est délimité par l'équilibre entre le flux de photons produits par l'amas stellaire et les recombinaisons sur les atomes d'hydrogène, ou bien bornées par la densité (ou de manière équivalente la matière) lorsque leur rayon est fixé par l'absence de gaz en bordure de la région H II. Dans ce dernier cas, des photons ionisants produits par les étoiles peuvent potentiellement s'échapper dans le milieu entre les régions H II.

Ces régions H II sont généralement entourées d'une couche de d'hydrogène neutre fortement irradié (les régions de photodissociation) dont les atomes peuvent se recombiner plus loin des étoiles pour former du gaz moléculaire H_2. Une partie de ce gaz moléculaire réside dans des régions où H_2 demeure sous forme moléculaire grâce à sa capacité d'auto-écrantage alors que d'autres molécules, comme CO, sont photodissociées sous l'effet d'intenses champs de radiation. Ces régions dites "sombres en CO" sont particulièrement difficiles à observer puisque CO est le traceur communément utilisé pour tracer le gaz moléculaire, H_2 n'ayant pas de dipôle permanent.

La structure et les mécanismes de chauffage et de refroidissement au sein de cette structure multicouche autour des étoiles varient fortement dans les environnements pauvres en métaux et soumis à de fortes conditions d'irradiation. En particulier, peu d'informations sont disponibles sur les mécanismes régissant la physique des premières galaxies à s'être créées dans un univers primordial, dénué de métaux et de poussière.

Les galaxies naines de faible métallicité proches et lointaines

Durant les 380 000 premières années après le Big Bang, notre Univers est complètement opaque; même les télescopes les plus sensibles ne peuvent obtenir aucune donnée sur cette période appelée les "âges sombres". Il faut attendre 400 millions d'années après le Big Bang pour assister à "l'Aube cosmique", c'est-à-dire à la formation des premières étoiles et des premiers ensembles d'étoiles : les galaxies. Ces premières galaxies sont constituées de centaines de milliards d'étoiles se formant dans un gaz "primitif", peu évolué chimiquement. Elles contiennent donc très peu d'éléments lourds, peu de poussières stellaires et leurs jeunes étoiles sont à l'origine d'un rayonnement intense qui irradie le gaz environnant.

Parmi ces galaxies primordiales se trouvent les principales suspectes pour comprendre une transition majeure de notre univers, la Réionisation, durant laquelle notre Univers est devenu presque complètement transparent au rayonnement Ultra-Violet (UV) en l'espace de 500 millions d'années.

De nombreuses questions persistent sur la physique en jeu dans ces galaxies primordiales : possèdent-elles une structure particulièrement propice à l'échappement de photons ionisants ? Peuvent-elles, à elles seules, expliquer la Réionisation de l'univers ? Les réservoirs de gaz et mécanismes de formations d'étoiles sont-ils radicalement différents dans le gaz primordial, privé d'éléments lourds ?

Pour répondre à ces questions, nous nous sommes particulièrement intéressés aux galaxies naines formant des étoiles présentes dans l'univers proche. Ces dernières sont caractérisées par :

- Leurs tailles compactes.
- Leurs faibles masses de gaz et d'étoiles.
- Leurs faibles luminosités.

Bien que ces galaxies naines locales ne soient pas des analogues parfaites des galaxies primordiales, elles ont l'avantage d'être facilement observables, avec de nombreuses raies d'émissions détectables par les télescopes actuels, notamment du domaine UV à l'infrarouge (IR).

Pour étudier la physique et la chimie des environnements de faible métallicité, plusieurs cibles sont disponibles dans l'univers très proche. En particulier, nous utiliserons dans cette thèse l'échantillon du Dwarf Galaxy Survey (DGS; Madden et al. 2013) qui rassemble les observations photométriques et spectroscopiques de 50 galaxies proches (0.5 – 191 Mpc) observées dans les domaines de l'infrarouge lointain et submillimétrique avec le télescope spatial *Herschel*. Toutes ces galaxies ont également été observées dans le domaine de l'infrarouge moyen avec le télescope spatial *Spitzer* et des mesures spectroscopiques sont disponibles pour tous les objets sauf 5.

Dans le contexte de ce manuscrit, nous nous sommes concentrés sur l'analyse du DGS pour répondre aux questions spécifiques suivantes, qui sont directement liées au contexte plus large présenté précédemment :

- Comment le contenu en métaux et en poussières affecte-t-il la porosité du MIS aux photons ionisants et quels paramètres contrôlent la valeur des fractions d'échappement des photons ?
- Quelle est la distribution de la masse de gaz dans les galaxies primitives et comment pouvons-nous estimer le réservoir de gaz moléculaire, en tenant compte de la composante "sombre en CO" ?

La richesse des traceurs infrarouges observés dans cet échantillon de galaxies est idéale pour obtenir des contraintes sur le MIS de faible métallicité en comparant l'émission observée à ce qui est prédit par des modèles du MIS de plus en plus complexes.

Modéliser le MIS avec des modèles multicomposantes

En simulant le transfert de radiation dans un gaz chauffé par des sources ionisantes, des modèles comme `Cloudy` (Ferland et al. 2017) produisent des observations synthétiques qui peuvent être directement comparées aux observations. Cette méthode a été utilisée avec succès pour prédire des

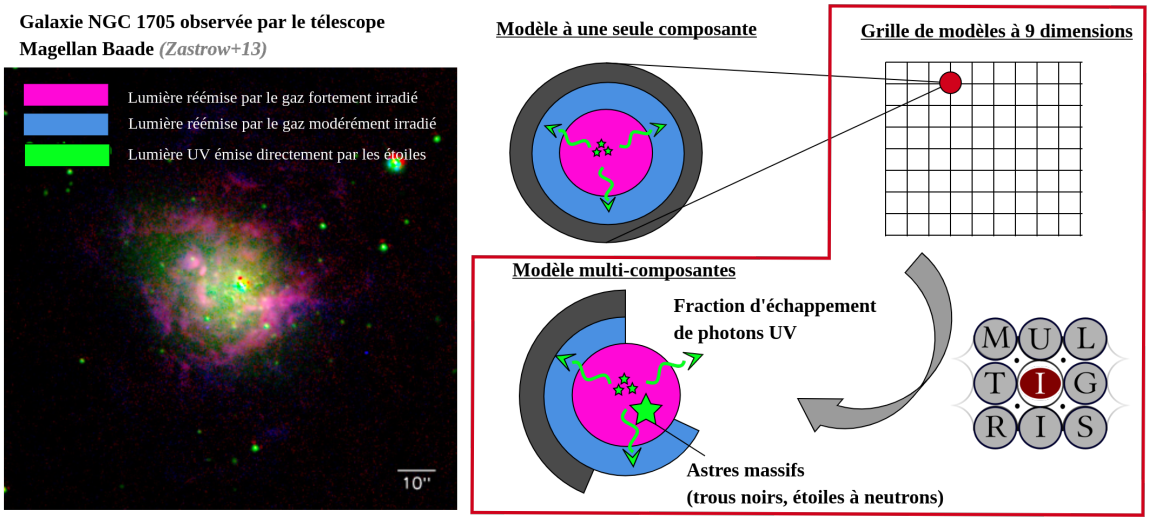

Figure 1: Aperçu de la stratégie de modélisation. **Gauche :** Image optique 3-couleurs de la galaxie NGC 1705 vue par le télescope Magellan Baade. Cette image montre de la lumière émise par les étoiles (en vert) et le gaz ionisé chauffé intensément, proches des étoiles (en rose) et modérément, un peu plus loin (en bleu). **Droite :** Illustration des différentes étapes de ce travail de thèse pour passer de modèles simples, à une composante, à des modèles multicomposantes en utilisant `MULTIGRIS`.

propriétés “simples” de galaxies (taux de formation d’étoile, âge des étoiles, composition chimique). Dans le cadre de ma thèse, pour la première fois, nous tentons de prédire des quantités “complexes” qui dépendent de multiples paramètres de nos modèles et sont mal reproduites par des modèles à une seule composante.

La stratégie adoptée pour combiner ces modèles à la géométrie simpliste (sphérique, à une seule composante) est illustrée dans la Figure 1. Pour cela, nous avons utilisé une représentation "topologique" introduite dans des études précédentes (Péquignot 2008; Cormier et al. 2012; Lebouteiller et al. 2017; Cormier et al. 2019; Lambert-Huyghe et al. 2022) qui reproduit l’émission de galaxies non résolues spatialement en combinant linéairement les prédictions associées à différentes composantes ou secteurs. L’automatisation de cette combinaison repose sur le développement d’un nouveau code que j’ai codéveloppé, `MULTIGRIS` (Lebouteiller & Ramambason 2022a) permettant la combinaison de modèles ainsi que la prédiction de densités de probabilité pour de nombreux paramètres de ces modèles combinés.

Ce code repose sur l’utilisation de nouveaux outils de statistiques bayésiennes, encore relativement peu utilisés pour l’étude du milieu interstellaire, bien qu’ils résolvent de nombreux problèmes liés à la comparaison modèles/observations. Ce code permet de combiner de manière dynamique différents modèles issus d’une large grille de 1 632 000 modèles que j’ai construite au cours de la première moitié de ma thèse. Cette grille fait varier neuf paramètres physiques associés à différentes propriétés physiques et chimiques du gaz et des étoiles.

En particulier, cette grille explore pour la première fois les effets liés à la présence d’étoiles binaires et d’astres massifs comme des trous noirs ou étoiles à neutrons qui irradient fortement le gaz et affectent sa structure. Enfin, cette grille varie l’épaisseur de la couche de gaz qui entoure les étoiles, un paramètre crucial pour étudier les fractions d’échappement de photons UV puisque les

modèles “épais” absorbent totalement les radiations UV. Cette large grille de modèles, ainsi que le code d’analyse statistique, ont été rendus public pour faciliter leur utilisation par la communauté astrophysique.

Résultats principaux

Étude de la fraction d’échappement des photons ionisants

Une première application de `MULTIGRIS` s’est focalisée sur l’étude de la fraction d’échappement de photons ionisants dans l’échantillon de galaxies naines de faible métallicité du DGS. Contrairement aux prédictions obtenues pour des quantités historiquement bien contraintes (par exemple, la métallicité, ou le taux de formation d’étoiles), nous avons constaté que la fraction d’échappement des photons ionisants est sensible à la fois aux hypothèses de modélisation faites pour prédire cette fraction dans notre grille de modèle et à la topologie choisie par `MULTIGRIS` (par exemple, le nombre et la distribution des secteurs autour d’une source ponctuelle).

Nos résultats principaux sont les suivants :

- Nos résultats confirment la porosité aux photons UV des régions H II à faible métallicité trouvées dans l’étude précédente de Cormier et al. (2019), mais en se basant cette fois sur une méthode statistique plus robuste.
- Les fractions de photons ionisants s’échappant des régions H II ($f_{\mathrm{esc,HII}}$) estimées dans le DGS sont élevées, atteignant des valeurs jusqu’à ~60% pour certains objets. Nos résultats démontrent la possibilité d’identifier les régions H II associées à des fuites de photons ionisants, en se basant uniquement sur leurs signatures spectrales en émission.
- Nous trouvons que $f_{\mathrm{esc,HII}}$ est corrélée avec le taux spécifique de formation d’étoiles et anti-corrélée avec la métallicité. Cela renforce l’image d’une de fuite de photons UV régulé par rétroaction stellaire, et fortement dépendante des propriétés galactiques globales.
- Nous avons identifié plusieurs rapports de raies IR et optiques qui pourraient permettre d’identifier les larges fractions d’échappement dans des échantillons de galaxies.

Alors qu’un tel échappement de photons UV est observée dans les études résolues des régions H II, elle n’est cependant pas observée à une échelle galactique dans l’univers proche. Cette absence de détection de fraction d’échappement pourrait être principalement liée aux limitations instrumentales qui empêchent la détection directe des photons UV dans les galaxies naines locales. Cependant, plusieurs effets, qui ne sont pas pris en compte dans nos modèles, pourraient également affecter les fractions d’échappement globales dans le DGS. En particulier, le rôle potentiel d’une composante diffuse de gaz pourrait réduire l’échappement de photons UV ionisants à grande échelle.

Étude du gaz moléculaire sombre en CO

Dans un second temps, nous avons réalisé une autre application de MULTIGRIS portant cette fois sur l'étude des réservoirs de gaz moléculaire dans les galaxies du DGS. Contrairement à l'étude précédente sur les fractions d'échappement, nous avons constaté que les topologies combinant un nombre discret de composantes ne sont pas assez complexes pour reproduire les signatures issues de la phase moléculaire.

Nous avons donc examiné une autre topologie, basée sur l'utilisation de distributions de type loi de puissance et loi de puissance brisée (comportant deux indices et un point de pivot), qui généralisent la combinaison discrète de composantes. Ces modèles nous ont permis de reproduire avec succès la raie d'émission CO(1-0) et d'estimer les masses de gaz moléculaire dans différentes phases : le gaz sombre en CO, associé au carbone atomique ionisé; $M(H_2)_{C^+}$, ou neutre; $M(H_2)_{C^0}$), et le gaz moléculaire associé à l'émission de CO; $M(H_2)_{CO}$.

Nos résultats principaux sont les suivants :

- Tous nos modèles prédisent que le gaz moléculaire dans le DGS est complètement dominé par $M(H_2)_{C^+}$, qui est en moyenne 100 fois plus importante que $M(H_2)_{C^0}$ et 1000-10000 fois plus importante que $M(H_2)_{CO}$. Nous trouvons que les galaxies DGS devraient se situer sur la relation de Schmidt-Kennicutt, qui relie le taux de formation d'étoiles à la masse de gaz moléculaire disponible, comme prédit précédemment par Madden et al. (2020).

- Nos résultats indiquent que [C II] 158μm est un bon traceur du gaz moléculaire total. En particulier, la robustesse de la relation $M(H_2)$-[C II] aux changements de topologies et la faible dispersion de nos prédictions suggèrent que [C II] est un très bon proxy pour prédire H_2 dans les environnements de faible métallicité. À l'inverse, l'intérêt de CO comme traceur du gaz moléculaire disparaît à faible métallicité (par exemple, Bolatto et al. 2013), puisque son émission n'est pas associée à une masse de gaz moléculaire importante.

- La raie CO(1-0) ne peut pas être facilement reproduite dans notre cadre statistique qui vise à reproduire simultanément toutes les raies provenant des différentes phases du MIS. Les configurations à un seul secteur et multisecteurs sont incapables de reproduire toutes les détections de CO dans notre échantillon. Alors que la représentation topologique multisecteurs peut décrire avec succès les phases ionisées et neutres, nous constatons qu'elle ne permet pas de reproduire l'émission provenant du gaz plus dense qui émet CO.

- L'utilisation d'une distribution en loi de puissance brisée pour contrôler la profondeur des nuages de gaz nous permet de reproduire l'émission de CO observée. L'émission est donc produite par un nombre relativement faible de régions denses qui contribuent peu à la masse moléculaire totale. Il en résulte de grandes fractions de gaz moléculaire sombres en CO, associées à un rapport $M(H_2)_{tot}/M(H_2)_{CO}$ plus élevé que ce qui était prédit par les études précédentes (Madden et al. 2020).

- Les prédictions de $M(H_2)_{CO}$ varient de manière significative en fonction des différentes hypothèses de nos modèles. Cela rend les prédictions du rapport $M(H_2)_{total}/M(H_2)_{CO}$ et du facteur de conversion α_{CO} très sensibles aux hypothèses de modélisation.

Calculer un facteur de conversion α_{CO} à l'échelle galactique est une tâche complexe. Nous trouvons qu'une dispersion significative est présente dans nos prédictions. Nous avons identifié deux cas

extrêmes correspondant à des distributions différentes du gaz d'où provient CO : (1) le gaz émettant CO est distribué de manière homogène et peut être modélisé par un seul secteur émettant CO, associé à des masses H_2 relativement importantes. C'est le cas décrit par les modèles à une ou plusieurs secteurs, conduisant à des valeurs de α_{CO} proches de la relation de Madden et al. (2020). (2) L'émission de CO est dominée par des petits nuages denses avec de grands A_V, à l'extrémité supérieure de la distribution en loi de puissance. Dans ce cas, les valeurs de α_{CO} sont proches des valeurs galactiques. Nos prédictions se situent entre la relation Madden et al. (2020) et la valeur Galactique, avec une dispersion relativement importante.

L'emplacement exact entre ces deux relations est probablement déterminé par nos hypothèses de modélisation concernant la distribution des nuages émettant CO. Des contraintes supplémentaires sont nécessaires sur la distribution de CO si nous voulons calculer avec précision la valeur de α_{CO}.

Perspectives et conclusion

Le travail présenté dans cette thèse constitue une première étape vers une modélisation plus réaliste du milieu interstellaire multiphase. Le nouveau formalisme développé au sein du code `MULTIGRIS` ouvre de nombreuses possibilités pour des futures applications et permet d'envisager des améliorations.

En particulier, les modèles topologiques présentés dans cette thèse peuvent être améliorés. Cela peut notamment passer par l'inclusion de données issues du continu stellaire et de la poussière qui permettraient de s'affranchir de certaines hypothèses que nous avons dû fixer pour générer les modèles `Cloudy`. Plus précisément, l'utilisation du continu stellaire permettrait de contraindre l'histoire de la formation d'étoiles dans les galaxies observées, en considérant des évolutions plus complexes que les simples flambées d'étoiles, à un âge donné, que nous avons considérées (par exemple, des épisodes de formation d'étoiles par flambées successives ou des formations d'étoiles continues).

Le continu de poussière, quant à lui, permettrait de contraindre des propriétés comme la distribution en masse, en taille, ou en température des grains de poussière interstellaire, ainsi que leur nature. Ces données sont cruciales pour comprendre les mécanismes de chauffage du MIS qui impactent la formation d'étoiles. Un autre axe d'amélioration des modèles serait une meilleure prise en compte du gaz diffus neutre et ionisé qui n'a pas été modélisé dans cette thèse.

Par ailleurs, des améliorations de nos outils numériques sont aussi possibles. Par exemple, le code présenté dans cette thèse pourrait être adapté afin d'adopter une structure hiérarchique permettant de considérer simultanément des échantillons de galaxies en inférant leurs propriétés physiques par le biais d'hyperparamètres qui prennent en compte les tendances globales, à l'échelle d'échantillons de galaxies (alors que le code actuel les traite indépendamment). Une autre voie prometteuse est l'utilisation du machine learning afin d'accélérer certaines étapes coûteuses en temps, comme l'interpolation en N dimensions sur nos grilles de modèles.

Enfin, la version actuelle du code peut déjà être utilisée pour un grand nombre d'objets. L'étude de l'échantillon du DGS peut être étendue à l'analyse d'autres paramètres que les fractions d'échappement et les masses de gaz, comme par exemple, les processus de chauffage dominants à faible métallicité (Lebouteiller et al. in prep). Parmi les études à venir, l'application à d'autres galaxies que le DGS permettrait de traiter des problématiques spécifiques. Par exemple, un projet en cours s'intéresse à

l'application de MULTIGRIS à la galaxie NC 253 (Beck et al., in prep), qui possède un trou noir actif en son centre, avec de possibles signatures spectrales associées.

Un autre projet consiste à appliquer le code à l'échantillon du Low-z Lyman Continuum Survey dans lequel 89 galaxies ont été rassemblées avec des mesures directes de leurs fractions d'échappement en UV (Marques-Chaves et al. in prep). Ces galaxies sont idéales pour étendre l'étude la fraction d'échappement dans le DGS et permettre de mieux comprendre si les larges $f_{\mathrm{esc,HII}}$ prédites à faible métallicité correspondent effectivement à des fractions d'échappement importantes aux échelles galactiques. Enfin, l'application de MULTIGRIS à un échantillon plus large de galaxies permettrait d'analyser statistiquement les tendances globales.

Nous comptons également appliquer le code à l'échantillon limité en volume RESOLVE observé avec le Sloan Digital Sky Survey dans le domaine optique. Cet échantillon est particulièrement intéressant, car il présente une bimodalité de distribution en métallicité et certaines galaxies possèdent des fractions d'AGN non nulles. Il est bien adapté pour y étudier les signatures spécifiques à la présence de sources X, mais également pour calibrer l'utilisation de distributions de probabilité pour décrire certains paramètres physiques.

Un outil tel que MULTIGRIS permet de préparer l'étude de galaxies à grand décalage vers le rouge qui sont observées proches ou directement dans la période de Réionisation avec des télescopes tels que le JWST et ALMA. L'étude des grands décalages vers le rouge reste extrêmement complexe en raison du faible nombre de détections (quelques raies seulement) qui doivent être complémentées par des priors sur, par exemple, la masse ou la métallicité. Enfin, les processus dominant la physique de l'univers primordial restent fondamentalement inconnus et les extrapolations à partir d'observations dans l'univers proche peuvent ne pas être valides. Pour pallier cette limite, l'utilisation de simulations permettrait d'explorer ces paramètres physiques non contraints et de calibrer de futurs outils d'analyse.

Enfin, l'application de MULTIGRIS à des galaxies simulées peut permettre d'explorer les effets de résolutions et de mieux comprendre les interactions entre différentes échelles spatiales, depuis les régions H II jusqu'aux échelles galactiques. En parallèle, de nouveaux échantillons d'observations multilongueur d'onde obtenues à une résolution spatiale sans précédent, tel que PHANGS, peuvent permettre d'éclairer notre compréhension globale des phénomènes de formation d'étoiles, et leur interaction avec le milieu interstellaire. Cet axe d'étude permet de créer des ponts entre les petites échelles spatiales, à laquelle se forment les étoiles, et les propriétés globales des galaxies.

INTRODUCTION

Spectra observed using ground-based telescopes (e.g., ALMA, VLT) and space telescopes (e.g., *Herschel*, *Spitzer*, and JWST) provide unique "fingerprints" of the interstellar medium (ISM), in which a large set of physical and chemical information is embedded about the different phases of the ISM: ionized gas, photodissociation region, and molecular gas. This multiphase structure in which different regions with different physical conditions co-exist results from complex dynamical processes (e.g., turbulence, self-gravity, thermal, and mechanical feedback), which control the evolution of the ISM and the way it condenses to eventually form stars.

The newly-formed stars interact with their surroundings as they deposit energy, momentum, and release chemical elements and dust at the end of their life in the ISM. The complex interplay between the ISM and the ionizing sources that formed out of it regulates the evolution of galaxies. Detailed studies of this multiphase structure are crucial to understand the physical mechanisms at play in the ISM of different environments, including in primordial galaxies, for which little information is available. While direct studies of the primordial universe have been limited by observational barriers (e.g., detection threshold, poor spatial resolution, extinction of the emission by the intergalactic medium), the study of nearby galaxies can help to improve our understanding of the physical mechanisms at low metallicity. Local star-forming dwarf galaxies are ideal laboratories to probe the conditions of star formation within the dust- and metal-poor ISM.

One key element in the study of chemically primitive environments is the impact of the metals and dust content on the gas reservoirs (e.g., neutral atomic and molecular), which fuel the star-formation. In particular, we wish to understand how the physical processes associated with star formation (e.g., heating and cooling processes, radiative and mechanical feedback, cloud formation and fragmentation) are modified at extremely low metallicity (see Chapter 1). Our main scientific questions are summarized in Chapter 2: does the neutral atomic gas participate in the star-formation process? how is the molecular gas distributed? Are metal-poor star-forming dwarf galaxies potentially responsible for the Reionization of the universe, and what physical conditions drive the escape fraction of ionizing photons?

To address these questions, we used multicomponent models (or "topological" models, see Chapter 4) such as the ones introduced by Péquignot (2008) and refined iteratively in several studies in our group (e.g., Cormier et al. 2012; Lebouteiller et al. 2017; Polles et al. 2019; Cormier et al. 2019; Lambert-Huyghe et al. 2022). Such models can be constrained based on the observed spectral signatures available for each galaxy. They rely on the linear combination of several components that mimic the contribution of different regions, having different physical and chemical conditions, to the total integrated emission observed in unresolved galaxies. While topological models may not correspond to the actual spatial distribution of the gas in the ISM, which is usually unknown and generally too complex, they can, however, provide a luminosity-weighted representation of the components that contribute the most to the emission of a given spectrum.

Each component of the topological models is drawn from a large grid of photoionization and photodissociation models, tailored for the application to starbursting galaxies, with a wide range of metallicities (see Chapter 5). In particular, we explored the low-metallicity regime by including metallicity bins down to $1/100\,Z_\odot$ and a possible contribution of an X-ray source, since the latter are

expected to be relatively more numerous and brighter at low metallicity (e.g., Ponnada et al. 2020; Saxena et al. 2020; Basu-Zych et al. 2016). We also include in our grid the contribution of density-bounded regions, stopped before they reach their Strömgren radius, from where ionizing photons can escape. This is motivated by a previous work (Ramambason et al. 2020) showing that the inclusion of such regions is necessary to reconcile the observed emission lines with the measured escape fractions of Lyman Continuum (LyC) photons, which are detected in a growing number of nearby galaxies (see, e.g., the recent Low-z Lyman Continuum Survey; Flury et al. 2022a,b).

In order to find the best combination of models to reproduce a given set of emission lines, we co-develop a new code, `MULTIGRIS` (Lebouteiller & Ramambason 2022a, Chapter 7), using Bayesian statistics (see Chapter 6) to infer physical parameters from emission line measurements. This new code allows us to overcome some previous limitations of classically used frequentist methods, like the χ^2 (see discussion in, e.g., Andrae et al. 2010; Lebouteiller & Ramambason 2022b). It enables a proper treatment of the upper limits and systematic uncertainties and allows the inclusion of informative priors. Most importantly, the outputs of the code are posterior probability distribution functions (PDF), which can be used to derive statistical trends and examine possible correlations or degeneracies between parameters.

This new code was first benchmarked (Chapter 8) and applied to reproduce and extend the results of previous studies (Péquignot 2008; Lebouteiller et al. 2017) regarding the prototypical low-metallicity galaxy I Zw 18 (Lebouteiller & Ramambason 2022b). This detailed study has allowed us to reproduce automatically the previous results obtained for this object and to examine the PDFs of various parameters, including its gas masses in different phases and its escape fractions of photons, within different energy ranges.

The first application in my thesis (Ramambason et al. 2022) has consisted in applying our code to a sample of 39 galaxies drawn from the Dwarf Galaxy Survey (DGS; Madden et al. 2013) to carry out a detailed study of their porosity to ionizing photons based on their mid- to far-infrared emission lines (see Chapter 9). This project builds on previous work from Cormier et al. (2019) which derived non-unity PDR covering factors in the DGS, hinting at an increased porosity of the ISM at low-metallicity. We revisit those results by converting the ill-defined PDR covering factors into estimates of the escape fraction of ionizing photons from H II regions. Our results confirm that substantial leakage (up to ~60%) of ionizing photons from density-bounded H II regions are expected at low metallicity. We further examined the dependencies of our estimates of escape fractions with various galactic parameters and spectral tracers.

A second project has focused on estimating the mass distribution of neutral and molecular gases in the DGS, including the relative contribution from the "CO-dark" molecular gas component. This CO-dark molecular gas consists of gas reservoirs in which H_2 remains self-shielded while CO is photodissociated by strong radiation fields (e.g., Wolfire et al. 2010). This component has been detected both in our Galaxy (e.g., Grenier et al. 2005; Ackermann et al. 2012) and in extragalactic environments, including in low-metallicity dwarf galaxies (e.g., Poglitsch et al. 1995; Madden et al. 1997), and is expected to increase at low metallicity (see Chapter 10). Focusing on the DGS, we found in Madden et al. (2020) that this CO-dark component dominates the molecular gas reservoirs, and translates into large CO-to-H_2 conversion factors at low metallicity. Using our new topological framework, we examine the robustness of those results to changes in the modeling assumptions, and specifically on the distribution of clouds. Our results confirm the large CO-dark molecular gas fractions in the

DGS and the interest of using [C II]158μm as a tracer of the total molecular gas. We find, however, that our CO-to-H_2 conversion factors are highly sensitive to the modeling assumptions and discuss their interest in low metallicity environments where CO barely traces any molecular gas mass.

The new statistical framework developed with `MULTIGRIS` represents a major step towards a more realistic modeling of the complex mutliphase ISM of galaxies and opens promising opportunities for future applications, described in Chapter 11. The code is available for public use and provided with extensive precomputed grids of Cloudy models, although it can also be used with other grids. This new tool is well suited to tackle questions regarding the physics at play in the multiphase ISM in nearby unresolved objects with many spectral signatures. It may be complemented by physically motivated priors, based for example on simulations, in order to make the most of the few detections already available for high-redshift galaxies. In particular, this tool is relevant in the context of the future spectroscopic survey of galaxies in the early universe, which will be detected with telescopes such as ALMA and the JWST.

Part I

Observational context & big questions

Chapter 1

The multiphase ISM

In this chapter, we aim to briefly describe the physics and chemistry that control the energetic balance in the ISM. In Section 1.1, we describe the dynamical processes that regulate the interactions between stars and gas. Specifically, we present the physical processes that allow a galaxy to channel its diffuse H I gas reservoir into dense, star-forming molecular clouds (MCs) and how the ISM gets enriched in heavy elements and dust. These physical mechanisms set the conditions of the **thermal equilibrium**, which will be assumed in our all models throughout the thesis. At thermal equilibrium, the ISM stabilizes in a multiphase structure (Section 1.2), with various heating and cooling mechanisms at play in each phase. Among other parameters, we focus on how the metallicity and the hardness of ionizing radiation affect the energetic balance in the multiphase ISM, with potential implications for the primordial universe (see Chapter 2).

1.1 Dynamical picture

The dynamical view presented in this section is mostly drawn from studies of the Milky-Way and nearby galaxies. We hence focus on describing how stars form in dense and cold molecular clouds, heat their surrounding medium, and release their metals and dust content in "classical" physical and chemical conditions (i.e., at solar-metallicity). To which extent this scenario remains valid in the primordial universe will be discussed in Chapter 2.

1.1.1 From diffuse gas to star formation

MCs form out of diffuse, warm atomic flows which are responsible for local increase of the gas density (e.g., Tacconi et al. 2020). Many processes can lead to the formation of such instabilities (e.g., passage of a spiral arm, disk instabilities, supernovae (SNe), stellar winds or turbulence). Most of the volume in galaxies is filled with hot and low-density gas, which makes up for a low fraction of the mass (e.g., Girichidis et al. 2020). Because the cooling times are long compared to the dynamical time in the hot phase (e.g., McKee & Ostriker 1977; Gnat & Ferland 2012), turbulent compression also plays a major role in the cooling processes. The effect of turbulence is two-fold: on the one hand, it assists cooling by creating over-densities, on the other hand, it can support the gas through effective

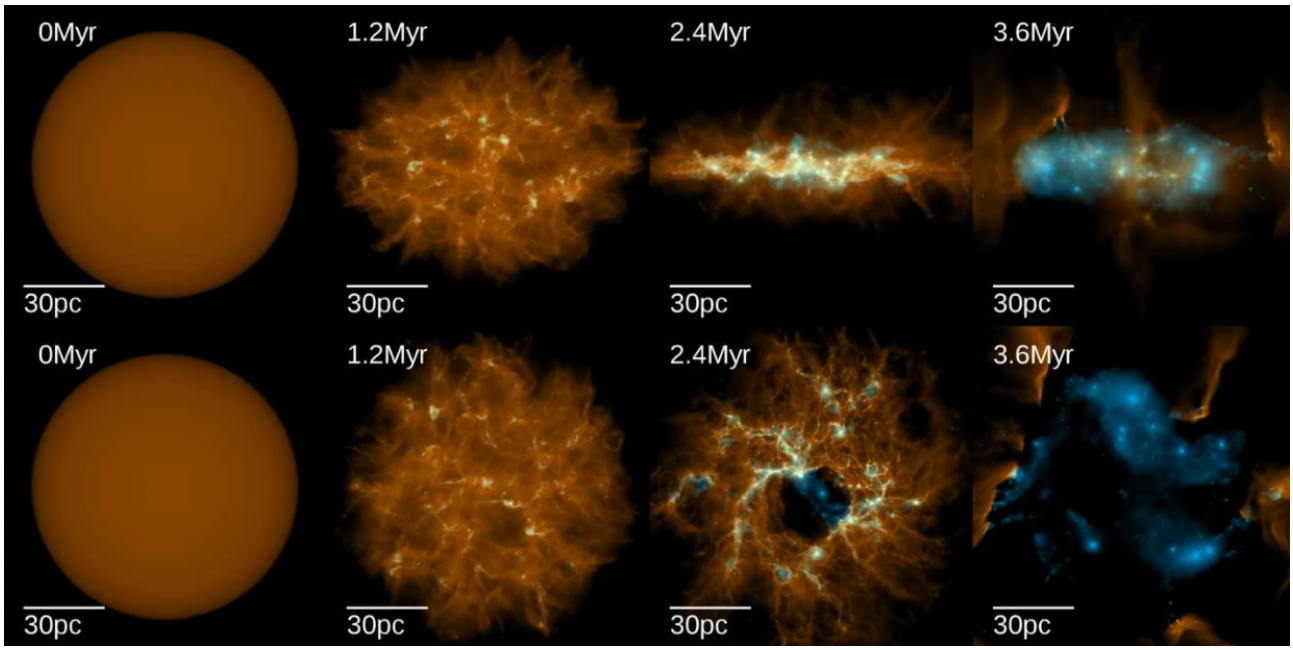

Figure 1.1: Evolution of the gas surface density (orange) and young stars (blue) for a simulated GMC of $3 \times 10^7\,M_\odot$ and an initial radius of 50 pc, from left (initial conditions) to right (after 3.6 Myr). The top row is showing the edge-on view of the disc and the bottom row, the face-on view. Star formation starts within a free-fall time (here 1.2 Myr) and stops after a few Myr, when the gas has been completely blown out by feedback. Figure taken from Grudić et al. (2018).

turbulent pressure (e.g., Mac Low & Klessen 2004). Since the warm H I gas is thermally unstable, the increase of density leads to a catastrophic cooling and a rapid contraction of the gas, reaching higher and higher densities, and forming molecules.

Stars form in the densest regions of massive (10^4 – $10^7 M_\odot$), dense ($n(H_2) \sim 10^2$ – 10^5 cm^{-3}), and cold (10 – 40 K) giant molecular clouds (GMCs) (see, e.g., reviews from Tacconi et al. 2020; Chevance et al. 2020a, 2022). The conditions of collapse in the cold gas depend on the balance between self-gravity and thermal pressure, which can be expressed as a characteristic length; the Jeans length:

$$\lambda_J = \left(\frac{\pi c^2}{G\rho}\right) \sim 2.2\,\mathrm{pc}\left(\frac{c_S}{0.2\,\mathrm{km\,s^{-1}}}\right)\left(\frac{n}{10^2\,\mathrm{cm^{-3}}}\right)^{-1/2}, \tag{1.1}$$

where c_s is the sound speed, G the gravitational constant, ρ the density of the gas, and n the gas number density. The associated Jeans mass, above which an isothermal parcel of fluid collapses due to its self-gravity, is given by:

$$M_J = \frac{4\pi}{3}\rho\left(\frac{\lambda_J}{2}\right)^3 \sim 34\,M_\odot\left(\frac{c_s}{0.2\,\mathrm{km\,s^{-1}}}\right)^3\left(\frac{n}{10^2\,\mathrm{cm^{-3}}}\right)^{-1/2}, \tag{1.2}$$

Once the GMC starts collapsing, the duration of collapse essentially depends on the density of the gas and is of the order of a few free-fall times:

$$t_{ff} = \sqrt{\frac{3\pi}{32G\rho}} \sim 3\mathrm{Myr}\left(\frac{n}{10^2\mathrm{cm}^{-3}}\right)^{-1/2}. \tag{1.3}$$

We note, however, that there is no clear consensus regarding the relative roles played by self-gravity versus turbulence in the cloud formation processes. Because the warm and cold ISM phase coexist in pressure equilibrium (Field & Saslaw 1965; McKee & Ostriker 1977), both turbulent motions and gravity may play an important role in shaping molecular clouds. Although MCs may have formed due to gravitational attraction in the galaxy, the resulting clouds are turbulent structures which are globally not bound by self-gravity (Heyer et al. 2009; Heyer & Dame 2015).

The global properties of the ISM set the properties of GMCs that condense out of it. Correlations are observed, for example, between the galactic gas pressures and the ones of GMCs, the surface densities, and the volume densities (e.g., Chevance et al. 2020a, 2022). In particular, the resulting star formation rates are strongly correlated with the total molecular mass, at galactic scale. A typical evolution of a star-forming MC is presented in Figure 1.1. After a period of 3 – 20 Myrs, the stars eventually disperse their molecular birth clouds. This disruption can be caused by protostellar outflows, ionized gas flows, and SNe.

Stars are tightly connected to the medium in which they form, evolve, and die. In particular, the stellar feedback deposits energy, momentum, mass, and metals in the surrounding ISM through photoionization, stellar winds and SNe. Therefore, the star formation history of galaxies plays a major role in evolution of the ISM (see, e.g., Chevance et al. 2020a). In particular, old stars contribute to the enrichment of the ISM in the most abundant elements as they release dust and metals at the end of their life.

1.1.2 Metal and dust enrichment

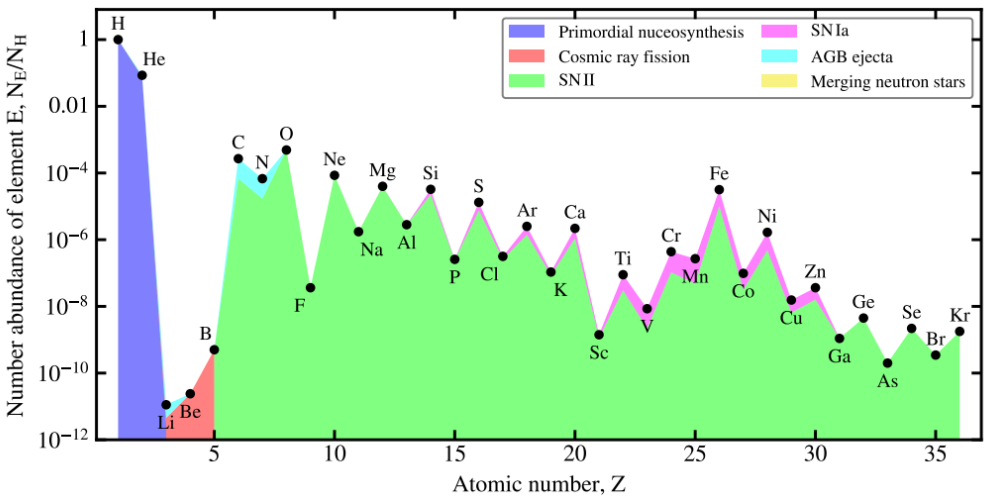

Figure 1.2: Nucleosynthesis origin of the main elements for the Solar abundances from Asplund et al. (2009). Taken from Galliano (2022).

Stars in their late stages return a fraction of the heavy elements or "metals" (i.e., elements heavier

than helium) that they have synthesized to the ISM. The injection of metals can occur through various processes, which are summarized in Figure 1.2:

- **Primordial nucleosynthesis** is responsible for the production of most of the light elements (e.g., H, He).
- **Type II SNe**, which result from the collapse of massive stars, account for the production of most metals.
- **AGB stars ejecta** are responsible for the production of a significant fraction of carbon and nitrogen.
- **Type Ia SNe**, which occur in binary systems in which one of the stars is a white dwarf, are responsible for the production of metals around iron.

Other processes such as cosmic ray fission and merging neutron stars may contribute to release the heaviest metals into the ISM, although they are not the dominant mechanisms. Star formation is also responsible for ISM enrichment of dust by AGB stars ejecta and type II SNe. Hence, both the metallicity and the dust content of galaxies are tightly linked to their star formation history.

While several studies have focused on understanding the chemical evolution, with the specific aim to constrain the global galactic metallicity (abundance of all elements heavier than helium), the individual elemental abundances (see, e.g., review from Maiolino & Mannucci 2019) and the properties or distribution of dust grains (see, e.g., Galliano 2022) in extra-Galactic environments, this is not the aim of this thesis. Nevertheless, the abundance patterns which we use in the models (see Section 5.2.2) are directly set by the formation mechanisms described in this section. Physically motivated variations as a function of metallicity for the metals and dust abundances profiles in the gas are needed to allow meaningful comparisons between models and observations and to investigate the low-metallicity and low-dust regimes expected in the primordial galaxies (see Section 2.1).

1.2 ISM at thermal equilibrium

In the presence of heating sources, the ISM energetic balance can be described in terms of heating and cooling volumetric rates. The thermal equilibrium is obtained when heating and cooling are balanced:

$$R_\Gamma = R_\Lambda, \tag{1.4}$$

where:

$$\begin{cases} \mathrm{R}_\Gamma \equiv \text{volumetric radiative heating rate } [\mathrm{erg\,cm^{-3}\,s^{-1}}] \\ \mathrm{R}_\Lambda \equiv \text{volumetric radiative cooling rate } [\mathrm{erg\,cm^{-3}\,s^{-1}}] \end{cases}$$

This is usually written in terms of cooling and heating functions with:

$$\begin{cases} R_\Gamma = n_{\mathrm{gas}}\Gamma; \\ R_\Lambda = n_{\mathrm{gas}}^2(T_{\mathrm{gas}}) \end{cases} \tag{1.5}$$

The global heating (Γ) and cooling (Λ) functions are hence defined locally and depend on the local physical conditions of the gas. At thermal equilibrium, the gas reservoirs stabilize at specific densities and temperature which we will refer to as "phases" (see Section 1.2.1), each dominated by specific heating and cooling processes (see Section 1.2.2). Additionally, the heating and cooling processes we describe depend on galactic parameters and, in particular, on the metallicity (see Section 1.2.3).

1.2.1 ISM phases

	Density	Temperature	Volume filling factor
HIM	$\simeq 3 \times 10^{-3}$ cm^{-3}	$\simeq 10^6$ K	$\simeq 50\,\%$
H II regions	$\simeq 1 - 10^5$ cm^{-3}	$\simeq 10^4$ K	$\lesssim 1\,\%$
WIM	$\simeq 0.1$ cm^{-3}	$\simeq 10^4$ K	$\simeq 25\,\%$
WNM	$\simeq 0.3$ cm^{-3}	$\simeq 10^4$ K	$\simeq 30\,\%$
CNM	$\simeq 30$ cm^{-3}	$\simeq 100$ K	$\simeq 1\,\%$
Diffuse H_2	$\simeq 100$ cm^{-3}	$\simeq 50$ K	$\simeq 0.1\,\%$
Dense H_2	$\simeq 10^3 - 10^6$ cm^{-3}	$\simeq 10$ K	$\simeq 0.01\,\%$

Figure 1.3: Characteristic of the main ISM phases (in terms of volume), taken from Galliano (2022).

Figure 1.3 shows the typical properties of the main phases that dominate the ISM in volume. Specifically, this table gathers the typical densities and temperatures for each phase (although local variations of those properties exist within each phase). The diffuse phase (i.e., HIM: Hot Ionized Medium, WIM: Warm Ionized Medium, WNM: Warm Neutral Medium, and CNM: Cold Neutral Medium) makes up for most of the volume of the ISM but represents only a small fraction of the mass and luminosity. On the contrary, the emission seen, for example, in the optical and infrared domains is dominated by relatively denser and irradiated regions, close to stellar clusters. In the context of this thesis, we will mostly use as constraints IR emission lines in star-forming galaxies. We hence specifically focus on the ISM structure around stellar clusters.

While the dense, multilayer, irradiated gas around star-forming regions only represents a small volume of galaxies, it dominates the observed continuum and line emission in starbursting galaxies. Thus, we assume that the integrated emission arising from a galaxy can be expressed as the sum of the most irradiated regions, close to the stars. The latter regions are identified by the green boxes on the ISM schematic presented in Figure 1.4.

H II regions

Around stellar clusters, the radiation produced by young O/B stars creates bubbles of ionized gas; the H II regions. In a homogeneous medium, each H II region fills a theoretical sphere called the Strömgren sphere (Strömgren 1939), whose boundaries are defined by the equilibrium between ionization of hydrogen and recombination of electrons by hydrogen atoms. Hence, for a cluster with an H-ionizing photon rate $Q(\mathrm{H}^0)$, the equilibrium can be written:

$$Q(\mathrm{H}^0) = \frac{4}{3}\pi R_\mathrm{S}^3 n_\mathrm{p} \times n_\mathrm{e} \alpha_\mathrm{B}, \tag{1.6}$$

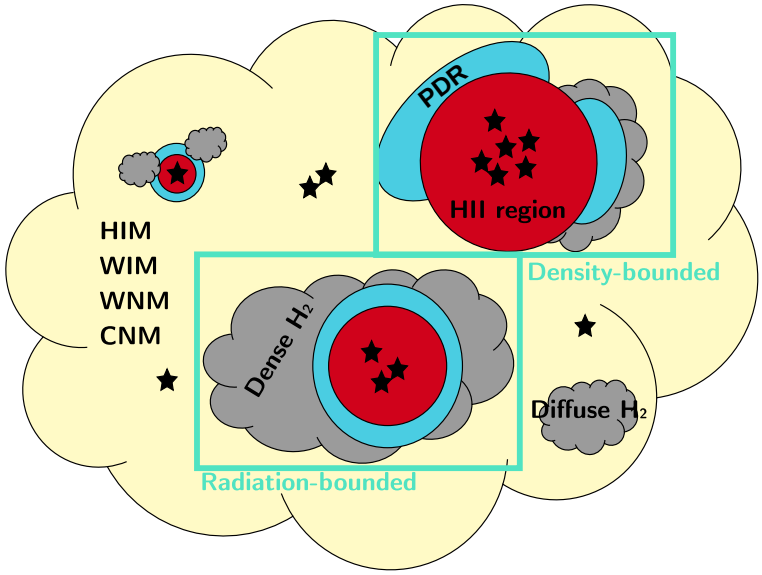

Figure 1.4: Schematic view of the ISM. The two green boxes identify two regions around stellar clusters: one radiation-bounded H II region surrounded by a complete PDR and a complete molecular H_2 cloud and one density-bounded H II regions, which have broken out of its host MCs and is only partially surrounded by the PDR and H_2 layers.

where R_S is the Strömgren radius, n_p the density of protons, n_e the density of electrons and α_B the case B recombination rate. In a sphere dominated by hydrogen ($n_p \approx n_e \approx n_H$), the Strömgren radius can be expressed as:

$$R_S = \left(\frac{3Q(H^0)}{4\pi n_H^2 \alpha_B} \right)^{1/3}. \tag{1.7}$$

When the ionizing front (IF) of the H II region coincides with the Strömgren radius, the H II region is said to be *radiation-bounded*. All the UV photons produced by the stellar cluster are consumed within the sphere to ionize the hydrogen. Nevertheless, in some cases the H II region can be smaller than its Strömgren sphere. This can happen, for example, when H II regions are bounded by the lack of matter or due to dynamical effects (e.g., blister H II regions). In this case, the region is said to be *density-bounded* (or matter-bounded) and ionizing photons may escape (see Figure 1.4). Such escape of ionizing photons from H II regions may have important consequences in the early universe, which will be discussed in Section 2.1.3.

Photodissociation regions

Connected to the radiation-bounded part of H II regions, photodissociation regions (PDRs) consist of thin layers of atomic gas which are strongly irradiated by massive stars. Further away from the stars, the hydrogen recombines to form H_2 is dense MCs. The PDRs are hence the interfaces between ionized regions and dense MCs. Although the volume filling factor of such regions is negligible compared to the diffuse neutral phase, those interfaces absorb most of the far-UV (FUV; 6 – 13.6 eV) radiation and re-emit it in the IR domain.

The first historical models of PDR and molecular gas relied on a “clumpiness” hypothesis and

assumed a structure in which high density clumps were embedded in a lower density interclump medium (e.g., Burton et al. 1990; Parmar et al. 1991). This modeling approach has successfully been applied to interpret observed cooling lines of atomic carbon and CO in resolved observations in local GMCs (Kramer et al. 2008).

Nevertheless, later studies have demonstrated that observations of PDR could also naturally be explained by the presence of a smooth density gradient without the need to invoke small-scale clumpiness. This density gradient can, for example, be produced by a constant pressure profile. Such models have been found to better reproduce PDR observations than constant density models (Bron et al. 2018; Joblin et al. 2018). Dynamical effects due to photoevaporation (the edge of a dense molecular cloud being exposed to stellar UV photons and can freely evaporate into a low pressure cavity) can also have a direct effect on the H/H_2 transition and impact the width of the PDR layer. In some extreme cases, with a sufficiently high velocity of the ionization front, the ionization front and dissociation front can be superimposed, leading to a direct transition from the ionized phase to the molecular phase (Maillard et al. 2021). Finally, we note that when a bright X-ray source is present, an additional X-ray dominated phase can co-exist with the PDR, in which the heating mechanisms are dominated by X-ray photons rather than FUV photons (XDR, Wolfire et al. 2022).

Dense H_2 clouds

At deep visual extinction ($A_V \sim 1$), corresponding to large physical depths, the radiation emitted by the incident sources is extincted by dust within the cloud and hydrogen atoms may form H_2 molecules. Due to the presence of H_2 absorption bands (e.g., Lyman and Werner) which absorb part of the ionizing flux, H_2 molecules remain partly self-shielded from dissociative radiations. Other molecules, such as CO may, however, be photodissociated in regions where H_2 exists and create the so-called "CO-dark" H_2 gas reservoirs. This latter phase is especially important in low-metallicity environments, and will be further discussed in Section 2.2.4.

At large A_V ($A_V \geq 1$), only energetic photons such as hard X-rays and γ-rays can propagate and heat the gas. Because of the latter heating source, the electronic temperature within MCs does not drop below a few tens of kelvin (see Figure 1.3). We note that the physical conditions in MCs are also impacted by dynamical mechanisms such as shocks, turbulence, or gravitational effects, which control the density distribution in MCs. For example, a smoothly increasing density profile is expected in turbulent-dominated clouds (Wolfire et al. 2010).

1.2.2 Thermal processes

Continuous radiative transfer in a multiphase ISM

Considering the density discontinuities which may arise in the presence of density-bounded H II region (see Section 1.2.1), shock fronts, or photo-evaporated PDRs (see Section 1.2.1), modeling the gas around stellar clusters as a single system with continuously connected phases is an obvious simplification. Nevertheless, this simplified "shell-like" structure composed of the three main phases around stellar clusters (H II region + PDR + dense H_2) allows a continuous treatment of the radiative transfers

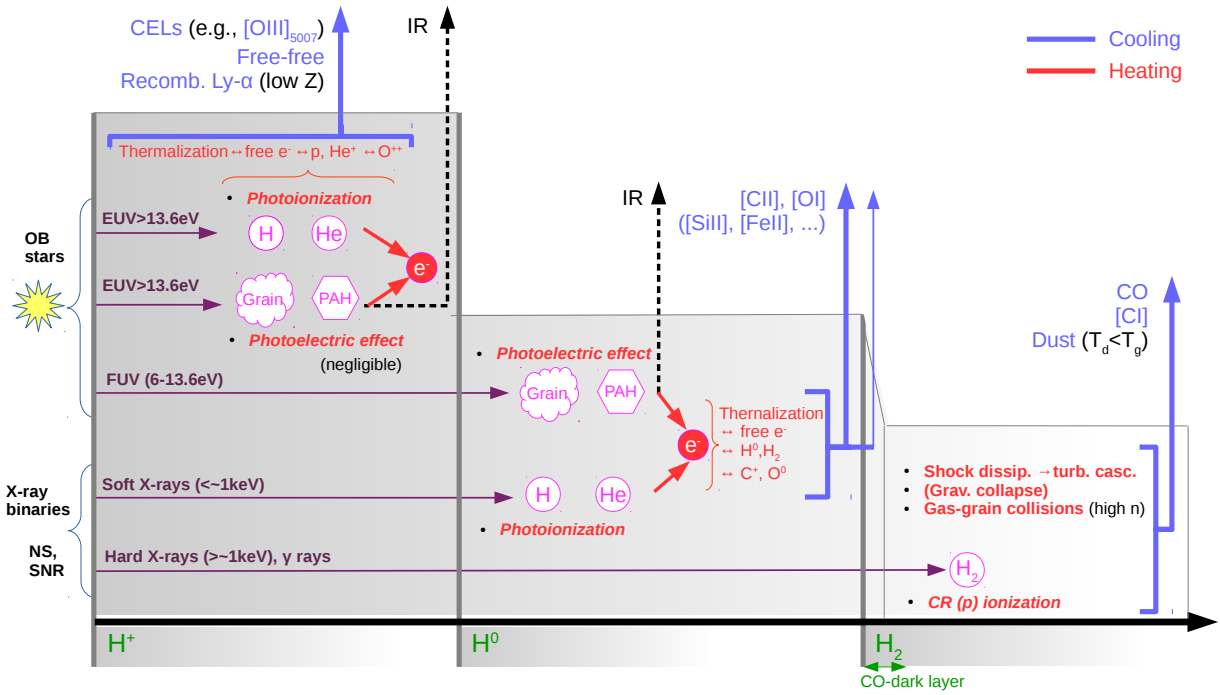

Figure 1.5: Schematics of the thermal processes in 1D photoionization and photodissociation codes (Lebouteiller; private communication). Cooling processes are in blue, while heating processes are in red.

(RT) in a multiphase medium, which is much needed to understand the heating and cooling processes responsible for emission lines. In a more realistic picture, additional phases could be described to account for other sources of ionization (e.g., XDR, shock-dominated regions) and out-of-equilibrium effects. We further discuss the interest of other modeling tools in Chapter 3.

The multiphase picture presented in the previous section is represented in Figure 1.5, along with the main heating and cooling mechanisms in each phase. This is, for example, the representative view adopted by 1D RT codes such as Cloudy (Ferland et al. 2017) which continuously computes the RTs throughout a multiphase medium. As shown in Figure 1.5, various ionizing sources irradiate the ISM with energetic photons that propagate throughout the gas. This radiative feedback leads to complex chains of chemical reactions, with different processes that contribute to either heating or cooling the gas.

We further describe the main heating and cooling mechanisms in the following. While our description is not exhaustive, it is designed to provide the key elements needed to explain the thermal equilibrium of the ISM around stellar clusters. For a more detail view, we refer to Lequeux (2005) and Draine (2011) on which this section is based.

Main heating mechanisms

Most heating mechanisms in the ionized gas involve inelastic Coulomb interactions between electrons, which thermalize in the gas. In a partially ionized medium, elastic collisions between electrons and neutral particles can also transfer energy to neutral particles throughout the medium, although

more slowly. The main heating mechanisms are the following:

- **Photoionization by stellar radiation**: The ionization of atoms and molecules by stellar radiation produces free electrons whose energy is the difference between the photon energy and the ionization energy E=hν. The average kinetic energy of the photoelectrons depends on the photoionization cross-section and the energy density of incident radiation. Photoionization of hydrogen and helium usually dominates the heating budget in H II regions.
- **Photoelectric effect from grains**: UV radiations from hot stars remove electrons from interstellar dust grains. The electrons that are released through this process usually have an energy which is too small to ionize atoms, hence, all the energy is used for heating the gas. The efficiency of the photoelectric effect depends on the grain properties and especially the integrated surface of dust grains. The smallest grains, such as Polycyclic Aromatic Hydrocarbons (PAHs) which dominate the grain surfaces in the ISM are responsible for most of the photoelectric heating (see Galliano 2022). This effect is particularly important in PDRs in which most PAHs reside. It is negligible in H II regions where PAHs are destroyed by UV radiations (Relaño et al. 2018; Chastenet et al. 2019) but can play a role in molecular gas in which small fractions of PAHs may still exist (e.g., Chastenet et al. 2019).
- **Photoionization by X-rays**: X-rays remove photoelectrons from atoms and ions, producing energetic primary photoelectrons. The latter are responsible for cascades that produce secondary ionization and produce other free electrons. Soft X-ray ($\lesssim$ 1 keV) heating is only important up to a certain depth in atomic gas, after which the cloud becomes opaque to their radiation. The hardest X-ray photons ($\gtrsim$ 1 keV) may traverse even the neutral atomic medium without being absorbed.
- **Cosmic ray ionization**: Cosmic rays (CR; high-speed particle being either an atomic nucleus or an electron) transfer energy to the bound electrons of atoms and molecules as well as to the free electrons in the ISM. The heating rate of CRs is poorly constrained in extra-Galactic environments. It is, with hard X-rays, the main heating mechanism in the interior of MCs.
- Other heating mechanisms may contribute such as: chemical heating (in dense gas), grain-gas thermal exchanges (in dense gas), hydrodynamic, and magnetohydrodynamic heating (mostly in the WNM and CNM), or interstellar shocks (in the WNM, the CNM, and the dense H_2 gas).

Main cooling mechanisms

As shown in Figure 1.6, the main cooling mechanisms are functions of the gas temperature. They are also proportional to the electron density times the density of the relevant specie. The dominant cooling mechanisms are also sensitive to metallicity effects, which will be investigated in Section 1.2.3.

The main cooling processes are:

- **Collisional excitation**: a free electron knocks a bound electron to an excited state; it decays, emitting a photon. This mechanism in especially important in H II regions and PDRs. Indeed, the spectra of H II regions are very rich in emission lines, mostly due to the presence

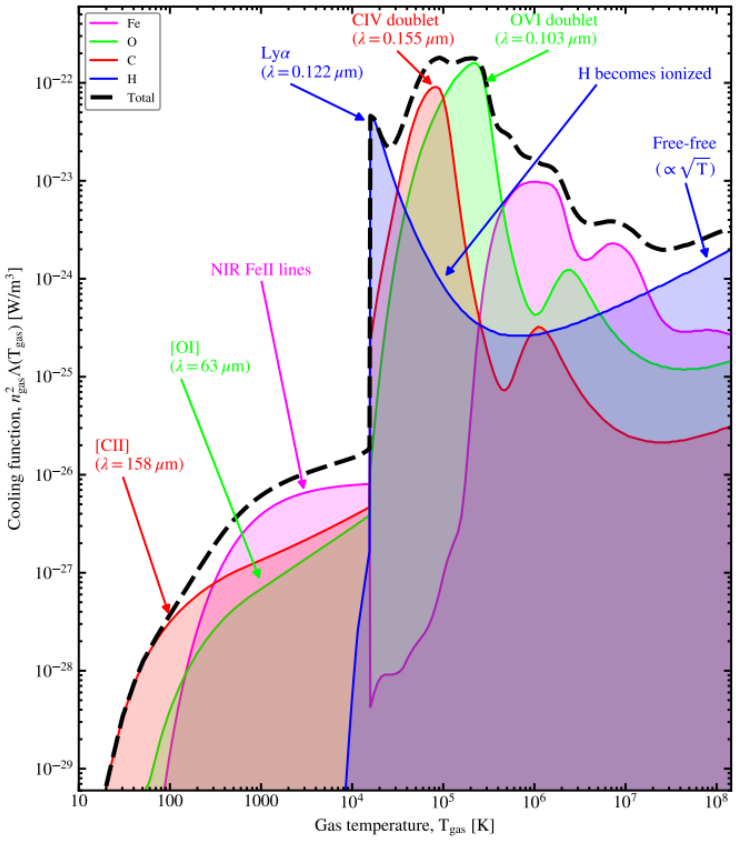

Figure 1.6: Cooling function (as defined in Equation 1.5) of the ISM at solar-metallicity for the main cooling pathways, taken from Galliano (2022).

of collisionally excited lines of metallic ions which emit from the IR to UV domains (e.g., [O III] $\lambda\lambda$4959,5007Å emission lines). In PDRs, the collisional excitation of neutral species (e.g., O^0) or of ions with low ionization potential (e.g., C^+, Fe^+, or Si^+) produces bright emission lines which significantly contribute to the cooling of the ISM (e.g., [C II] [O I], [Si II], [Fe II]emission lines in the IR domain). In MCs, collisional excitation of neutral species and molecules are the dominant cooling mechanism (e.g., [C I], CO emission lines), with dust.

- **Electron-ion recombination**: a free electron recombines with an ion; the binding energy and the free electron's kinetic energy are radiated away. Recombination of helium and hydrogen atoms efficiently cool H II regions through the emission of recombination lines (e.g., Lyα, Hα Hβ). Recombination processes are particularly important in metal-poor environments.

- **Free-free emission**: a free electron is accelerated by an ion, emitting a photon. This process is important in H II regions in which free-free emission from thermalized electron is often detected in radio.

- **Dust**: dust is only able to cool the gas if collisions between dust grains and hydrogen molecules occur often enough to keep them thermally well-coupled. Hence, this cooling mechanism is only important for high gas densities (above $\sim 10^5 - 10^6$ cm^{-3}.

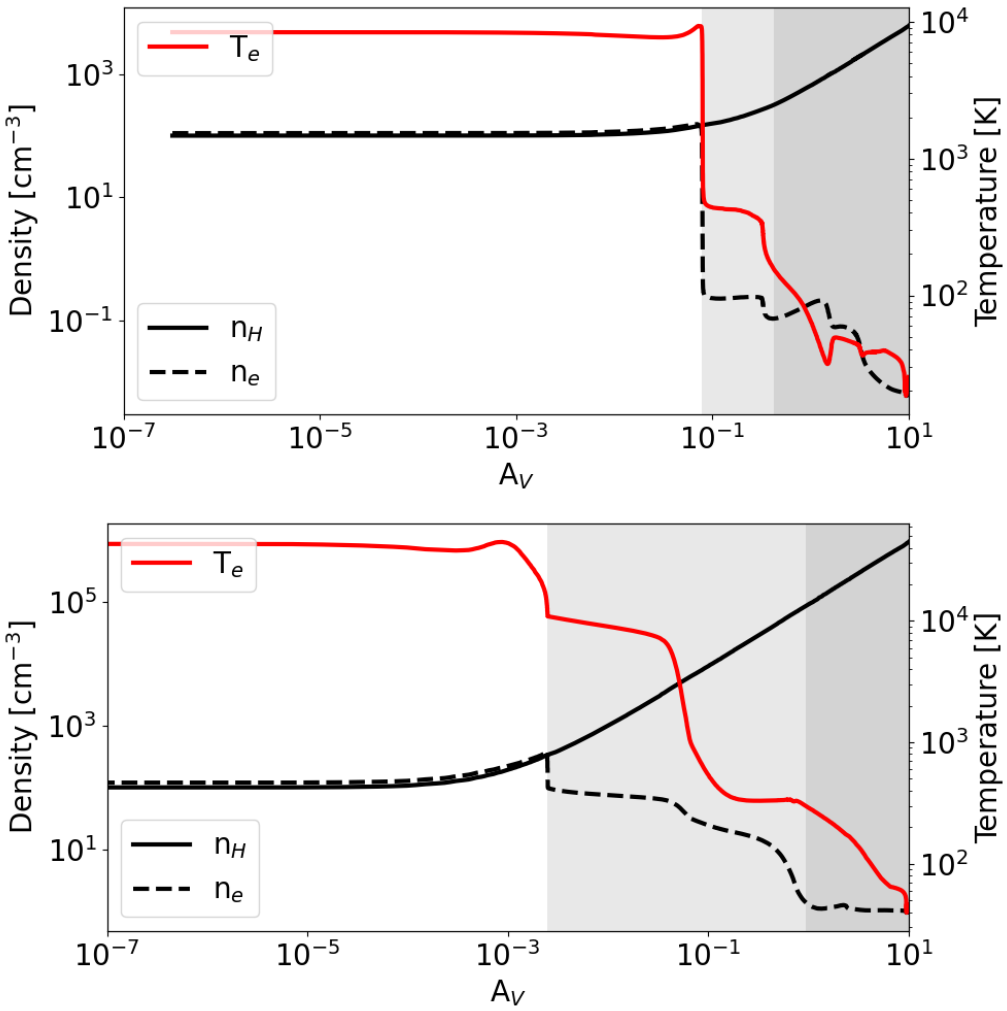

Figure 1.7: Density and temperature profiles at equilibrium for two typical `Cloudy` models. **Top**: Solar-metallicity case with no X-ray source. **Bottom**: Low-metallicity case (1/10 $Z_\odot$) with an X-ray source (L_X/L_*=10%).

1.2.3 Impact of metallicity on thermal equilibrium

As described in Section 1.2.2, the heating and cooling mechanisms that dominate within each phase of the ISM drastically vary with physical and chemical conditions, resulting in specific ISM signatures. Any attempt to disentangle the emission of ISM phases presented in Section 1.2.1 ultimately depends on our ability to understand the physical mechanisms at play in each phase, and in turn identify the specific cooling signatures of each region.

These thermal mechanisms are relatively well-known at solar metallicity. However, they are expected to vary significantly in more extreme environments (e.g., low-metallicity, presence of X-ray source). In this section, we briefly discuss the dominant heating and cooling terms in each phase. To illustrate this, we examine the differences in two arbitrary models which were chosen to represent two significantly different cases (see Figure 1.7): a "typical" model with a solar-metallicity model and no X-ray source and an "extreme" model, with a low-metallicity (1/10 $Z_\odot$) and a significant (L_X=10%L_*) X-ray contribution. Although the latter case may not correspond to actual conditions that could be found in galaxies, it is representative of the most extreme cases that can be found in the sample of

low-metallicity dwarf galaxies that will be studied in this thesis. Moreover, we wish to highlight the differences that may arise in drastically different environmental conditions. Both models are computed with `Cloudy` (Ferland et al. 2017) assuming a radial density profile which will be further described in Section 5.2.3, and drawn from the model grid that will be described in Chapter 5.

H II regions

As described in Section 1.2.2, the dominant heating mechanism in H II regions is the photoionization of hydrogen and helium atoms. Which of hydrogen or helium ionization dominates the heating essentially depends on the hardness of the ionizing spectrum.

As shown in Figure 1.8, for a typical stellar spectrum without any X-ray contribution, the main heating mechanism is the photoionization of hydrogen. At solar-metallicity, there is also a contribution of photoelectric heating on dust grain, although it is not dominant in the H II region. We note that helium ionization only represents a marginal contribution ($< 1\%$) of the heating. The picture is quite different at low metallicity and when an X-ray source is present, with hard photons capable of ionizing He^{2+} (see Figure 1.9). In that case, the main ionizing term is the secondary photoionization of helium, which dominates over the hydrogen ionization in the H II region.

In terms of cooling at solar-metallicity, the dominant mechanisms are the metal emission lines, the free-free emission, and the recombination lines (see Section 1.2.2). Specifically, in the solar-metallicity case (see Figure 1.8), the cooling is dominated by the deexcitation of O^{2+} (e.g., [O III]$\lambda\lambda$4959, 5007Å emission in the optical or [O III] λ88μm in the IR) and S^{2+} (with, e.g., [S III] $\lambda\lambda$9068,9532Å lines in the optical or [S III] $\lambda\lambda$18,33μm in the IR). Near the ionization front, other species with lower ionization potentials may also contribute significantly to cooling (e.g., O^+, N^+). Another important coolant is the free-bound net cooling rate of hydrogen. Finally, cooling by dust and free-free emission may also contribute in smaller proportion. In the low-metallicity case with an X-ray source (see Figure 1.9), the same mechanisms are responsible for cooling gas, but the hydrogen recombination largely dominates over other processes with bright Lyman α emission.

PDRs

The depth and size of the PDR is strongly dependent on the physical and chemical conditions. As shown in Figure 1.7, models with lower metallicity and hard ionizing sources will produce much thicker PDRs than in the solar-metallicity case. Additionally, the structure and thermal processes in PDRs also vary depending on the density profile which is adopted. This choice will be further discussed in Section 5.2.3. At solar metallicity, the main heating mechanism in PDRs is the photoelectric effect on dust grains (see Section 1.2.2). Because a small fraction of the PDR gas remains partially-ionized, the photoionization of hydrogen and helium atoms can also contribute, although the ionization fractions in PDRs are low, especially in solar metallicity environments. As shown in Figure 1.7, the electronic density drops abruptly at the IF down to a few tenths of cm^{-3}.

This is not the case, however, in the low-metallicity model with an X-ray source, in which the electronic density remains high ($\sim 100\,cm^{-3}$) and follows a much shallower drop throughout the PDR. Hence, while the heating in the PDR is largely dominated by photoelectric effect heating (see

Figure 1.8) in the solar-metallicity case, the situation is much more complex in the low-metallicity, X-ray irradiated case. In the latter case (see Figure 1.9), the main heating sources are photoionization of helium and hydrogen followed by the ionization of metals such as O^+. Deeper in the cloud ($A_V \sim 0.1$), the thermal heating due to collisions between H_2 molecules becomes predominant.

In terms of cooling, the main cooling pathway is the emission of collisionally excited metallic lines (see Section 1.2.2). Among the main coolants at solar metallicity, we find the [C II]158μm line in the IR, the [O I] lines (e.g., [O I] λ6300Å in the optical, [O I]63,145μm in the IR), and a marginal contribution from silicon lines (e.g., [Si II]), iron lines (e.g, [Fe II]) and dust. At low-metallicity, cooling through the [C II] emission is reduced, and the main coolants are hydrogen recombination, [Fe II], [Si II], and [O I] lines. The fact that [Fe II] and [Si II] become relatively more important at low metallicity can be explained by the lower dust abundances, which result in less of Fe and Si atoms being locked onto dust grains. Hence, they are more abundant in the gas phase and in turn contribute more to the cooling.

Dense H_2 clouds

In the molecular gas, the heating is mostly dominated by cosmic ray heating, regardless of the metallicity or the presence of an X-ray source. We note that the predictions derived for the molecular phase strongly depend on the assumed CR rate, which is most of the time not well-known. Additionally, in the low-metallicity case, we note that the thermal heating due to collisions between H_2 molecules becomes predominant. This is due to the fact that this model reaches high gas densities ($\geq 10^5 cm^{-3}$) at which collisions between molecules indeed become important.

In the solar-metallicity case with no X-ray source, the cooling in the molecular phase remains dominated by ionic species, namely: [C II], [C I], and [O I] emission lines. At large A_V ($A_V \geq 1$), cooling through CO emission lines also contributes significantly. Other molecules may contribute to the cooling in the dense gas, such as ^{13}CO, C_2H, CS, or HNC. In the low-metallicity environments with hard radiation fields and dense MCs (see Figure 1.9), molecules such as OH and H_2O may also contribute to the cooling.

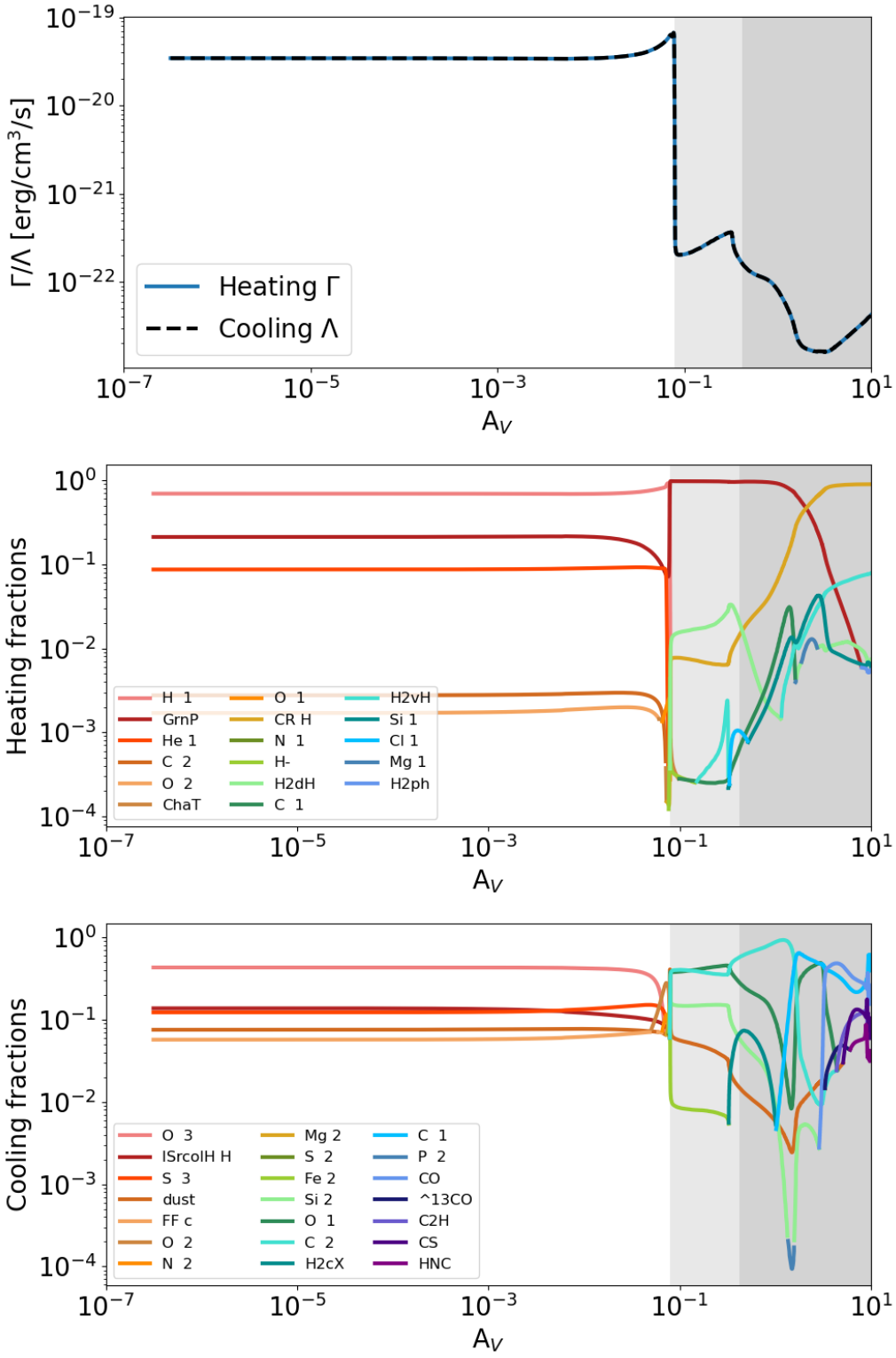

Figure 1.8: Heating and cooling at solar-metallicity, without an X-ray source. **Top**: Volumetric heating and cooling rate at thermal equilibrium. **Middle**: Heating fractions in the H II region (A_V layer in white), PDR (light gray), and molecular zone (dark gray). **Bottom**: Cooling fractions in the H II region (white), PDR (light gray), and molecular zone (dark gray). **Main acronyms**: GrnP = photoelectric heating, ChaT = charge transfer heating, CR H = cosmic rays heating, H2dH = heating due to photodissociation, H2vH = heating due to collisions, H2ph = H_2 photoionization heating, ISrcolH= net free-bound cooling minus free-free heating, FF c = free-free cooling, H2cX = H_2 cooling.

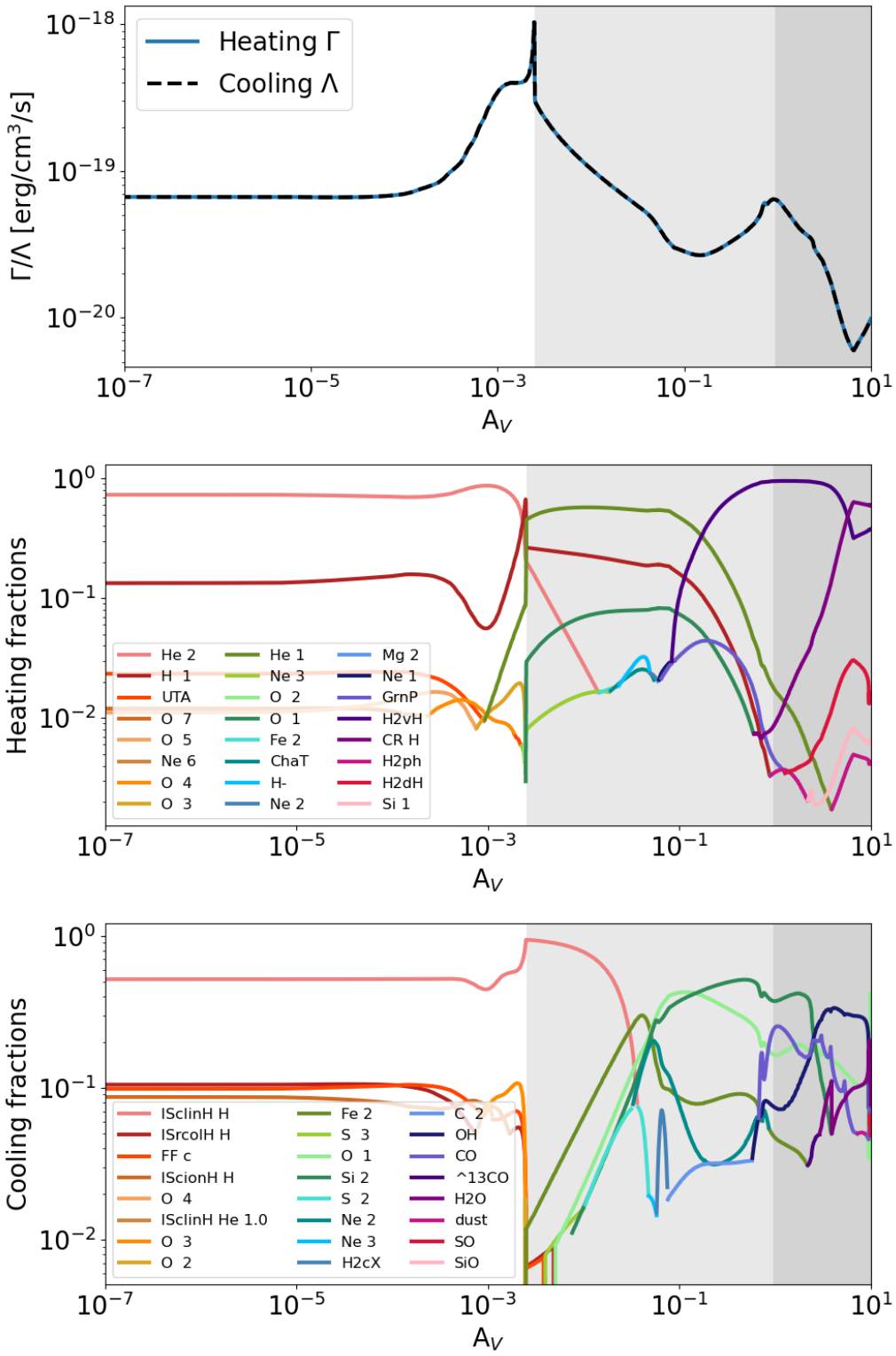

Figure 1.9: Heating and cooling at low-metallicity (Z=1/10 $Z_\odot$, with an X-ray source with L_X/L_*=10%). See Figure 1.8 for the plot description. **Main acronyms**: UTA = unresolved transition array, ChaT = charge transfer heating, GrnP = photoelectric heating, H2vH = heating due to collisions, CR H = cosmic ray heating, H2ph = H_2 photoionization heating, H2dH = heating due to photodissociation, ISclinH = total cooling by hydrogen recombination, ISrcolH = net free-bound cooling minus free-free heating, FF c = free-free cooling, IScionH = net collisional ionization cooling, H2cX = H_2 cooling.

Chapter 2

Low-metallicity dwarf galaxies near and far

In the previous chapter, we described the current picture of the multiphase ISM. In Section 1.1, we described the dynamical processes that shape the ISM and control its main properties, including its metal and dust content. In Section 1.2, we further described the structure of the ISM at thermal equilibrium, the main heating and cooling mechanisms, and how those mechanisms may drastically change in low-metallicity, highly irradiated environments. This motivates the study of extreme environments in which such physical conditions could be found.

In the current chapter, we provide a general overview of the current knowledge of the ISM physics in low-metallicity dwarf galaxies, both in the local and high-redshift universe. In Section 2.1, we describe the main properties of the ISM of primordial dwarf galaxies and their potential role in the cosmological context. In Section 2.2, we show the interest of targeting low-metallicity dwarf galaxies at low-redshift and discuss their use as analogs to high-redshift galaxies. In Section 2.3, we present two relevant categories of dwarf galaxies that will be referred to throughout the thesis. The plan of the practical work following introductory chapters 1 and 2 is described in Section 2.4.

2.1 Primordial dwarf galaxies

2.1.1 Galaxy evolution

The formation of the first stars and galaxies, around 200 – 400 billions years after the Big Bang, marks the beginning of the Epoch of Reionization (EoR; $z \approx 6 - 15$; see Figure 2.1). This epoch corresponds to the transition from a completely neutral and opaque universe to a warm ($\approx 10^4$K) to a completely ionized medium in which photons can propagate (see, e.g., Gnedin & Madau 2022). The first galaxies assembled from a pristine gas, not yet enriched in metals and dust, which condensed into dark-matter (DM) halos. Hence, primordial galaxies most likely exhibited very different properties from the modern galaxies that formed at later ages. Following the hierarchical structure formation scenario, dwarf galaxies are the first to form in low-mass DM halos (Press & Schechter 1974; Dekel & Silk 1986). Hence, the primordial dwarf galaxies are the first building blocks which may then create larger galaxies and clusters of galaxies by undergoing merger events.

This picture has been completed by the emergence of alternative scenarios of formation for mas-

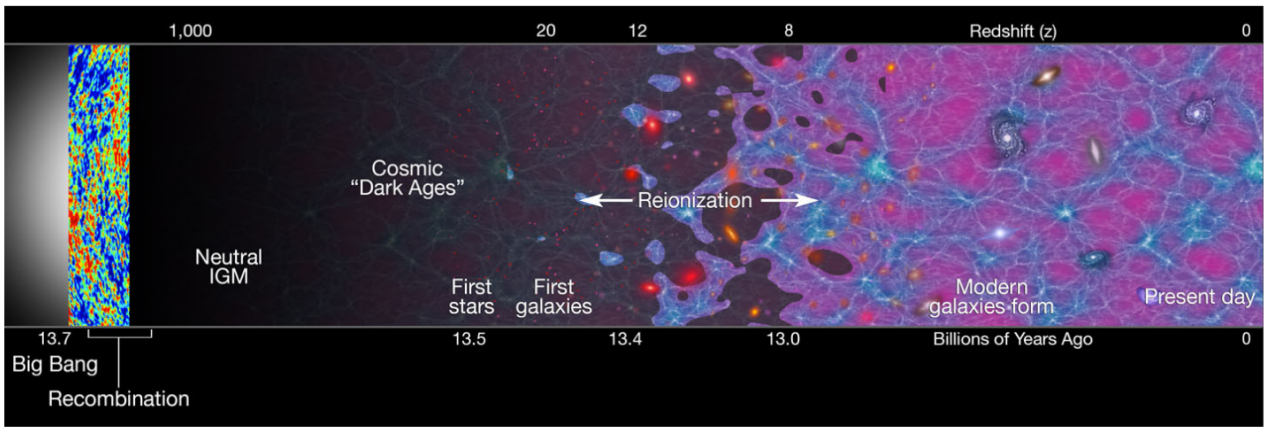

Figure 2.1: Galaxy formation and evolution timeline. After recombination at $z \approx 1100$, hydrogen in the intergalactic medium remains neutral until the first stars and galaxies are formed at $z \approx 15$. These primordial objects produce energetic UV photons capable of ionizing local bubbles of hydrogen gas. As the abundance of these early galaxies increases, H II bubbles overlap and progressively occupy larger volumes. The hydrogen reionization process is completed at $z \approx 6$, approximately 1 Gyr after the Big Bang. Taken from Robertson et al. (2010).

sive galaxies at early epoch (e.g., through cold accretion stream; Dekel et al. 2009) and evidences of a "downsizing" effect (Cowie et al. 1996) indicating that massive galaxies completed the bulk of their star formation faster than lower-mass galaxies. These findings have challenged the sequential picture of galaxy evolution, going from low-mass dwarf galaxies at early epochs to massive spiral and elliptical galaxies at later stages, in favor of an evolutionary scheme in which massive galaxies are formed fast and dominate the star formation in the early universe. Because of their rapid evolution timescales, massive galaxies are difficult to observe in a chemically young phase. On the contrary, since dwarf galaxies assemble and form their stellar masses over a long cosmological time, they can be relatively more easily caught at different stages of their evolution.

Additionally, dwarf galaxies can have complex star formation histories with multiple episodes of quenching and re-ignition of star formation (e.g., Rey et al. 2020). In some local dwarf galaxies, the star formation is delayed which allows them to be observed in a relatively unevolved chemical state, even though their gas reservoir assembled long ago. While evidence suggests that most local dwarf galaxies have already experienced earlier bursts of star formation (e.g., McQuinn et al. 2010; Anders et al. 2004), as opposed to what would be the case in purely pristine environments, the ISM of gas-rich, star-forming dwarf galaxies remains relatively metal-poor and dust-poor. The lowest metallicity galaxies can hence found among the latter dwarf galaxies.

2.1.2 Star formation in pristine gas

Because the first stars form in a primordial gas which has not yet been enriched by stars, dwarf galaxies may help to solve questions regarding the chemical and dynamical evolution of galaxies through cosmic times. In particular, they have been proposed as proxies to estimate the primordial chemical abundances. Specifically, they have been used to derive the primordial value of helium abundance, which was mainly formed during the first few minutes after the Big Bang (e.g., Izotov

et al. 1999; Fernández et al. 2019; Hsyu et al. 2020). Hence, dwarf galaxies are valuable objects to understand the cosmological enrichment in metals and dust (e.g., Nanni et al. 2020) that led to the present-day massive and metal-rich galaxies.

Another pressing question is the possible existence of pathways to form stars in pristine atomic gas. While it is commonly accepted that star formation is triggered by the collapse of cold MCs (see Section 1.1.1), a debate has been installed as to whether low-metallicity star formation may be initiated within a cold dense *atomic* gas, provided that the timescale to reach thermal equilibrium becomes shorter than the molecular gas formation timescale (Glover & Clark 2012c,b; Krumholz 2012; Krumholz & Dekel 2012). While this scenario would provide an explanation for the lack of detected molecular gas in low-metallicity dwarf galaxies (e.g., Taylor et al. 1998; Shi et al. 2015; Cormier et al. 2014; Leroy et al. 2007), it is also possible that H_2 is dissociated by shocks or ionizing radiation in the star formation aftermath, or fully consumed by star formation.

Star formation in a metal-free gas can also be associated with the emergence of massive stellar objects, such as Population III (Pop III) stars. Pop III stars are the first generation, zero-metallicity stars that have been proposed early on in Rees (1978b) in the context of the cosmic microwave background, and in Rees (1978a) to explain the missing mass in galaxy clusters. Such massive stars may provide the seeds required to form massive black holes through direct collapse (Smith et al. 2018; Haemmerlé et al. 2020; Nakajima & Maiolino 2022). Although Pop III stars remain theoretical objects which cannot be directly observed individually, their spectral signature should be detectable with future facilities such as the ELT or the JWST (Katz et al. 2022a; Nakajima & Maiolino 2022). If present, such very massive stars (up to 300 $M_\odot$; Bromm et al. 2009) are expected to strongly affect the heating mechanisms that dominate primordial galaxies, with a particularly pronounced effect in faint galaxies (Berzin et al. 2021).

Star formation processes in a quasi-pristine gas may also be affected by the presence of compact objects, whose distribution and properties are particularly poorly constrained at low metallicity. Such compact objects (e.g., neutron stars and black holes) may experience mass transfers in binary systems, leading to the creation of hot accretion disks. Several studies have suggested that the production of High-mass X-ray binaries (HMXB) might be more numerous in extremely-low-metallicity environments (Mirabel et al. 2011; Basu-Zych et al. 2016; Ponnada et al. 2020), with potential implications on the heating mechanisms that dominate primordial galaxies. Finally, the presence of massive black holes in metal-free environments is also expected to affect the star formation mechanisms, with scenarios indicating that they may locally trigger star formation (Aykutalp et al. 2020). Although many efforts have been invested in the characterization of the ionizing sources expected in the early universe, the lack of observational constraints makes it difficult to identify the exact nature of such sources. (Simmonds et al. 2021; Umeda et al. 2022).

2.1.3 Primordial galaxies and the reionization of the universe

Regardless of the nature of the sources that produce ionizing photons, primordial dwarf galaxies may have played a crucial role during the reionization process. The presence of intense sources of radiation combined with a metal- and dust-deprived ISM may result in the escape of Lyman continuum (LyC) photons into the intergalactic medium (IGM). This LyC-leakage potentially provides a significant

contribution to the reionization process. Indeed, even though dwarf galaxies are the least massive ones, they are, however, the dominant type of galaxies in numbers. Recent simulations indicate that a population of numerous, low-mass, LyC-leaking galaxies with an average escape fractions of 10–20% would be sufficient to reionize the universe by redshift $\sim$ 6 without invoking any other contribution from ionizing sources such as active galactic nuclei (AGN; Robertson et al. 2013, 2015).

Under favorable assumptions concerning ionizing photon production and accounting for a sub-dominant contribution from AGN, Finkelstein et al. (2019) find that even a population-averaged escape fraction below 5% throughout the bulk of the EoR would be enough to match observational constraints. Although the minimum value of the required escape fraction fluctuates depending on the assumptions made in simulations, most recent studies agree on the dominant role of faint, low-mass galaxies to provide the bulk of ionizing photons. Among the latest simulations of the reionization (see Gnedin & Madau 2022 for a complete review) most results agree on the predominant role played by stellar sources in faint galaxies. For example, using the state-of-the-art `Cosmic Dawn II` (with a 100 Mpc box and resolving halos down to $10^8 M_\odot$) to study the relative contribution of galaxies with different halo masses, Lewis et al. (2020) found that galaxies with UV magnitudes between -12 and -19 dominate the reionization between $z = 6$ and 8. Using the `SPHINX` simulations, Rosdahl et al. (2022) found that the escape fraction of UV photons peaks for a population of intermediate-mass ($M_* \approx 10^7\ M_\odot$), intermediate-brightness ($M_{1500} \approx$ -17), and low-metallicity galaxies ($Z \leq 5 \times 10^3\ Z_\odot$).

While the progress of computing power has allowed simulations accounting for larger cosmological volumes, resolving smaller halos, and incorporating complex chemical networks, the task of obtaining observational constraints from the reionization epoch remains difficult. A few recent observations, making the most of facilities such as ALMA (e.g., Hashimoto et al. 2019; Harikane et al. 2020; Tadaki et al. 2022), NOEMA (e.g., Meyer et al. 2022), or the JWST (e.g., Adams et al. 2023; Naidu et al. 2022; Atek et al. 2022) have been pushing the observational limits of detection for faint galaxies after $z > 6$ and up to $z \approx 16$. Nevertheless, such observations have remained limited to a few objects for which only photometric data and/or a few spectral lines are detected. Meanwhile, the results obtained in samples of local and intermediate-redshift dwarf galaxies can help us understand the properties of the ISM in low-metallicity environment.

2.2 Local dwarf galaxies

The lowest gas-phase metallicities in the local universe are found among low-mass galaxies (see Section 2.2.3 for a more detailed discussion). Hence, local dwarf galaxies stand as ideal laboratories to study the mechanisms at play in conditions as close as possible to primordial ones. The results obtained for nearby galaxies may help to gain insights into the ISM properties and the metallicity effects that allow for large escape fractions of ionizing photons in the context of the reionization. The study of dwarf galaxies may also improve our understanding of the physical mechanisms at play in low mass, chemically young objects, with important implications for the high-redshift universe. The numerous samples of dwarf galaxies observed at low and moderate redshift have allowed to perform detailed studies of their ISM. A precise classification scheme was developed for dwarf galaxies based on these observations in the local universe.

2.2.1 Definition

After discussing in Section 2.1 the theoretical specifics of primordial dwarf galaxies, we now intend to present observational criteria that allows us to identify and select dwarf galaxies, with a focus on the nearby ones (i.e., $z \lesssim 0.3 - 0.4$). From an observational point of view, "dwarf galaxies" are galaxies which are significantly smaller and less luminous than the Milky-way-like spiral galaxies and the massive elliptical galaxies. Several definitions can be found in the literature, all related to the following features with respect to more massive galaxies:

- Their compact physical sizes.
- Their low gas and stellar masses.
- Their faint luminosities.

In the recent review from Henkel et al. (2022), dwarf galaxies are defined as galaxies having a visual magnitude above $M_V \approx -18$ but being brighter than a typical globular cluster in the Milky-Way ($M_V << -8$). This definition roughly corresponds to a threshold in terms of stellar mass of $10^9\,M_\odot$, which is classically used in the literature. It is, however, slightly more stringent than the criterion used in other studies (e.g., DGS galaxies have magnitudes down to $M_V \approx -20$). In particular, this definition from Henkel et al. (2022) excludes the Large Magellanic Cloud (LMC) which is slightly too luminous. Regardless of the exact cut in luminosity/stellar mass, we will loosely refer to "dwarf galaxies" as a family of objects sharing common ISM properties, which we now describe.

2.2.2 Morphological classification

Dwarf galaxies form in shallow gravitational potential wells; as a result, their gas reservoirs are quite sensitive to feedback and environmental effects (see, e.g., Henkel et al. 2022). The ISM of dwarf galaxies is most often unstructured and sometimes highly stretched by interactions with other galaxies. They are present under a wide variety of morphologies, with some overlap between those categories (the following is based on reviews from Kunth & Östlin 2000 and Grebel 2001):

- **Dwarf Spirals (dS)** which exhibit clear arm patterns. They are symmetric with both a disk and a bulge component, and are usually more massive than the other types of dwarfs. They are found in galaxy clusters or groups and in the field. They show rotation and are gas rich.
- **Dwarf Irregular (dIrr)** whose surface brightness is well described by exponential profile. Such galaxies are also gas rich, with an extended reservoir of neutral gas, and most often show ongoing star formation in H II regions.
- **Dwarf elliptical galaxies (dE)** which are more gas-poor galaxies and most often host older populations of stars than the irregular dwarf galaxies. They are found close to more massive galaxies. They show no rotation and little gaseous content, with often asymmetric distribution and distinct kinematics.

- **Dwarf spheroidal galaxies (dSph)** exhibit a spheroidal shape and are the faintest and least massive of the dwarf galaxies. They are characterized by visual magnitude above -14 mag. They show no rotation, no central concentration, and have very little gaseous content. Their H I gas is diffuse, usually spatially offset from the optical emission, and they do not show emission associated with H II regions.

- **Blue compact dwarfs (BCD)**, sometimes referred to as H II regions galaxies. They consist of several concentrated stellar clusters and exhibit high central surface brightness. Their name refers to the blue colors (measured in the B-band) of young, hot, and massive stars due to recent bursts of star formation. They are usually found isolated. They show rotation and a large reservoir of gas.

- **Tidal Dwarfs** are a specific category of dwarf galaxies that form out of debris from a merger event or interactions with more massive galaxies. Hence, their characteristics (mass, size, gas content, ...) depend essentially on their progenitors.

In the context of this thesis, we will mostly discuss the properties of **gas rich, star-forming dwarf galaxies which consist of dIrr and BCDs**. Moreover, we will focus on the study of their integrated galactic properties, derived most of the time for spatially unresolved emission lines. Hence, we do not further discuss the structure and dynamics of dwarf galaxies, but instead focus on their peculiar chemistry and its impact on their stellar and gas content.

2.2.3 Primitive chemistry

Mass-metallicity relation

The gas-phase metallicity strongly correlates with the stellar masses and the bolometric luminosities (see, e.g., Kewley et al. 2019). Hence, low-mass, faint galaxies are associated with a low-metallicity ISM. These mass-metallicity (MZ) and luminosity-metallicity relations were first observed in sample of dIrr and BCDs (Lequeux et al. 1979; Kinman & Davidson 1981), before being also observed in more massive disk galaxies. This relation holds for a broad range of stellar masses down to $\sim 10^6\,M_\odot$ (Lee et al. 2006).

In the local universe, the MZ relation is quite steep for dwarf galaxies and up to stellar masses of $10^{10.5}\,M_\odot$, after which the relation flattens. Its slope changes at a characteristic "turnover mass" which depends on galactic parameters (in particular the star formation rate (SFR); Curti et al. 2020). One possible explanation of the steep MZ slope observed in low-mass galaxies is the presence of strong galactic-winds, which may remove metals from low-mass galaxies (e.g., Matteucci & Recchi 2001). Alternative scenarios include low star formation efficiencies in low-mass galaxies caused by supernova feedback (Brooks et al. 2007) and a variable integrated stellar initial mass function (Köppen et al. 2007). At local scales, dwarf galaxies also distinguish themselves from their massive counterparts by showing little radial metallicity variations or even inverted metallicity gradients (Lagos & Papaderos 2013). This may be explained by their short crossing timescales, which allows an efficient mixing of the gas (e.g., Carton et al. 2018)

Whether the MZ relation holds at high-redshift has been investigated up to $z \sim 3$ and several authors have reported a tight correlation between gas-phase metallicity, stellar mass, and an anticorrelation with the sSFR that does not evolve with redshift (e.g., Mannucci et al. 2010; Henry et al. 2013; Sanders et al. 2021). Nonetheless, evidence suggests that at a given stellar mass, high-redshift galaxies tend to be less enriched than local ones, especially among the lowest masses. This effect might be due to an increased relative gas content in high-z galaxies (i.e., higher dilution by inflows; Lagos et al. 2016), a higher efficiency of gas outflows/winds (Chisholm et al. 2015; Tremonti et al. 2004; Lian et al. 2018), or the reduction in the stellar yields driven by a mass dependency of the initial mass function (IMF; Lian et al. 2018).

The lowest metallicity galaxies

In order to probe the physics of star formation in pristine gas (see Section 2.1.2), we would ideally like to detect star-forming galaxies with a zero-metallicity. Unfortunately, such observations are most likely impossible to reach, since evidence suggests that the IGM and the protogalaxies at low-z are themselves already enriched to significant amounts (Rafelski et al. 2012). The best alternative is hence to search for the lowest metallicity objects in the nearby universe. In that view, many efforts have been deployed to identify samples of Extremely Metal Poor (XMPs) galaxies (12+log(O/H) $\leq$ 7.35; see, e.g., Guseva et al. 2015; Almeida et al. 2016; Hsyu et al. 2018; Laseter et al. 2022). Although the exact values depend on the proxies used to estimate the metallicity (see Section 8.2.1), the lowest metallicity galaxies found to date have metallicities down to 12+log(O/H) $\approx$ 7 (Izotov et al. 2019; Laseter et al. 2022). In the local universe, observations of the neutral atomic medium probed by FUV absorption lines toward the massive stars suggest that the H I regions might be even more metal poor than the ionized H II regions (e.g., Kunth et al. 1994; Aloisi et al. 2003; Lebouteiller et al. 2013). Nevertheless, even in the atomic gas, the estimates are suggestive of a minimum metallicity threshold around 12+log(O/H) $\approx$ 7 in the local universe.

The existence of a minimum metallicity was first explained by a self enrichment scenario, in which stars formed in H II regions locally enhance the metallicity of their host galaxy (Kunth & Sargent 1986). Another explanation, motivated by the relatively uniform abundances measured in local dwarf, is that they formed from a previously enriched intergalactic medium which, itself, puts a lower limit on the content of metals in XMPs. Telfer et al. (2002) reported that the IGM was already enriched to metallicities above 6.5 at $z \sim 1.6 - 2.9$. Hence, the quest for the lowest metallicity objects would be limited by the intrinsic metallicity of the IGM at their epoch of formation.

We note, however, that several studies have proposed that galactic outflows could play a predominant role in removing metals from a previously metal rich medium (Matteucci & Recchi 2001). Finally, we cannot rule out the existence of a reservoir of extremely low-metallicity gas which is not star forming and hence not seen in emission. To solve this question, future facilities such as the Square Kilometer Array (SKA) will be extremely useful to identify small galaxies based on their H I content rather than optical emission.

Elemental abundances and dust properties

In the low-metallicity regime, abundance patterns experience drastic changes that need to be accounted for in order to use spectral line fluxes as signatures of physical quantities. The abundance patterns found in local dwarf galaxies are the product of their complex star formation histories (see Section 1.1.2).

For BCDs in the local universe, those variations have been extensively studied (e.g., Izotov & Thuan 1999; Izotov et al. 1999, 2006; Nicholls et al. 2017; Berg et al. 2019). In particular, the nitrogen-to-oxygen ratio N/O has been found to be rather constant at low-metallicity (12+log(O/H) < 7.6), suggesting primary production by massive stars as the main contribution to nitrogen enrichment at low-metallicity (Izotov & Thuan 1999; Nicholls et al. 2017). It then increases at higher metallicity due to the contribution of intermediary-mass stars to the production of nitrogen.

Another important change at low-metallicity in the modification of the depletion patterns. A depletion is the difference between the gas-phase abundance of an element and its intrinsic (gas phase + solid phase) abundance. It arises from elements being locked onto dust grains, which are effectively removed from the abundances in the gaseous phase. Studies comparing stellar abundances to galactic measurements in BCDs have also demonstrated that depletion onto dust grains is reduced in low-metallicity environments (Nicholls et al. 2017). This finding is in line with other studies using FUV absorption spectroscopy to measure depletion patterns (e.g., in the LMC/SMC; Tchernyshyov et al. 2015).

These low depletion patterns are not surprising considering that dwarf galaxies are notoriously dust-poor (Madden et al. 2013; Rémy-Ruyer et al. 2014; Galliano et al. 2021). Specifically, it has been shown that the dust-to-gas mass ratio drops at low-metallicity due to changes in the mechanisms that control dust production. While dust is mainly produced by grain growth in the ISM above a critical metallicity of 12+log(O/H) $\approx$ 8, the low-metallicity regime is instead dominated by dust production from condensation in type II SNe ejecta and AGB stars (e.g., Rémy-Ruyer et al. 2014; Galliano et al. 2021).

Finally, several studies have hinted at a significantly reduced fraction of PAHs in dwarf galaxies (e.g., Jenkins & Wallerstein 2017). This effect has been associated both with a reduced production of PAHs at low-metallicity and an increased destruction due to strong radiation fields (Wu et al. 2006; Gordon et al. 2008; Sandstrom et al. 2010). A detailed overview of the typical metals and dust variations found among local BCDs is provided in Section 5.2.2.

2.2.4 Impact of low-metallicity on the gas reservoirs

Ionized ISM around stellar clusters

Evidence suggests that radiation fields might be harder in low-metallicity galaxies (Croxall et al. 2017; Kehrig et al. 2006). This finding may be explained by changes in terms of ionizing sources which are associated to harder spectra at low-metallicity or changes in how radiation propagates through the gas because photons are selectively destroyed by dust grains depending on their energy. Several studies also found that low-metallicity H II regions are associated with globally higher electronic tempera-

tures and higher ionization parameters (e.g., Pérez-Montero 2014; Morisset et al. 2016; Thomas et al. 2019). Nevertheless, the opposite trend between ionization parameter and metallicity has also been found in other studies (e.g., Ji & Yan 2022) and will be further examined in Section 8.2.3.

Moreover, spectral tracers which correspond to highly ionized species have been reported in samples of dwarf galaxies (e.g., [Fe IV], [Fe V], [Ne V] and even [Fe VI], [Fe VII]; Izotov et al. 2001, 2004a,b, 2021b). The origin of such lines is debated, with two main hypotheses being either the presence of X-ray sources (e.g., Lebouteiller et al. 2017; Schaerer et al. 2019; Simmonds et al. 2021) or the presence of fast radiative shocks (e.g., Allen et al. 2008; Izotov et al. 2012). To be complete, metal-free Pop III stars (introduced in Section 2.1.2) have also been invoked as an alternative scenario to reproduce the UV lines, like He II, observed in some of the most metal-poor local dwarf galaxies (I Zw 18; Kehrig et al. 2015; Ucci et al. 2019 and SBS 0335 052E; Kehrig et al. 2018; Wofford et al. 2021).

Cold and warm neutral atomic medium

In the neutral atomic medium, the energetic balance strongly depends on the metallicity. In particular, the photoelectric effect heating depends on the dust-to-gas mass ratio and on the dust grain properties, which are known to vary in low-metallicity galaxies (see Section 2.2.3). For instance, the drop in abundance of the small grains that carry aromatic features results in a lower heating efficiency. At a metallicity of $1/10\,Z_{\odot}$, Bialy & Sternberg (2019) find that the cold neutral medium (CNM) is on average colder, as compared to the solar-metallicity case, due to a reduced photoelectric effect heating efficiency.

The propagation of hard photons in the neutral gas is also affected in a non-trivial way. On the one hand, it has been suggested that dwarf galaxies may have reduced cosmic ray rates (Bell 2003). On the other hand, the impact of energetic photons associated to X-ray sources or cosmic rays may also have an enhanced effect and might significantly contribute to the neutral gas heating at low-metallicity (Lebouteiller et al. 2017; Bialy & Sternberg 2019).

Molecular gas

Tracing molecular gas in dwarf galaxies is a complex task because H_2 has no emission lines corresponding to transitions in the cold molecular gas and because of the difficulty to detect CO emission. Indeed, most dwarf galaxies are faint CO emitters due to the intense radiation fields which efficiently penetrate into their molecular clouds and photodissociate CO. Hence, very little CO is detected, even in galaxies showing vigorous star formation (e.g., Taylor et al. 1998; Cormier et al. 2014, 2017; Leroy et al. 2007). Additionally, high resolution studies have shown that CO emission tends to reside in clumps of a few parsecs at low-metallicity (Rubio et al. 2015; Oey et al. 2017; Schruba et al. 2017; Shi et al. 2020; Wong et al. 2022). This view is also confirmed by hydrodynamical simulations hinting at a highly fragmented structure of the CO emitting clouds (e.g., Glover & Clark 2012a).

Hence, the estimates of molecular gas masses assuming a standard Galactic CO-to-H_2 conversion factor are strikingly small in low-metallicity galaxies. These findings are hard to reconcile with the intense star formation rates observed in some of these galaxies. This apparent tension between the

available star-forming H_2 reservoirs and the star formation rates can be solved with different, possibly combined, scenarios. One possible option is that the star formation efficiencies are increased in low-metallicity environments (e.g., Turner et al. 2015), leading to vigorous star formation even when little molecular gas is present. An alternative option, already mentioned in Section 2.1.2, would be the existence of pathways to form stars directly in the cold atomic medium. Finally, another scenario, which we will tackle in this thesis, is the presence of large CO-dark molecular gas, in which H_2 remains self-shielded while CO is photodissociated (Madden et al. 1997; Lebouteiller et al. 2019; Chevance et al. 2020b; Madden et al. 2020).

2.2.5 Local dwarf as analogs to primordial galaxies?

Local low-metallicity dwarf galaxies and primordial dwarf galaxies are distinct objects with different physical and chemical properties. Hence, the use of local dwarfs as "analogs" to their high-redshift counterparts in terms of physical and chemical properties is to be taken with caution. Nevertheless, when restraining the analogy in terms of observed emission, the few observations available in the high-redshift universe are showing striking similarities when compared to local dwarf galaxies. In the infrared domain, the few detections of galaxies with $z = 4 - 6$ obtained with ALMA (e.g., Harikane et al. 2020; Tadaki et al. 2022), NOEMA (e.g., Meyer et al. 2022), and APEX (e.g., De Breuck et al. 2019) at $z > 6$ have reported bright dust continuum associated with compact structures, high infrared surface brightnesses, and high [O III]/[C II] and [O III]/[N II] ratios, which are reminiscent of the properties observed in local dwarf galaxies (Madden et al. 2013).

In the optical domain, Izotov et al. (2021a) have shown, by focusing on a sample of $\sim$ 25000 compact star-forming galaxies at $z < 1$ from the Data Release 16 (DR16) of the Sloan Digital Sky survey (SDSS), that the subsample of galaxies with the highest equivalent width of Hβ (>100Å) shared similar galactic properties with high-redshift ($z > 1.5$) star-forming galaxies in terms of: stellar masses, far-UV absolute magnitudes, SFR, sSFR, LyC-photon production efficiencies, UV continuum slopes, [O III] $\lambda\lambda$4959,5007Å/[O II] $\lambda\lambda$3726, 3728Å, [Ne III] 3868Å/[O II] $\lambda\lambda$3726, 3728Å ratios, and equivalent widths. Similarly, Schaerer et al. (2022) have shown that the spectral properties of 3 galaxy candidates from the SMACS J0723.3-7327 field at $z \sim 8$ show strong similarities with known samples of low-z BCDs. While this analogy remains to be further explored (e.g., with the upcoming JWST observations), those findings strengthen the interest in local and intermediate-z dwarf galaxies as possible analogs to primordial galaxies.

However, as previously mentioned in Section 2.1, the physical and chemical mechanisms expected in the primordial universe are peculiar and most likely differ from what is known in the local universe. Whether those similarities in terms of emission lines diagnostics can be translated in terms of common physics and chemistry remains unclear. In particular, the star formation histories in primordial galaxies are expected to strongly differ from the ones in the local universe, since no old population of stars can be present. On the contrary, hints of old stellar populations (>1 Gyr) have been reported in numerous local BCDs (e.g., Aloisi et al. 2001, 2003, 2005, 2007). Other important differences are expected regarding the nature and strength of ionizing sources (e.g., Pop III stars, X-ray sources), the distribution and properties of the dust (e.g., Sommovigo et al. 2020), the dominant heating and cooling mechanisms, and the properties of outflows or inflows throughout the history of the galaxy. All the information gathered in the local universe regarding how dwarf galaxies assemble and form

their stars may not be easily extrapolated to galaxies that formed in a completely different environment. Despite the differences, local dwarf galaxies, which are known to be metal-poor, are still the best option available to understand the physics at play at low metallicity environments that may also be relevant at high-redshift.

2.3 Relevant samples of dwarf galaxies

While numerous samples of dwarf galaxies have been gathered, we limit this section to surveys of BCDs in the local universe, which have collected extensive spectral data in the infrared and optical domains. Those samples are ideally suited to test and apply new modeling tools such as the ones developed in this thesis. In particular, spatially unresolved samples of local dwarf galaxies for which numerous emission lines are available in the IR and optical domains are ideally suited to benchmark new methods to disentangle the emission arising from different phases of the ISM.

In the context of this work, we have focused on a sample of local ($\lesssim$ 100 Mpc) star-forming dwarf galaxies (from the Dwarf Galaxy Survey, presented in Section 2.3.1). The main interest of this sample is to gather IR tracers of both the ionized and neutral gas reservoirs, allowing us to perform a global study of the multiphase ISM. Although rather small ($\sim$ 50 galaxies), this sample has well-known stellar and chemical properties, with some objects close enough to spatially resolve the ISM observations. In Section 2.3.2, we briefly introduce another category of BCDs, located slightly further ($z \lesssim 1$). Although the UV and optical tracers available in emission for the latter objects have, for now, limited their study to the analysis of their H II regions, several of them have been identified as confirmed LyC-leaking galaxies, making them interesting targets for future analysis (see Section 11.2.1).

2.3.1 Star-forming dwarf galaxies in the local universe

Dwarf Galaxy Survey

To investigate the physics and chemistry of low-metallicity environments, several targets are available in the very nearby universe ($\leq$ 100Mpc). The Dwarf Galaxy Survey (DGS; Madden et al. 2013) (see Figure 2.2) gathers some of the most relevant targets in the nearby universe. This sample gathers photometric and spectroscopic observations of 50 nearby (0.5 – 191 Mpc) galaxies observed in the far-infrared (FIR) and submillimeter domains with the *Herschel* Space Telescope (Pilbratt et al. 2010). All of these galaxies were also observed in the mid-infrared (MIR) domain with the *Spitzer* observatory, and spectroscopic measurements are available for all but 5 objects.

A few galaxies in this sample are dwarf irregular galaxies (dIrrs; see classification from Section 2.2.2). This is for example the case of the Small Magellanic Cloud (SMC) and NGC 6822, the two closest of such type. In some cases, like for the SMC, this irregular structure can be the result of interactions with another galaxies. Most of the DGS galaxies are BCDs galaxies (as defined in Section 2.2.2). These galaxies are known to be actively forming stars, resulting in high surface brightnesses and blue colors. The ISM of the DGS galaxies is hence dominated by active regions of star formation, which account for most of the emission in the optical and IR domain. Nevertheless, they are also

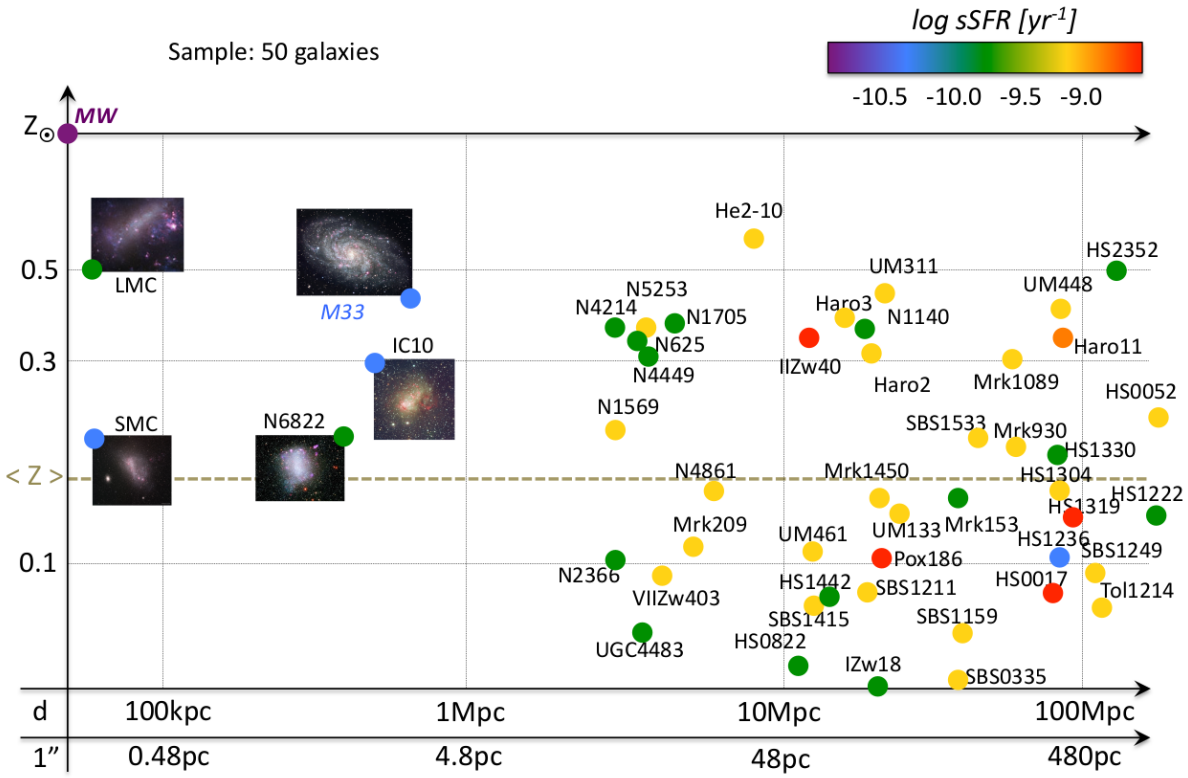

Figure 2.2: Illustration of the *Herschel* Dwarf Galaxy Survey from Madden et al. (2013). The LMC and SMC are the two closest irregular dwarf galaxies for which H II regions and molecular clouds can be resolved with *Herschel*. M33 is a late-type spiral galaxy, NGC6822 an irregular dwarf and IC10 a BCD. In those 3 objects, the molecular complexes can be resolved with *Herschel*. All the other galaxies are BCDs which are unresolved with *Herschel*. We note however that resolved observations have been obtained with other instruments, for example in the optical.

associated with large H I reservoirs that dominate their ISM mass (e.g., with masses ranging from 10^7 to $\sim 10^{11}$ $M_\odot$ in the DGS; Rémy-Ruyer et al. 2014).

As shown in Figure 2.2, the DGS sample spans a large range of sub-solar[1] metallicities ranging from 12+log(O/H)=7.14 (~1/35 $Z_\odot$) up to 8.43 (~1/2 $Z_\odot$), derived from empirical strong line methods (Madden et al. 2013; Rémy-Ruyer 2013). Their dust-to-gas mass ratios also span a large range of values and have been carefully investigated in Rémy-Ruyer et al. (2014, 2015) and Galliano et al. (2021) using continuum measurements from *Spitzer* and *Herschel*. This sample is ideally suited to study local, low-metallicity galaxies with numerous emission lines available. In particular, the spectral features arising from both the ionized and neutral ISM in the DGS have been presented in Madden et al. (2013). The sample has already been modeled using *Spitzer* and *Herschel* in Cormier et al. (2019). We further present and extend this modeling approach in Chapter 9. Meanwhile, other studies have instead focused on detailed studies on individual objects from the DGS.

[1]We use the solar value from Asplund et al. (2009) of 12+log(O/H)=8.69.

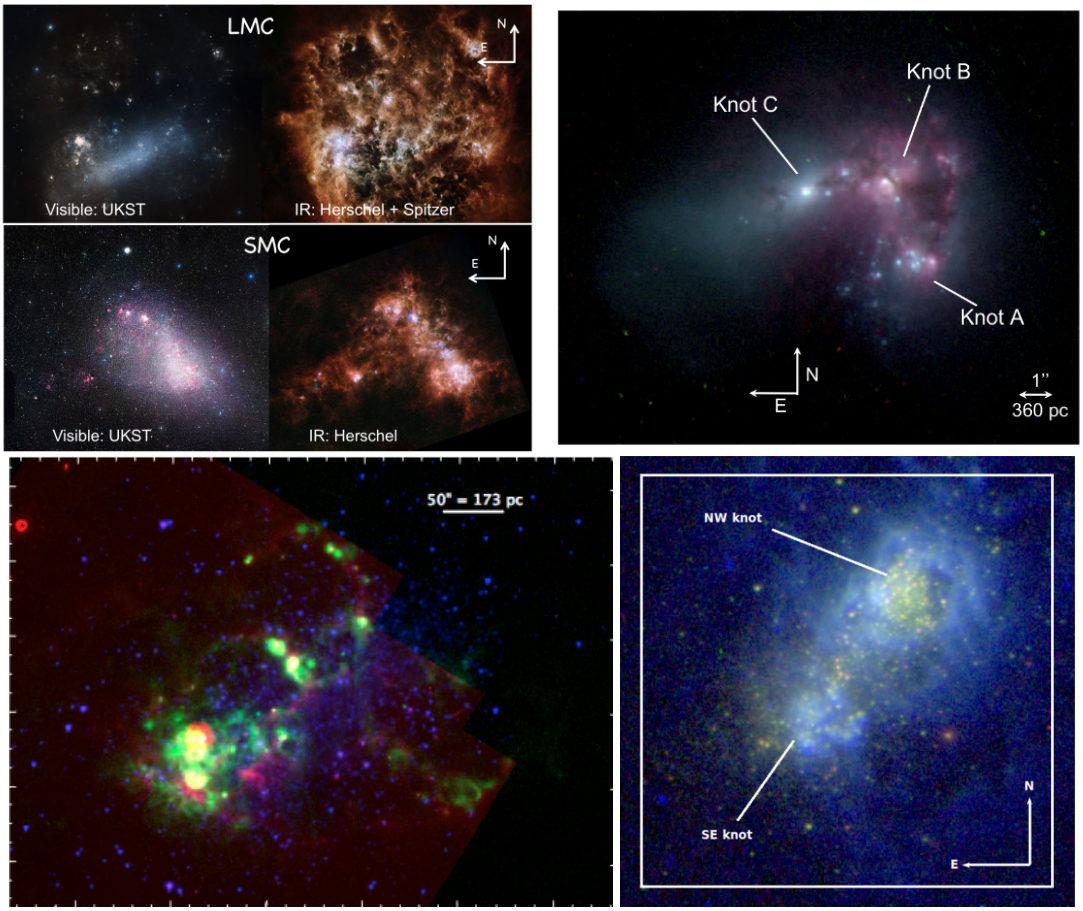

Figure 2.3: **Top left**: LMC and SMC observed by the UK Schmidt Telescope and by the space telescopes *Herschel* and *Spitzer* in IR. Taken from Chevance (2016). **Top right**: HST/ACS 3-color image of Haro 11 with labels for the 3 star-forming knots (blue: F435, green: F550, red: FR656N/Hα). Taken from Cormier et al. (2012). **Bottom left**: Three-color image of IC 10 with *Spitzer*/IRAC4, 8 μm in red, Hα in green, and *Spitzer*/IRAC1, 3.6 μm in blue. Taken from Polles et al. (2019). **Bottom right**: Color composite *Hubble* Space Telescope image of I Zw 18 in three band passes (blue = WFPC2/F658N, green = WFPC2/F555W, red= WFPC2/F814W). Taken from Kehrig et al. (2016)

.

Large and Small Magellanic Clouds (LMC/SMC)

With their sub-solar-metallicity (Z_{LMC}=1/2 $Z_\odot$ and Z_{SMC}=1/5 $Z_\odot$) and their proximity to the Milky-Way (50 kpc for the LMC; Walker 2012, 61 kpc for the SMC, Hilditch et al. 2005), the Magellanic clouds are ideal objects to study the mechanisms at play at moderately low-metallicity with a relatively large spatial resolution (see Figure 2.3). The chemistry of the Magellanic clouds has been studied in detailed using absorption features detected in stellar spectra (e.g., Jenkins & Wallerstein 2017). The authors were able to derive depletion patterns (defined in Section 2.2.3) in the SMC, in which they find significant differences with respect to the Milky-Way, possibly associated with the low-metallicity and the different grain composition and distribution (Weingartner & Draine 2001a).

The proximity of the Magellanic Clouds has also allowed detailed studies of their stellar content, which include massive young stellar objects having a strong impact on the surrounding gas (Oliveira et al. 2019). Several studies have focused on disentangling the different regions and phases of their

ionized and neutral gas reservoir (e.g., Kennicutt et al. 1995; Pellegrini et al. 2012; Lambert-Huyghe et al. 2022). Other studies have focused on the feedback mechanisms at play around massive star-forming regions and their impact on the PDR/molecular gas reservoirs in sub-regions of the LMC such as, for example, 30 Doradus (Chevance et al. 2016, 2020b; Lee et al. 2019; Wong et al. 2022), N159W (Lee et al. 2016; Tokuda et al. 2022), N11 (Lebouteiller et al. 2019), and N44 (Barman et al. 2022). The cold H I distribution in the SMC has been recently mapped within the GASKAP-H I survey (Pingel et al. 2022; Dempsey et al. 2022). The CO distribution has also been mapped out in the northern region of the SMC, with most of the emission arising from compact clouds embedded within envelopes of lower density.

IC10

The irregular dwarf galaxy IC 10 (see Figure 2.3) has a metallicity of 12+log(O/H) = 8.26 (Lequeux et al. 1979), or $\sim$ 2.7 times lower than the solar-metallicity and between that of the SMC and LMC. It is located at 700 kpc. This galaxy is ideally suited to link between small scale analyses (such as the ones presented in the Magellanic Clouds) and the global integrated properties, since it is situated far enough to be fully mapped. Its ionized gas structure was studied in detailed in Polles et al. (2019) which have put in evidence resolution effects on the ISM properties that were inferred, such as the porosity to UV photons.

Haro 11

Haro 11 (see Figure 2.3) is a BCD located at 84 Mpc, with M_B = 20.3, and metallicity $1/3\,Z_\odot$ (Guseva et al. 2012). It is composed of 3 main star-forming regions, or knots, described in Kunth et al. (2003), with a morphology presumably resulting from a merger event (Östlin et al. 2001). This galaxy is thought to host very young stellar clusters (Adamo et al. 2010) associated with a complex ISM structure that corresponds to a low PDR covering factor and a contribution from a diffuse low-ionization gas component (Cormier et al. 2012). In the local universe, this galaxy is one of the few objects for which direct LyC leakage ($f_{esc} \approx 3.2\%$) has been directly detected in the UV (Leitet et al. 2013). Recent observations with HST/COS have revealed that this leakage is inhomogeneous and most likely originates from the knot C (Östlin et al. 2021). This galaxy also shows Lyα emission, possibly associated with the presence of Ultra-Luminous X-ray sources (ULX; Gross et al. 2021). Finally, the molecular gas content of Haro 11 has recently been mapped with ALMA in (Gao et al. 2022) which found that most of the mass is clustered around the central star-forming regions (knots B and C). The estimated gas masses are suggestive of a high star formation efficiency and a high sSFR (Gao et al. 2022).

I Zw 18

I Zw 18 (see Figure 2.3) is the lowest metallicity galaxy in the DGS sample, and it is among the lowest metallicity known in the nearby universe (12+log(O/H) = 7.14 or 1/35 $Z_\odot$; Madden et al. 2013). I Zw 18 is a prototypical BCD galaxy, often put forward as a practical test case for simulations and models of low-metallicity environments. Its ISM has been modeled in details thanks to deep optical

and IR observations (Péquignot 2008; Lebouteiller et al. 2013; Kehrig et al. 2016; Lebouteiller et al. 2017). In particular, the impact of its bright ULX (Thuan et al. 2004) on the surrounding metal-poor gas has been investigated in several studies, leading to an active debate upon its possible role in powering the He II lines observed in I Zw 18 (Lebouteiller et al. 2017; Kehrig et al. 2018; Rickards Vaught et al. 2021; Umeda et al. 2022). Finally, its molecular gas content has also been investigated in Leroy et al. (2007) which reported a stringent upper limit on CO emission, hinting at an extremely low content of CO-bright molecular gas. Recently, Zhou et al. (2021) reported a marginal detection of CO(2-1) obtained with NOEMA. The very low CO content found in I Zw 18 remains to be reconciled with its intense star formation, making it an ideal test case to study star formation in primitive gas.

2.3.2 BCDs and Green Peas at intermediate redshift ($z \lesssim 1$)

While very nearby samples such as the DGS can be studied with many details thanks to their proximity, other samples of somewhat further galaxies are can also be very useful, in particular to study the origin of the LyC-leakage. Specifically, low-z ($z \lesssim 1$) compact star-forming galaxies have recently received a lot of attention due to the possible analogy drawn between their properties and the ones of primordial galaxies (see Section 2.2.5). In particular, a growing number of luminous BCDs, which were targeted with HST/COS in the UV, have revealed LyC-leakage which could be directly observed with substantial LyC leakage reaching values as high as 72% (Bergvall et al. 2006; Leitet et al. 2013; Borthakur et al. 2014; Leitherer et al. 2016; Izotov et al. 2016a,b, 2018a,b; Wang et al. 2019; Izotov et al. 2021c; Wang et al. 2021; Flury et al. 2022a,b). As further described in Section 9.1, the direct observation of LyC-leakage is challenging and restricted to a narrow range of redshift which make LyC-leaking galaxies relatively unique objects to study the mechanisms which may power LyC-leakage in the early universe.

Another interesting population of extremely star-forming galaxies have been detected at $z \sim 0.2$ and have been classified as "Green Pea" galaxies (Cardamone et al. 2009; Jaskot & Oey 2013; Henry et al. 2015; Hogarth et al. 2020; Jaskot et al. 2019). Such objects have compact sizes, very large star formation rates (up to 30 $M_\odot$ yr^{-1}), low stellar mass ($\approx 10^{9.5} M_\odot$), low-metallicity (log(O/H) + 12 $\approx$ 8.7), and large UV luminosities ($\sim 3 \times 10^{10}$ $L_\odot$) (Cardamone et al. 2009). Their denomination refers to their intense emission of the [O III] $\lambda\lambda$4959,5007Å line which is particularly dominant at optical wavelengths, when redshifts move the Hα line out of the optical spectral window. Their properties are somewhat similar to the most luminous members of known BCDs samples.

2.4 Focus of the thesis

In the context of this manuscript, we focused on the following specific questions which directly relate to the wider context presented in Section 2.1:

- How do the metal and dust content affect the ISM porosity to ionizing photons and how do they relate to the escape fraction of photons? (see Section 2.1.3)
- How different is the gas mass distribution in primitive galaxies and how can we estimate the molecular gas reservoir, accounting for the so-called "CO-dark" H_2 gas? (see Section 2.1.2)

To investigate these questions, we have chosen to perform a detailed analysis of a sample of a sample of well-known nearby dwarf galaxies, the Dwarf Galaxy Survey (DGS; Madden et al. 2013), which we presented in 2.3.1.

The wealth of infrared tracers observed in this sample of galaxies is ideally suited to derive constraints on their low-metallicity ISM by comparing the observed emission to what is predicted by ISM models, with an increasing level of complexity. In Chapter 3, we review the existing tools to model the ISM and its emission. We then introduce a new topological approach in Chapter 4, which combines models drawn from a grid of Cloudy models (Chapter 5), tailored for the application to low-metallicity BCDs.

Those combinations of models span a large parameter space, which we explore using Bayesian tools presented in Chapter 6. Specifically, we developed a new code which performs a multicomponent probabilistic grid search (`MULTIGRIS`, Lebouteiller & Ramambason 2022a), which we describe in Chapter 7. The two main applications of this new code have consisted in investigating the ISM porosity to UV photons (see Chapter 9) and deriving the gas masses associated to different ISM phases, including the CO-dark molecular gas (see Chapter 10). This new code opens interesting opportunities to further progress in the comparison between models and observation, which we briefly describe in Chapter 11.

Part II

Modeling the interstellar medium

Chapter 3

State-of-the-art ISM models

In Chapter 2, we presented the main questions that we wish to tackle in this thesis. Specifically, we discussed the interest of studying the primordial universe, with a focus on primordial dwarf galaxies (see Section 2.1) and how local dwarf galaxies may help to reduce observational difficulties and improve our understanding of the physical mechanisms in low-metallicity environments. In Section 2.4, we highlighted specific questions that can benefit from the study of local dwarf galaxies, in which we can disentangle the emission arising from their multiphase ISM, described in Chapter 1. Such analysis relies on the comparison of observations with the predictions made using refined modeling tools.

The current chapter provides an overview of the multiple tools available to model or simulate the gas and dust in galaxies. This field has greatly benefited from the latest advances in computing power. We will limit our description to tools that can be of interest to model the multiphase ISM can help in designing the specific strategy we need. We will specifically discuss the possibility to compare their outputs with observations. The tools presented here come with a wide range of complexity in terms of geometry, chemical network, and treatment of physical processes such as RTs. In Section 3.1, we give here a short overview of the existing tools to motivate the choice of our modeling approach in Section 3.2.

3.1 Modeling tools

3.1.1 Stationary chemical models

Laboratory studies of the transitions (e.g., deexcitation or recombination processes) experienced by chemical elements under a wide range of physical conditions have led to the development of astrochemical models, which can predict, among other things, the expected emission lines within complex chemical networks. The atomic data (e.g., collisional rate, energy thresholds) are stored in large databases (e.g., `CHIANTI` database; Dere et al. 1997; Del Zanna et al. 2021, `KIDA` database; Wakelam et al. 2012, `UMIST database` McElroy et al. 2013). This data can be used by astrochemical codes to predict the expected gas-phase abundances and associated emission line fluxes. Such models usually assume a stationary medium at thermal equilibrium, which is fixed by a set of physical conditions.

Alternatively, the most recent codes can account for time-dependent processes. Although astrochemical codes incorporate complex chemical networks, they, however, do not consider the effects linked to the transfer of radiation within the gas (they can be referred to as 0D models). Hence, the validity of their prediction is limited to small patches of gas in which the stationary approximation is valid. They are, however, not well suited to model gas over large regions or at galactic scales.

3.1.2 1D RT codes

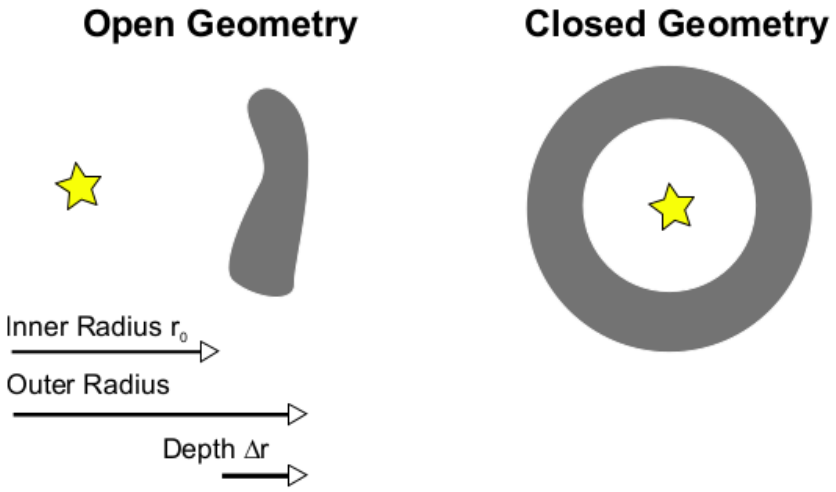

Figure 3.1: Open vs. closed geometry in `Cloudy` (Ferland et al. 2013, 2017) taken from the *Hazy* documentation.

Photoionization and photodissociation codes simulate the RT within a cloud of gas irradiated by an incident source. Such codes are often 1-dimensional and assume a simple geometry (i.e., plan parallel or spherical as illustrated in Figure 3.1) but can incorporate complex chemical networks. They simulate the physical and chemical processes which power emission lines trough cooling, under various physical conditions. Among the most classically used photoionization and photodissociation codes, `Cloudy` (Ferland et al. 2013, 2017) allows a consistent treatment of RT throughout a multiphase medium (i.e., ionized and neutral phases, as described in Section 1.2). Other codes are available, such as the `Meudon PDR` code (Le Petit et al. 2006), which is specifically designed to study irradiated PDR, or `Mappings V` (Allen et al. 2008), which also includes the contribution from shocks. Although the assumed geometry is simple (1D, plane parallel, or spherical), those codes have been extensively used to analyze emission lines arising from galaxies. They are controlled by a limited number of free parameters, which can easily be varied to examine their impact on the predicted emission lines.

Even though the predictions reflect the assumptions made by the user (e.g., abundance patterns, density law, geometry), the comparison with observations is straightforward and allows constraining many physical parameters. Such models have been used in several galactic studies focusing on, e.g., the chemical abundances (Asari et al. 2016; Pérez-Montero 2014; Fernández-Ontiveros et al. 2021), the hardness of the radiation field (Lebouteiller et al. 2017; Pérez-Montero & Díaz 2007; Perez-Montero & Vilchez 2009; Pérez-Montero et al. 2019; Kumari et al. 2021; Fernández et al. 2022), or the presence of AGNs (Satyapal et al. 2021). They have also been used to tackle specific questions related to the structure of the multiphase ISM (introduced in Section 2.4), such as the escape fraction

of ionizing photons (e.g., with `MAPPINGS V`; Bassett et al. 2019, with `Cloudy`; Jaskot & Oey 2013; Stasińska et al. 2015; Plat et al. 2019; Ramambason et al. 2020; Pérez-Montero et al. 2020; Harikane et al. 2020), or the gas mass distribution between different ISM phases (e.g., with `Cloudy`; Cormier et al. 2012, 2015; Cormier et al. 2019; Lebouteiller et al. 2017; Polles et al. 2019; Madden et al. 2020; Lambert-Huyghe et al. 2022, with `Meudon PDR`; Lee et al. 2016, 2019; Chevance et al. 2016, 2020b).

Thanks to their short execution time of the order of minutes, photoionization and photodissociation models can conveniently be used to create large grids of H II regions spanning different physical conditions (such as the grids available on the 3MdB database; Morisset et al. 2015). This database is easily accessible through SQL requests and allows the users to directly access the main outputs from `Cloudy` models, without having to rerun them. This is especially useful for applications that require to use numerous models to cover large parameters space. We note, however, that specific applications may require on-the-fly modeling (e.g., within MCMC algorithms; see Chapter 6). In this case, the computing time required to create each model can become prohibitive. Additionally, the run time can be made significantly longer when including additional gas phases such the neutral atomic and molecular phases. We discuss alternative solutions using machine learning techniques in Section 11.1.2.

3.1.3 3D codes

Pseudo-3D models

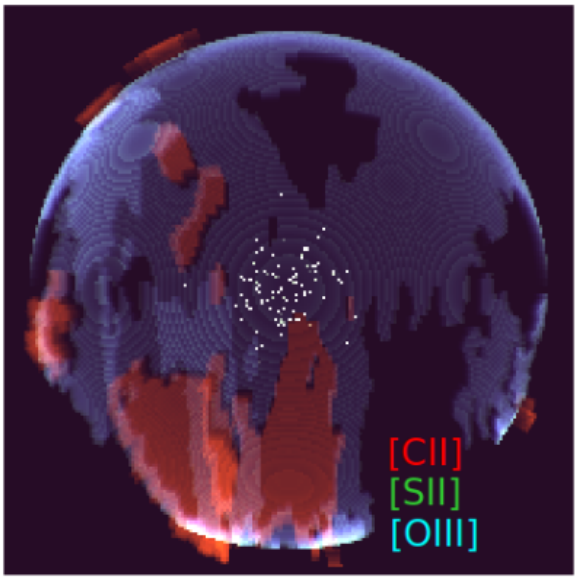

Figure 3.2: Example of a pseudo-3D representation taken from Lebouteiller & Ramambason (2022b): combination of `Cloudy` models which are distributed randomly in 3D according to their relative covering factors and then projected in 2D. The colors respectively represent the emission from [C II], [S II] and [O III]. Screening and extinction effects are ignored. Stars are added for visual purposes only.

Although 1D codes are highly popular and easy to use, they do not account for the 3D structure of the ISM and the 3D distribution of energetic sources. RT codes are notoriously difficult to scale to

3D due to the drastic increase of computing time. An alternative solution proposed in `PyCloudy 3D` (Morisset 2011, 2013) is to build pseudo-3D models by running a set of models, changing parameters following angular laws and interpolating the results in a 3D cube. This can be used for visual analysis and in order to derive predictions based on a more realistic distribution of the gas. An example showing the multiphase and inhomogeneous distribution of the gas that can be achieved with custom visualization tools is presented in Figure 3.2.

3D RT codes

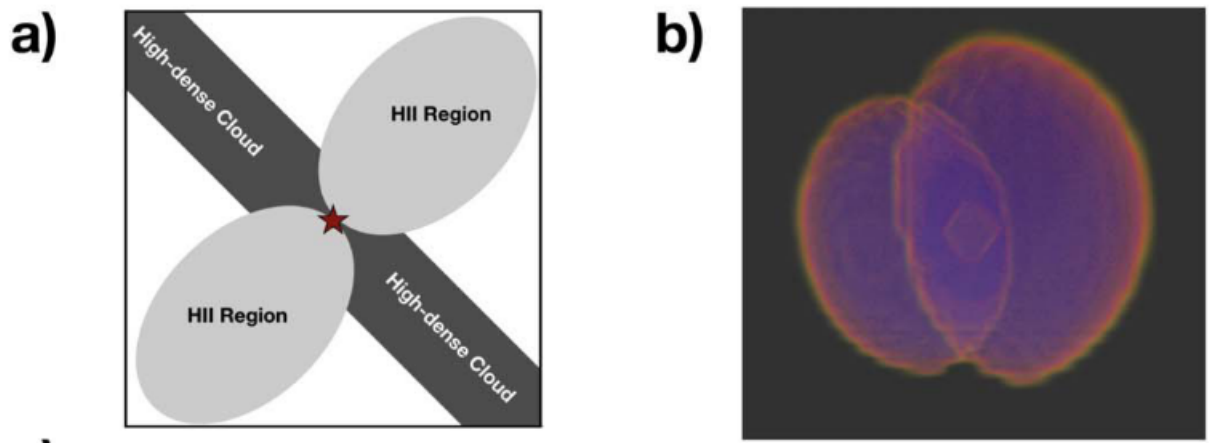

Figure 3.3: 3D model of a bipolar H II region taken from Jin et al. (2022): (a) Schematic figure of a bipolar H II region model. A bipolar H II region consists of two major components: a high-dense sheet-like cloud, the low-dense ionized gas. (b) Three-dimensional visualization of the modeled bipolar H II region.

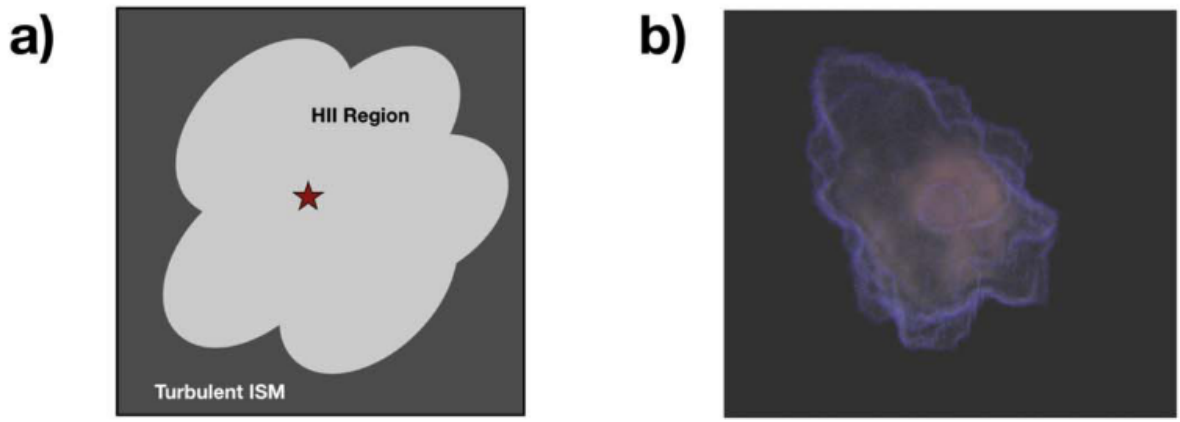

Figure 3.4: 3D model of a fractal H II region taken from Jin et al. (2022): (a) Schematic figure of a fractal H II region model. A fractal H II region is embedded in the turbulent ISM. (b) Three-dimensional visualization of the modeled fractal H II region.

3D RT have been implemented in some recent state-of-the-art codes based on Monte Carlo samplers (e.g., `MOCASSIN`; Ercolano et al. 2005, `TORUS-3DPDR`; Bisbas et al. 2015, `SOC`; Juvela 2019, `RASCAS`; Michel-Dansac et al. 2020, `ART`^{2}; Li et al. 2020, `M`^{3}; Jin et al. 2022, or ray-tracing (e.g., `3D-PDR`; Bisbas et al. 2012, `LOC`; Juvela 2020). Such codes usually account for simplified chemical

networks in order to compensate for the complex geometry and reach a reasonable computing time. The most advanced codes such as M^3 (Jin et al. 2022) are now able to account for the highly complex geometries expected for, e.g., extending bipolar H II regions (see Figure 3.3) or fractal H II region (see Figure 3.4). Nevertheless, building large grids of 3D models covering different geometries and physical conditions remains limited by the large computing time needed. Most importantly, 3D RT require a priori knowledge of the gas and star distribution in the modeled cloud, which applies only to objects with well-known geometries. An interesting pathway consists in extracting the outputs from zoom-in simulations and post-process them with either 1D or 3D RT.

3.1.4 Simulations

Simulations offer the greatest complexity level in terms of geometry of the gas and distribution of sources. Contrary to photoionization and photodissociation codes in which density profiles are assumed, the gas distribution is based on physical laws applied to a dynamically evolving medium that can produce complex ISM structure in mock galaxies or star-forming regions. Subsequently, galaxy simulations incorporate dynamical effects and can track time-dependent parameters. The advance in computing facilities have led to the development of large scale cosmological simulations (e.g., EAGLE; Schaye et al. 2015, Illustris TNG; Weinberger et al. 2017; Pillepich et al. 2018, MUFASA; Davé et al. 2016, 2017, SIMBA; Davé et al. 2019; THESAN; Kannan et al. 2022) and have significantly increased the spatial resolution which can be achieved in zoom-in simulations.

Nevertheless, there are still important challenges that make the comparison between simulations and observations difficult. First, on-the-fly chemistry and RT are difficult to implement in simulations due to large computational cost. Comparing simulations with observed emission lines requires post-processing simulations with a 1D or 3D RT codes (see Sections 3.1.2 and 3.1.3). Such studies are most of the time restricted to the analysis of a few interesting emission lines such as the resonant Lyα lines (e.g., with RASCAS; Mauerhofer et al. 2021; Maji et al. 2022) or a few hand-picked emission lines in the infrared (e.g., Ramos Padilla et al. 2021, 2022; Olsen et al. 2017, 2021; Katz et al. 2022b; Bisbas et al. 2017, 2021), optical (e.g., Hirschmann et al. 2017; Bassett et al. 2019; Garg et al. 2022) or UV domain (e.g., Hirschmann et al. 2019). The analysis of large suites of emission lines with post-processed simulation remains difficult to perform, although numerous efforts are being made to include more and more species and account for a realistic multiphase structure of the ISM in the post-processing (see, e.g., Ramos Padilla et al. 2022).

While the comparison of results obtained with the different methods presented in Section 3.1 is a complex task, exploiting the complementarity of different modeling approaches and comparing possible discrepancies is a crucial future step to be taken. We will briefly discuss possible bridges that can be developed in Section 11.2.2.

3.2 Topological models for unresolved galaxies

As described in Section 3.1.2, single-component models have been extensively used to constrain the physical conditions of H II regions and galaxies. The "average" values that are derived from single-component models are somehow representative of the physical conditions in the component which

dominates the total emission. It provides a good first order approximation of parameters which are not too sensitive to changes in the topology of the gas. However, there is obviously much to gain in combining models compared to using a single 1D model to reproduce the emission arising from different regions of the ISM, which may have drastically different physical conditions. To what extent the use of a simplified geometry in single-component models limits the analysis of more complex, topology-dependent quantities will be discussed in the following subsection.

Although the rapid progresses in 3D RT and simulations might provide, in a near future, powerful complementary tools to classical photoionization and photodissociation models, they are not yet adapted to the analysis of unresolved observations. While assuming an oversimplified geometry, the photoionization models presented in Section 3.1.2 are flexible enough to be combined using various approaches that somehow palliate for their lack of complexity. This approach is especially well-suited to study unresolved observations in which spatial information cannot be recovered, especially in the IR (due to the large point spread function) and/or for high-z galaxies (due to distance). While a few powerful instruments can now provide spatially resolved spectra of local galaxies in the optical (e.g., with VLT/MUSE) and NIR/MIR domain (e.g., with JWST), these observations remain possible for very sparse samples. For most galaxies, the spectra available are spatially unresolved and their analysis must be tailored to deal with this lack of information.

Resolution effects are especially important when studying the structure of the ISM. Indeed, the resolution impacts the observables tracing the spatial distribution of the gas and the mass distribution between different phases (see Section 1.2). The physical processes (e.g., stellar and non-stellar feedback, shocks, turbulence) associated to spectral signatures arising from the ISM involve energy cascades associated with multiscale processes, as described in Section 1.1. Those processes control the energy distribution in the ISM and, in turn, the gas distribution around stellar clusters. The resulting ISM structure is stratified and inhomogeneous. Nevertheless, in a spatially-unresolved galaxy, emission lines are integrated to produce a single spectrum, with a loss of information regarding the spatial distribution of physical parameters.

Topological models aim to provide a parameterized representation of the distribution of matter and phases in a galaxy. In Figure 3.5, we show an illustration of a topological representation associated with the main ISM phases around star-forming regions (see Figure 1.4 and its description in Section 1.2). This approach relies on the assumption that a single spectrum for an unresolved galaxy contains enough information (e.g., various lines emitted from various species in various phases) and that the relative emission varies with the structure which we wish to parameterize. This information can consist of emission line fluxes (used in this thesis) but could eventually include velocity or velocity dispersion (i.e., the line profile).

In essence, topological models are not designed to determine the precise distribution of the gas but only the relative contributions of each phase and sector. As shown in Figure 3.5, we model a galaxy as a sum of representative star-forming regions which can be combined into one single representative model, assuming that the star clusters which dominate the emission share the same properties. While this assumption has a strong impact on the results that are derived (see discussion in Section 11.1.1), it is a classically used hypothesis in galaxy modeling, especially for star-forming dwarf galaxies dominated by young starbursts.

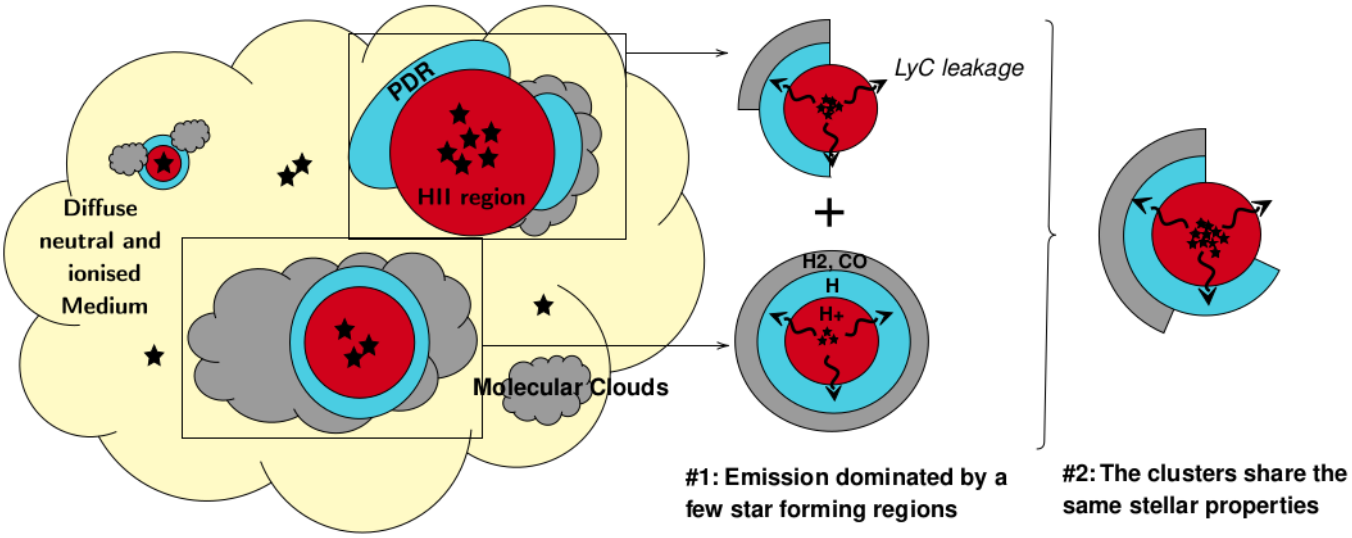

Figure 3.5: Schematic view of the ISM of a starburst galaxy and associated representative topology. We show one possible representative topology associated with the emission of phases around star-forming region (as presented in Figure 1.4). The upper H II region is modeled with a 3-sector configuration, while the lower one is represented with a single sector.

3.2.1 Iterative improvements of topological models

Photoionization and photodissociation models can be combined using different approaches, ranging from very simple two-component scenarios (e.g., Ramambason et al. 2020) up to much more complex topologies. In Figure 3.6 taken from Lebouteiller & Ramambason (2022b), we illustrate different topological models that have been used in previous studies in our group. All the following examples are based on combinations of `Cloudy` (Ferland et al. 2013, 2017) models.

Cormier et al. (2012) (panel a) used a combination of a H II region connected to a PDR to reproduce the emission lines (17 IR emission lines) arising from spatially-unresolved observations of the dwarf galaxy Haro 11 (see Section 2.3.1). The PDR covering factor was defined a posteriori to match the observed [O I] and [C II] emission lines. Additionally, a diffuse ionized gas component was added to fully explain the emission of [N II], [Ne II], and [C II]. Structure-wise, the volume filling factors of the diffuse component, H II region, and the PDR are respectively of 90%, ~0.2%, and 0.01%, hinting at a "porous" ISM (small filling factor of the dense phase) which could be linked to the relatively low-metallicity of Haro 11.

In Cormier et al. (2019) (panel b), the previous method was improved and applied to 38 galaxies drawn from the Dwarf Galaxy Survey (Madden et al. 2013) presented in Section 2.3.1. The combination accounted for either one or two components, whose parameters were determined simultaneously, with a PDR associated only to the component with the highest ionization parameter. The PDR covering factor was, again, defined as an a posteriori scaling factor applied to match the [O I] and [C II] lines. We note, however, that this covering factor was tabulated and considered as an additional free parameter within a large grid of models. This has allowed an automatic search of the best fitting models based on a χ^2 minimization. One of the main results in Cormier et al. (2019) is that the PDR covering factor increases with metallicity. This finding hints at an increased porosity to ionizing photons at low metallicity. Nevertheless, the conversion of the PDR covering factors into $f_{\rm esc}$ estimates

could not be performed, as the models did not include any density-bounded regions. The study presented in Chapter 9 is a direct continuation of their work, which aims to reproduce and extend their results using a new statistical method.

In Polles et al. (2019) (panel c), density-bounded regions were included for the first time to analyze spatially resolved observations of the starbursting galaxy IC 10 (see Section 2.3.1). We note that this study was limited to models with single-component H II regions with no PDR included. In this case, the parameter used to probe the ISM porosity was the physical depth of the models. Contrary to the previously mentioned studies, the models were compared to emission lines extracted from different regions of the galaxy rather than to the integrated emission. Their results have demonstrated that the derived porosity depends on the spatial scale, with most clouds being matter-bounded at small scales (~60 to 200 pc), while larger regions become more and more radiation-bounded at galactic scales. This result highlights the complexity of the ISM in which the energy produced by feedback is deposited at various spatial scales and over different dynamical timescales, hence producing a highly inhomogeneous internal structure (see Section 1.1).

In Lambert-Huyghe et al. (2022) (panel d), models were also applied to observations of the LMC and SMC (see Section 2.3.1). The observations were spatially resolved (~ 5 pc) and individual pixels were modeled as a combination of `Cloudy` models. The approach used to quantify the contribution of the PDR to the emission was slightly different from the one used in Cormier et al. (2012) and Cormier et al. (2019). Among other, it was motivated by the fact that the PDR covering factor is ill-defined at the scale of a pixel. Hence, the PDR models and H II regions models were computed independently and combined using a weight (c_{PDR}) to represent the relative contributions of PDRs along each line of sight. We find the derived c_{PDR} to be spatially anti-correlated with the [O III]88μm line emission, emitted only by H II regions, but uncorrelated with the two PDR tracers used in this study ([C II]158μm and [O I]145μm). Hence, we interpret c_{PDR} as a local indicator of the *relative* contribution of H II regions vs. PDR phases along each line of sight. Those results underline the complexity of disentangling different ISM phases in the case of resolved observations.

Finally, in Lebouteiller et al. (2017) (panel e), a complex model combining up to four components and including the contribution from density-bounded regions, and an X-ray source was used to model the prototypical, low-metallicity dwarf galaxy I Zw 18 (see Section 2.3.1). This study was inspired by the work of Péquignot (2008) with a conversion of the `Nebu RT` code to `Cloudy`, and with the inclusion of new IR emission lines, X-ray constraints, and dust treatment. Structure-wise, the model showed that the covering factor of the sector reaching into the molecular gas must be lower than ~ 0.1%, corresponding to sub-parsec size clumps, thereby confirming the porosity of the ISM in an extremely metal-poor environment. While the model is relatively more complex compared to other efforts, the topology has not been inferred statistically but rather from a partly-manual convergence process. A first application of our new code presented in Lebouteiller & Ramambason (2022b) has consisted in reproducing and extending the results obtained for I Zw 18 but in an automated way. Those results will be presented in Section 9.4.1 and 10.3.4.

3.2.2 Limitations of the χ^2 method

While the topological models used in our team have been iteratively improved in order to reach high levels of complexity, the statistical methods used to compare predictions and observations have remained limited to manual tuning of parameters (Lebouteiller et al. 2017) or frequentist approaches. More specifically, previous studies (Cormier et al. 2012, 2015; Cormier et al. 2019; Polles et al. 2019; Lambert-Huyghe et al. 2022) have extensively relied on the use of the χ^2 method in order to find the best-fitting combinations of models. We report here some of the most urgent limitations associated with the use of the χ^2 method, that are summarized in Lebouteiller & Ramambason (2022b):

- Upper limits were often ignored or else treated in a binary way (i.e., either the model is above or below the upper limit).
- The χ^2 distribution is difficult to translate robustly into a confidence/credible interval.
- The number of degrees of freedom is difficult to estimate, even for linear models (Andrae et al. 2010). Some tracers may trace the same physical parameter while a given tracer may constrain several parameters, and some parameters are correlated, in which case two tracers may constrain more than two parameters. This difficulty increases when priors need to be considered.
- The type of constraints (ratios and/or absolute values) and choice of tracers (ignoring or adding some tracers) often produce erratic probability density functions with solutions that are not compatible with each other within errors.
- Due to the large number of free parameters and to the difficulty in assessing similar or degenerate solutions, it was difficult to consider combinations of more than two components (i.e., either dealing with more sectors and/or more stellar populations).
- Priors could not be used to inform the model of known quantities (with their uncertainties).
- Degeneracy and correlation between parameters are difficult to account for.

The work presented in this thesis is an attempt to overcome limitations of the statistical method used to compare models to observations and the many caveats of single-component models. To further make progress on the statistical method, we developed a new framework based on a Bayesian statistics, which will be presented in Chapter 7. This code is designed to constrain the parameters associated with new, flexible topological models which we now present.

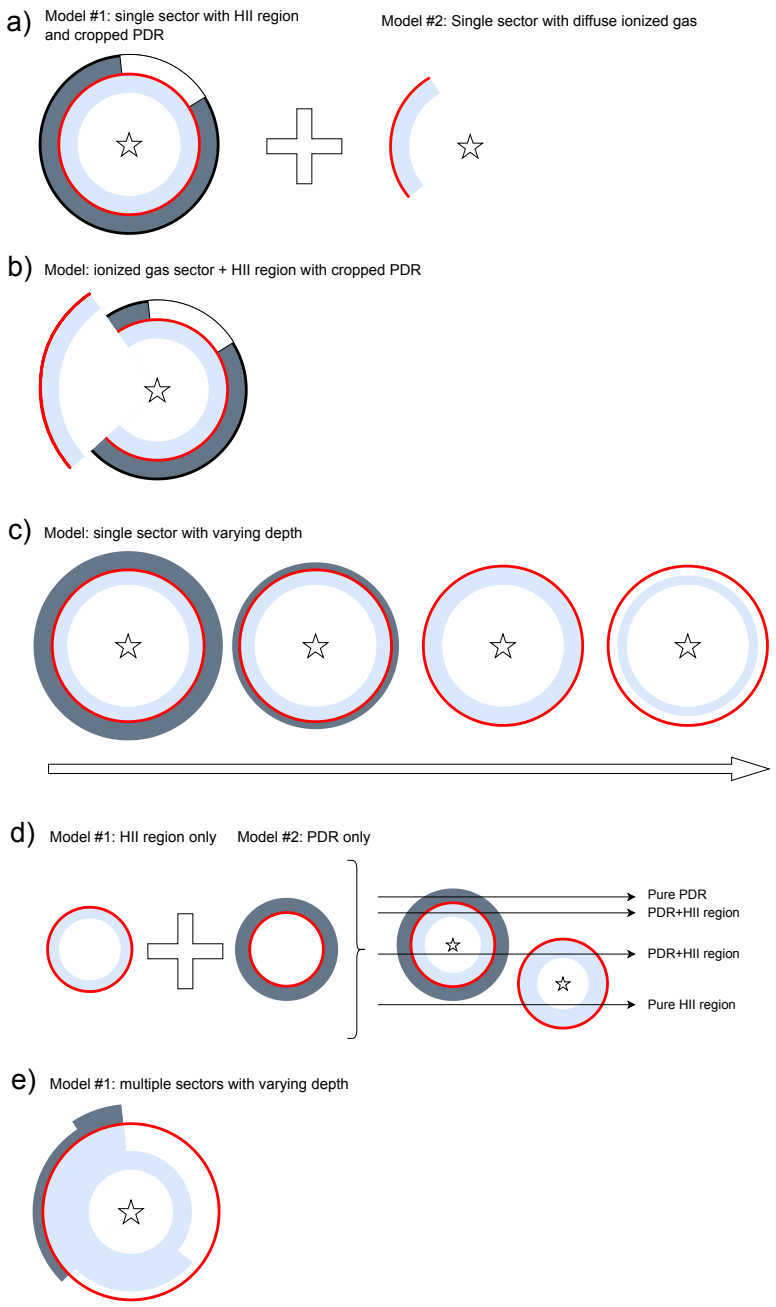

Figure 3.6: Illustration of the different topological configurations considered in previous studies, taken from Lebouteiller & Ramambason (2022b). For each approach, we show the relative contributions of sectors, each corresponding to a specific `Cloudy` model. These relative contributions represent the contribution of each 1D model to the integrated galactic emission. The light blue shell corresponds to the ionized gas, the dark gray shell to the neutral gas, and the red arc to the ionization front. In Cormier et al. (2012) (a), the model used an H II region connected with a PDR, with the PDR covering factor scaled a posteriori (cropped shell in white), and an additional model is considered a posteriori to account for the diffuse ionized gas. In Cormier et al. (2019) (b), two sectors are considered, with the PDR associated only with the high-ionization parameter sector. In Polles et al. (2019) (c), models use a single sector with a varying depth. In Lambert-Huyghe et al. (2022) (d), observations are compared to a mix of pure H II region and PDR components drawn from the same `Cloudy` radiation-bounded model. In Lebouteiller et al. (2017) (e), up to four sectors were considered with different depths.

Chapter 4

Defining flexible topological models

In Chapter 3, we presented existing tools that allow to model the multiphase ISM and discuss their pros and cons. We then introduced a modeling strategy that has been iteratively improved in our group and that we will use in this thesis: the topological models. Topological models can be defined under various assumptions regarding the number of components and priors used to constrain or link their parameter values. In previous studies, such assumptions were chosen *a priori* by the user and could not be changed easily.

One of the advantages of using a Bayesian framework is that we can compare different configurations (described later in Section 7.5) in order to choose the best-fitting one. To fully take advantage of that new possibility, it becomes important to define topologies in a flexible way. This chapter describes with more details the approach used to combine sectors into a new type of topological models.

4.1 Configurations

In this section, we present general configuration set-ups which can be used with any kind of models that simulate the exposition of gas to an incident source of radiation (e.g., any code from Section 3.1).

Currently, those configurations can be used with the two `Cloudy` grids provided with the code that will be introduced in Chapter 7: one small grid drawn from the BOND database (Asari et al. 2016) and our own SFGX grid, which will be described in Chapter 5. Slightly more complex configurations, that include priors on some parameters, will be used for the applications on the SFGX grid presented in Chapter 9 and 10. Ongoing works using other grids will be described in Section 11.2.1.

Table 4.1: Example of pre-defined configurations in `MULTIGRIS` (see Chapter 7)

	`MULTIGRIS` syntax
Single-cluster, multisector (1CnS)	BEGIN configuration <name> n_comps n groups 0,1,..,n END configuration
Multicluster, multisector (mCnS)	BEGIN configuration <name>

Continued on next page

Table 4.1 – continued from previous page

	Syntax
	n_comps n×m groups [[0,1,..,n_1], [n_1+1,..., n_2], ...] END configuration
Distributions of parameters	BEGIN distributions parameter 0: nearest neighbor interpolation, 1: linear interpolation, plaw: power-law or broken power-law(*), normal: normal or double-normal(*) END

(*): a broken power-law or a double-normal are considered when two sets of parameters are given after the keywords plaw and normal, respectively.

4.2 Single cluster, multisector

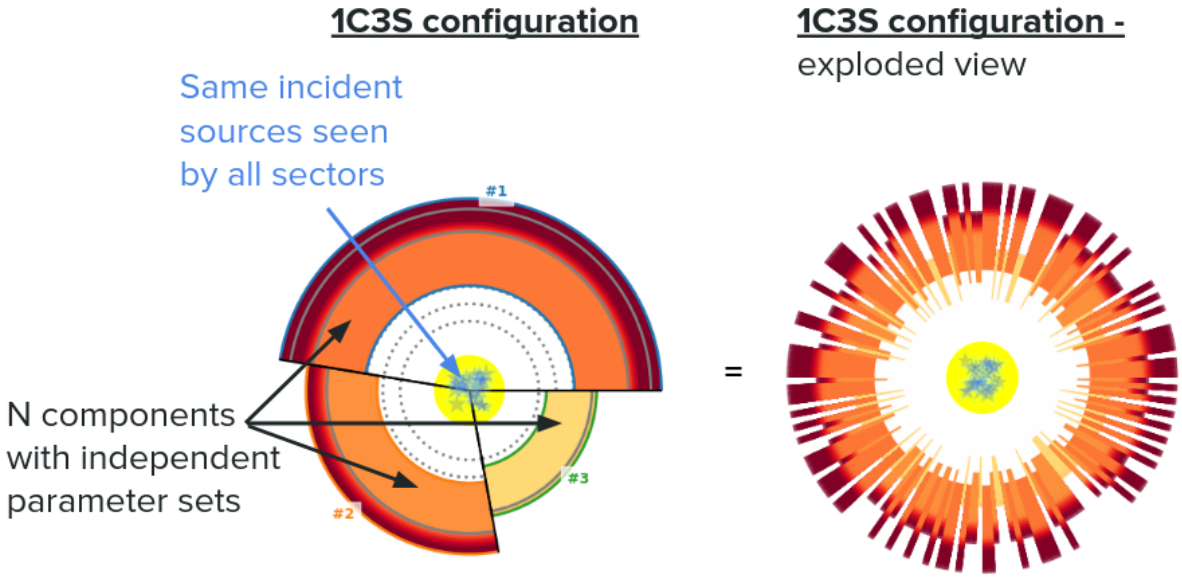

Figure 4.1: Single cluster, multisector representative topology. The left plot shows the integrated covering factors while the right plot shows the same topology but with a different, randomized, geometry. The color scales from yellow to red as the gas volume density increases. Different inner radii indicate different ionization parameter (U) values, with dotted arcs showing U = 0, 1, ... The gray arcs show the photoionization and photodissociation fronts. The depth is not drawn to scale.

A single cluster, multisector configuration is defined by a model having several components considered as sectors arranged around one single cluster. This configuration mimics the case of a galaxy which is dominated by one population of stars originating roughly from the same burst. We consider a single incident SED which is seen by all sectors. As shown in Figure 4.1, this configuration is defined by the number of components ($N_{\rm comps}$) which, in this case, corresponds to the number of sectors. For each sector, the primary parameters (see, e.g., Section 5.2 for a description of primary parameters of the SFGX grid) are independent and can vary freely. On the contrary, we force the parameters

controlling the incident source of radiation to be the same for all sectors. We emphasize that the exact arrangement of sectors around the source is a purely arbitrary choice, which has no incidence on the predicted integrated quantities. Hence, the exploded view represented on the right-hand side of Figure 4.1 is strictly equivalent to the non-exploded view in terms of topology.

4.3 Multicluster, multisector

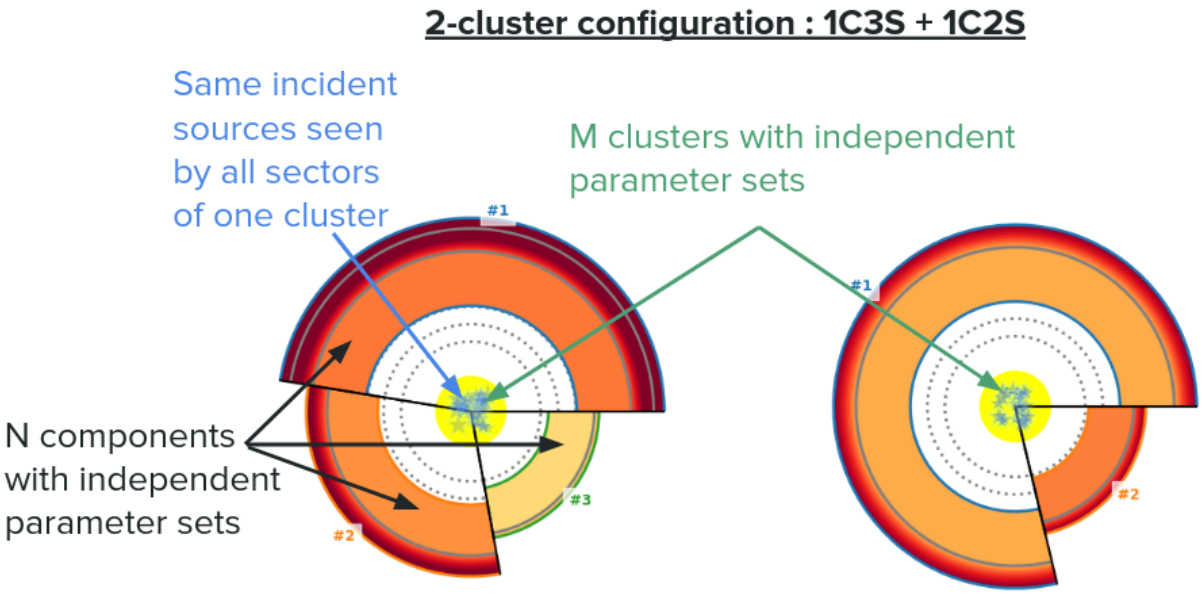

Figure 4.2: Multicluster, multisector representative topology with the same representation as in Figure 4.1.

A multicluster, multisector configuration is defined as a model having several components considered as sectors arranged around several clusters, all independent of each another. This configuration mimics the case of a galaxy in which several populations of stars significantly contribute to the emission. It is for example well suited to model a galaxy having an older stellar population and a young burst. As shown in Figure 4.2, this configuration is defined by the number of components ($N_{\mathrm{comps,i}}$) distributed around the i^{th} out of M independent clusters. We stress that each cluster can itself include a different number of sectors. In this configuration, all the parameters are independent, including the parameters controlling the incident SEDs. Each cluster consists in a single cluster, multisector model.

4.4 Distributions of parameters

Although our multisector models allow us to reach a higher level of complexity than single 1D models, the topology we consider remains oversimplified. To avoid considering models with a very large number of free parameters, we must limit ourselves to combinations of a few clusters/sectors only. Another possibility is to consider a distribution of parameters instead of simple combinations of clouds. The code currently supports normal and power-law distributions, which can be applied to any parameters. In this case, the inferred parameters are the parameters controlling the distribution (e.g., slope and boundaries of the power law, or center and width of the normal distribution).

This approach is inspired by the Locally Optimally emitting Clouds (LOC) models presented in Richardson et al. (2014, 2016) which consider clouds following distributions of density and ionization parameters. The use of a distribution to describe the gas density is indeed motivated by several theoretical studies. In particular, in a gravoturbulent ISM, it has been shown that molecular clouds follow a log-normal distribution at low density but are better described by a power-law distribution in the dense, star-forming regime (e.g., Offner et al. 2014; Burkhart 2018; Burkhart & Mocz 2019). A power-law distribution in mass of CO emitting clouds have also been reported in observational studies of nearby molecular clouds (e.g., Marsh et al. 2016). This approach opens promising avenues which will be discussed further discussed in Section 10.3.5.

Predefined configurations are proposed for the SFGX and BOND grids provided with the code. We will describe in the following chapter the SFGX grid of `Cloudy` models which will be used in the applications from Chapters 9 and 10. Predefined configurations are proposed for the SFGX and BOND grids provided with the code. We now describe the SFGX grid of Cloudy models which will be used in the applications from Chapters 9 and 10.

Chapter 5

Grid of models SFGX

In Chapter 4, we described the different configurations that can be considered in our topological models. All those configurations consist of combinations of either a discrete number of components or of a very large number of components linked through distributions of physical parameters. In all cases, each component is drawn from a pre-computed grid of models, which is an input of the code.

In the context of this thesis, we developed a large grid of models, specifically tailored to allow the study of the multiple phases of the ISM (see Chapter 1) in low-metallicity dwarf galaxies (see Chapter 2). To model the emission arising from the ISM such galaxies, we built a large grid of models (see Section 5.1), designed to study unresolved Star-Forming galaxies with X-ray sources (SFGX). This grid covers a wide range of parameters expected in samples of star-forming galaxies, including in the lowest-metallicity regime, going as low as $1/100\,Z_{\odot}$, and includes a potential contribution from X-ray sources (see Section 5.2). We also store other important quantities (see Section 5.3) which could easily be extracted from models and that will be useful in the applications from Chapters 8, 9, and 10.

5.1 Overview

This grid was created by running 96 000 `Cloudy` v17.02 (Ferland et al. 2017) models using the "cumulative" option which saves the physical properties and predicted emission lines at each depth of the models. Each model was then cut *a posteriori* in 17 sub-models (see Section 5.2.3). The final grid gathers a total of 1 632 000 submodels, which are available for use here: https://gitlab.com/multigris/mgris_sfgx.

For each Cloudy model, we saved the cumulative intrinsic and emergent emission lines of 516 emissions lines and instrumental bands (the emission is integrated accounting for the instrument band pass profile). The full list line is provided in Appendix (see Table A1). We also save physical and chemical properties of each model obtained at the last iteration of `Cloudy` runs in a HDF5 file. The choice of a HDF5 storage allows us to manage the numerous outputs from `Cloudy` into different categories (or datasets) which can be accessed separately without reading the whole HDF5 file. Each HDF5 contains 6 datasets, which are presented in Figure 5.1. Although the HDF5 datasets are compressed using *bzip2*, the resulting size of each file is still of ~50Mb. Hence, our grid of models has a total size of nearly 5Tb.

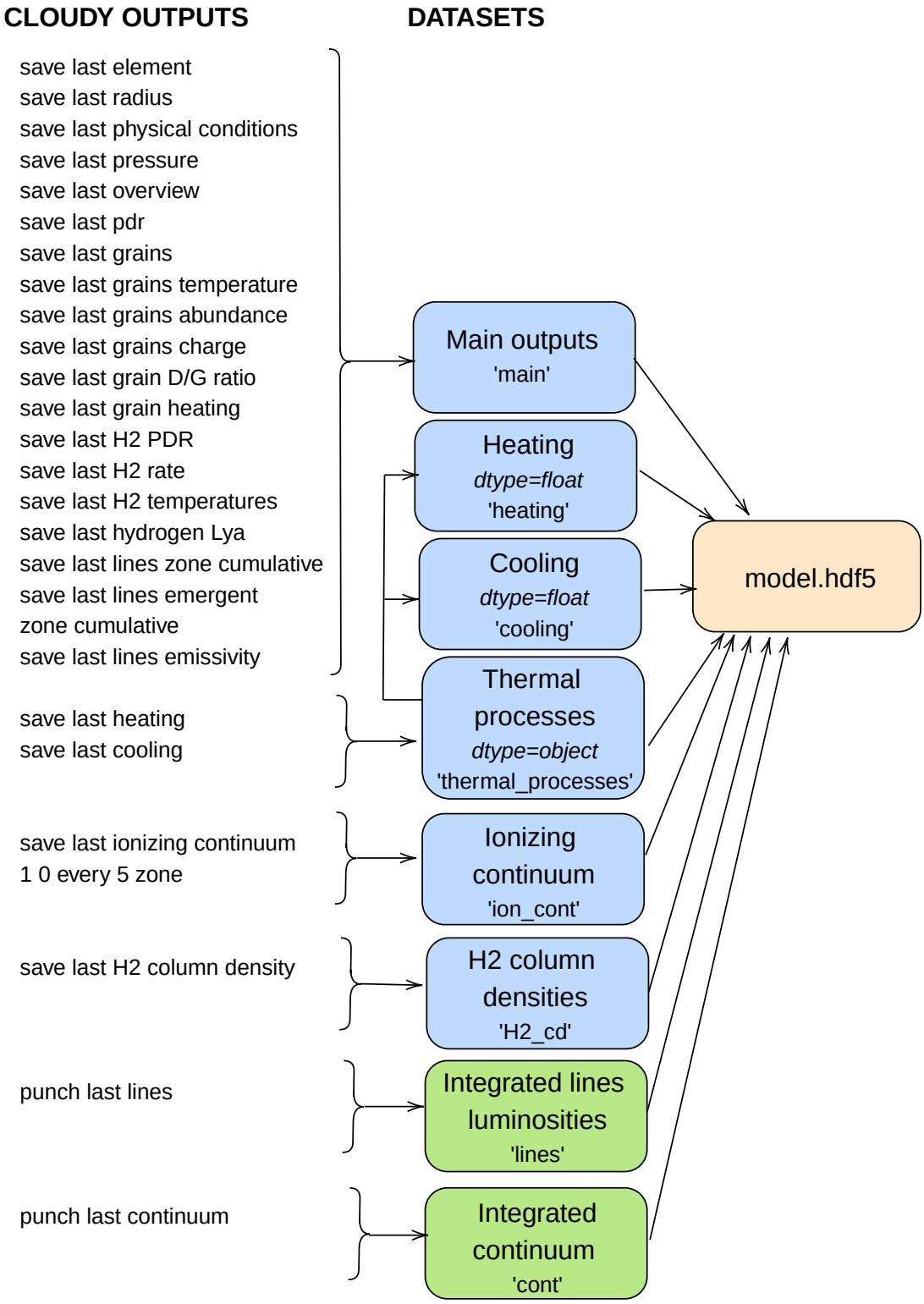

Figure 5.1: Structure of the datasets in each HDF5 file associated to a `Cloudy` model. The blue squares corresponds to outputs stored *at each depth* of the model. The different output files are then merged on the column "#depth" before being stored in the data set. The green squared correspond to integrated quantities saved only for the maximum depth of a model. The keyword "last" saves the last iteration of `Cloudy` runs.

The choice of using a large grid of models and storing numerous outputs is motivated by two aspects:

- First, we aim to cover a large range of physical conditions. As opposed to single-component models (see 3.1.2), covering the range of typical averaged physical expected conditions is not sufficient when performing combinations of models (see Section 3.2). Indeed, a given sector representing only a small part of the galaxy, can have extreme physical conditions (e.g., high density, high ionization parameter) without affecting significantly the integrated emission due to a small covering factor.

- Second, the time spent running a `Cloudy` model is one of the limiting aspects of our analysis. With a run time of ∼30 min per model, the current grid represents ∼48 000 h of computing time. With that in mind, the trade-off we make by storing numerous outputs, which require a large storage, allows us to avoid having to re-run new models for future applications. New techniques based on machine learning algorithms are currently being tested as a way to provide a finer sampling of the parameter space without having to re-run models. They will be discussed in Section 11.1.1.

We now present the various physical and chemical parameters of our SFGX grid. We separate them in two categories: the primary parameters are the input parameters which are used to identify models, while the secondary parameters are physical quantities which we can derive from the set of primary parameters. The secondary parameters are extracted from the `Cloudy` outputs stored in HDF5 files and tabulated. The selection of secondary parameters we present here was made to match the purpose of our two main scientific applications: the study of the escape fraction of ionizing photons (see Chapter 9) and the CO-dark molecular gas component (see Chapter 10). Additional secondary parameters could be added to this list by post-processing the HDF5 files without having to rerun the `Cloudy` models.

5.2 Primary parameters

Each model consists of a spherical shell of gas placed at a fixed inner radius of the incident radiation source. We use a closed, spherical geometry (see Figure 3.1) taking into account the transmitted and reflected radiation, assuming a unity covering factor of the gas. The RT is computed along each line of sight (1D) continuously throughout the H II region, PDR, and molecular zone. We summarize the main input parameters and the range of values spanned for each parameter, leading to a total number of 1 632 000 models in Table 5.1.

Table 5.1: Input parameters of Cloudy models.

	Fixed parameters
Initial Mass Function (IMF) of the stellar population	BPASS v2.1[1] (Eldridge et al. 2017) Broken power-law IMF α_1 (0.1 $M_\odot$ - 0.5 $M_\odot$) = -1.30 α_2 (0.5 $M_\odot$ - 100 $M_\odot$) = -2.35 Single burst, including binary stars
X-ray component	Multi-color blackbody (Mitsuda et al. 1984), T_{out}=10^3K, variable T_{in}
Other fields	Cosmic microwave background (at z = 0) Cosmic ray background[2] $6\times10^{-16}s^{-1}$
Grains	SMC graphite/silicate model (Weingartner & Draine 2001a)
q_{PAH}	Single value scaled with metallicity based on the analytical fits from Galliano et al. (2021)
X/O profiles[3]	$N^{(1)}$, $C^{(1)}$, $Ne^{(2)}$, $S^{(2)}$, $Ar^{(2)}$, $Fe^{(2)}$, $Si^{(3)}$, $Cl^{(4)}$
Other elements	table ISM from Cloudy
Density radial profile[4]	n_H = $n_H\times$ (1+N(H)/[10^{21} cm^{-2}])
Stopping criteria	A_V=10 or T_e=10K
	9 varied parameters
$\log(L_*/L_\odot)$	[7, 9]
L_X/L_*	[0, 0.001, 0.01, 0.1]
$\log(T_{X,in})$[5] [K]	[5, 6, 7]
$\log(Z)$ [$Z_\odot$][6]	[-2.186, -1.885, -1.186, -0.885, -0.664, -0.487, -0.186, 0.115]
Z_{dust}	Three values scaled with metallicity based on the analytical fits (median, upper and lower values from the 95% envelop) from Galliano et al. (2021)
Age of burst [Myrs]	[1,2,3,4,5,6,8,10]
$\log(n_H)$ [cm^{-3}]	[0,1,2,3,4]
$\log(U_{in})$[7]	[-4, -3, -2, -1, 0]
cut[8]	[0,4], step=0.25

(1) Metallicity bins in BPASS are Z = 10^5, 10^4, 0.001, 0.002, 0.003, 0.004, 0.006, 0.008, 0.010, 0.014, 0.020, 0.030 and 0.040 (i.e., 0.05% Solar to twice Solar). We note that $Z_\odot$=0.02 while the default value we use here is $Z_\odot$=0.014.
(2) Following Cormier et al. (2019) we use a cosmic ray rate of ~ 3 times the default rate from Indriolo et al. (2007).
(3) References for elemental abundances. (1): Nicholls et al. 2017, (2): Izotov et al. 2006, (3): Izotov & Thuan 1999, (4): fixed at -3.4
(4) The radial density follows a constant pressure law in the H II region and scales with the hydrogen column density above 10^{21} cm^{-2} following the same prescription as in Cormier et al. (2019). This law is further described in Section 5.2.3.
(5) We note that T_X is varied only for models with an X-ray source (i.e., $L_X/L_* > 0$). This results in a total of 86400 models with an X-ray source and only 9600 models without any X-ray source.
(6) In all models, we assumed that the stellar metallicity and the gas-phase O/H ratio are scale with the factor with respect to solar values from BPASS and Cloudy: $\log(Z_*/Z_\odot) = \log([O/H])$.
(7) The input ionization parameter, U_{in}, is set by varying the inner radius for a given luminosity. It is described in Section 5.2.3.
(8) The "cut" parameter controls the stopping depth of a given model. It is described in Section 5.2.3.

5.2.1 Incident radiation fields

In our grid we consider a stellar component and an X-ray component as represented in Figure 5.2. The stellar population consists of a single burst population from BPASSv2.1 (Eldridge et al. 2017) including the contribution from binary stars. We also include a contribution from cosmic rays and the cosmic microwave background at redshift 0 (see Table 5.1).

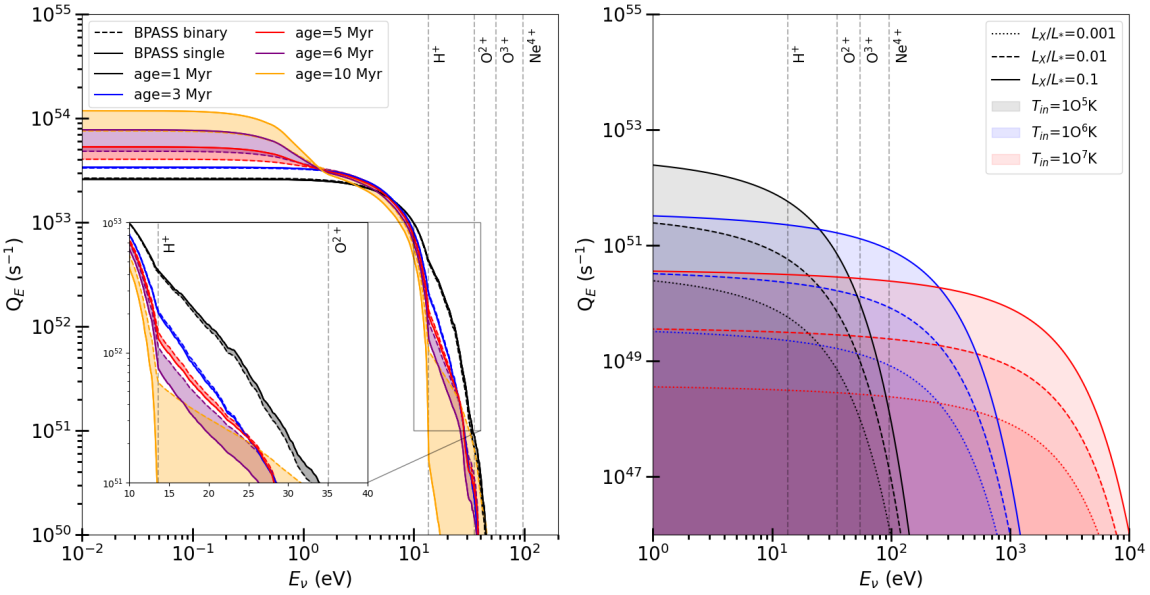

Figure 5.2: Stellar (left) and X-ray (right) incident SEDs of `Cloudy` models for a solar metallicity and bolometric luminosity of $10^9\,L_\odot$. Q_E indicates the number of photons produced above a given energy E_ν. The vertical dashed lines represent the ionization potentials of H^+ (13.6eV), O^{2+} (35.1eV), O^{3+} (54.9eV), and Ne^{4+} (97.1eV). On the left-hand side, the dashed lines represent BPASS models with binary stars, while the solid lines of the same color correspond to BPASS models without binary stars for a single stellar population of the same age. The change in the spectra due to the inclusion of binary stars is represented by the shaded area in between both lines. The insert shows a zoom around the Lyman edge at 912Å where the additional contribution of ionizing photons is visible for ages above 3 Myr and increases with the age of the burst. On the right-hand side, the colors represent different inner temperature of the multicolor blackbody and the different line styles to different percentage of stellar luminosity: 0.001%, 0.01%, and 0.1%.

Stellar source parameters

The stellar evolutionary tracks incorporate the effect of mass transfers between members of binary systems and the stellar initial mass function (IMF) includes a distribution of binaries tuned to reproduce the binary fractions observed in the local Universe. We use a broken power-law IMF with two indices of -1.3 and -2.35 with a change of slope at $M = 0.5\,M_\odot$ which is the default IMF in Eldridge et al. (2017). Although very massive stars ($> 100\,M_\odot$) may exist in local, low-metallicity galaxies (Crowther et al. 2010; Wofford et al. 2021), such objects remain largely unconstrained in models.

Since we do not need to invoke very massive stars to reproduce the infrared lines considered in this study, we choose to use the BPASS default mass cut-off at 100 $M_{\odot}$.

In Figure 5.2, we illustrate the effects of including binary stars by comparing the single-star and binary-stars BPASS (Eldridge et al. 2017) SEDs. A clear difference is visible at ages above 3 – 4 Myrs where models including binary stars produce more ionizing photons. This feature has also been pointed out in Xiao et al. (2018): while single-star and binary-star populations produce fairly similar hydrogen- and helium-ionizing spectra at young ages, the inclusion of binaries produces a shallower drop in ionizing flux than for their single-star counterparts at later ages. While the changes in the ionizing spectra are primarily driven by stellar age effects, the time evolution of the ionizing photons production is metallicity-dependent, as shown in Figure 5.3: low-metallicity models will produce more ionizing photons over a larger period of time.

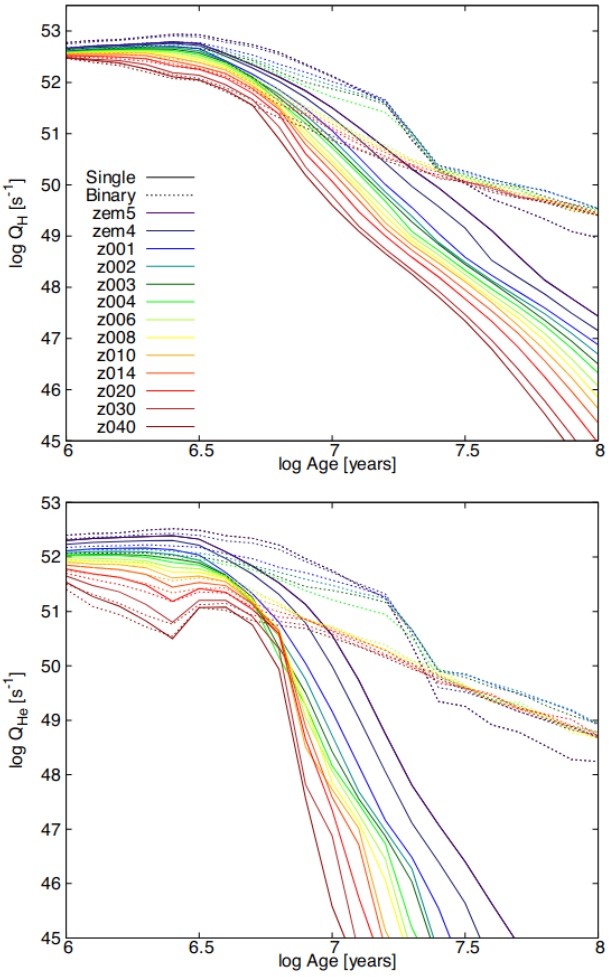

Figure 5.3: Time evolution of the ionizing photon production rate at for H, He at different metallicities, taken from Xiao et al. (2018). The dashed lines indicate the binary-star populations and the solid lines are for the single-star populations. The metallicity increases from Z = 10^5 (zem5) to 0.040 (z040) as the color coding changes from blue to red.

Since we consider only a single burst, our grid spans ages below 10 Myr, after which the Hα emission drops. We note, however, that for the lowest metallicity models, the contribution from binary stars delays this drop in Hα and that ages up to ~30 Myrs (Xiao et al. 2018) should be considered. Considering that our solutions favor ages below 6 Myr and that the addition of a new bin in stellar

age would significantly increase the size of our grid of models, we postpone this to future works. In practice, an older population of stars is present in many of the galaxies in our sample but single bursts of later ages alone would not match the spectral signatures that we study here.

We note that the assumptions we make by fixing a single and instantaneous burst for the star formation history (SFH) only allows us to constrain the age of the population of stars dominating the emission. In practice, assuming a continuous SFH might provide a more realistic spectrum to represent a stellar population in which several bursts of stars might simultaneously contribute to the total emission. A continuous SFH was for example used in Cormier et al. (2019) which modeled the DGS galaxies using a continuous SFH of 10 Myr from Starburst99 (Leitherer et al. 1999). We note, however, that the age until which this continuous SFH was computed was fixed for all galaxies and that only the stellar metallicity was varied. Switching to a continuous SFH instead of a single-burst would result in a change in our ionizing spectrum, as young O/B stars would provide ionizing photons over longer timescales. In particular, because our new code is well suited to perform the combination of several clusters of stars, several improvements could be tested to account for a more realistic stellar population based on the combination of an old stellar population with a younger burst. They will be discussed in Section 11.1.1.

X-ray source parameters

In Figure 5.2 we see that the inclusion of binaries with ages between $1 - 10$ Myr has a most profound effect on lines with ionization potentials between 13.6 up to ~54 eV (see Table 5.1). Indeed, X-ray binaries provide high energy photons which are able to power the emission of several species with very high ionization potentials. For example, X-ray binaries have been invoked in Schaerer et al. (2019) to explain the presence of He II emission which is severely underpredicted with stellar population models. Indeed, it is well-known that these high ionization lines cannot be reproduced by classical photo-ionization models (e.g., Stasińska et al. 2015), even when including effects of binary stars (see Stanway & Eldridge 2019). Several recent studies argue that the addition of an X-ray source is needed to simultaneously reproduce the emission lines arising from different ISM phases (e.g., Simmonds et al. 2021; Olivier et al. 2022; Umeda et al. 2022).

Similarly, we find that in order to produce both [O IV] and [Ne V] emission, we need an additional X-ray source that produces harder ionizing photons than BPASS models alone. The exact nature of such compact objects is unknown (e.g., high-mass X-ray binaries, intermediate mass black hole or AGN). Moreover, the X-ray spectra themselves are widely unconstrained and rely on strong modeling assumptions (Simmonds et al. 2021). Regardless of their nature, X-ray sources are present in the sample of BCDs (e.g., Otí-Floranes et al. 2012; Prestwich et al. 2015; Gross et al. 2021; Ott et al. 2005; Reines et al. 2011; Thuan et al. 2004; Kaaret et al. 2011; Kaaret & Feng 2013; Brorby et al. 2014; Prestwich et al. 2013) we wish to study and need to be included in our models.

We hence choose to use a general prescription to model a compact object surrounded by an accretion disk. To do so, we used a multicolor blackbody spectrum as defined in Mitsuda et al. (1984). This spectrum is defined by an outer temperature fixed at 10^3 K and an inner temperature varying from 10^5 to 10^7 K. The luminosity is set with respect to the stellar luminosity and varies from 0% (no X-ray component needed) to 10% of the stellar cluster luminosity. The right-hand side panel of Figure 5.2 shows the X-ray component when varying the inner temperature of the disk and the relative lumi-

nosity. Although simplistic, our prescription for the X-ray source is more general than that used in previous studies, which considered single blackbody spectra with fixed parameters (e.g., Lebouteiller et al. 2017; Cormier et al. 2019)

5.2.2 Metals and dust

Chemical abundances

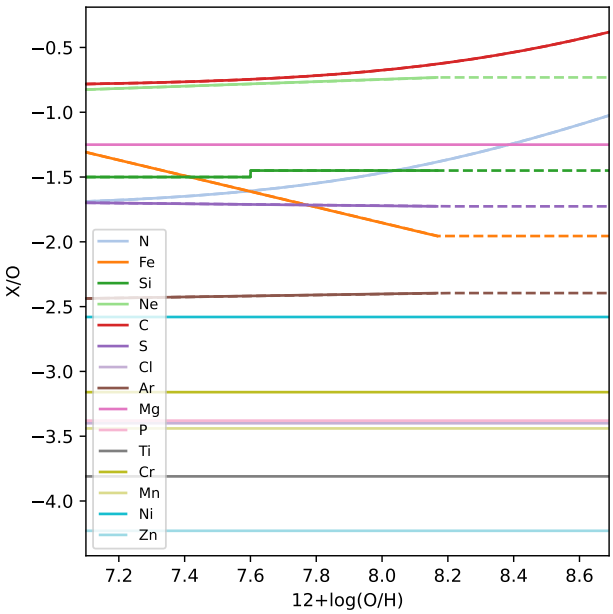

Figure 5.4: Abundance profiles used to compute the grid of `Cloudy` models. Species in bold are the ones for which we use the abundances profile from the ISM table of `Cloudy`, with a simple scaling with metallicity.

Our grid is designed to match the gas-phase abundance patterns of low-metallicity dwarf galaxies that belong to the group of BCD galaxies. The prescriptions used in our models are summarized in Table 5.1. The abundances for nitrogen and carbon are based on Nicholls et al. (2017) who derive analytical curves accounting for primary and secondary production. Their fit is derived from a large sample of stellar measurements in the Milky Way spanning a wide range of metallicities ($6 \leq 12+\log(O/H) \leq 9$). Neon, sulfur, argon, and iron abundance profiles follow the regressions of Izotov et al. (2006), based on a sample of low-metallicity BCDs. To avoid extrapolating at high metallicities, we used flat profiles for metallicities above 8.2 for those four elements. Because of the low statistics of Izotov et al. (2006) for chlorine, we fixed the [Cl/O][1] value to the median of their sample (-3.4). The silicon abundance profile is based on Izotov & Thuan (1999). For all the other metal (i.e., elements heavier than He) we used the values from the Cloudy ISM table, assuming they scale linearly with the oxygen abundance. The profiles used in our grid are given in Figure 5.2. We report the values taken from the table ISM along with the corresponding references in Table 5.2.

[1]Brackets notations indicate ratio relative to solar abundances.

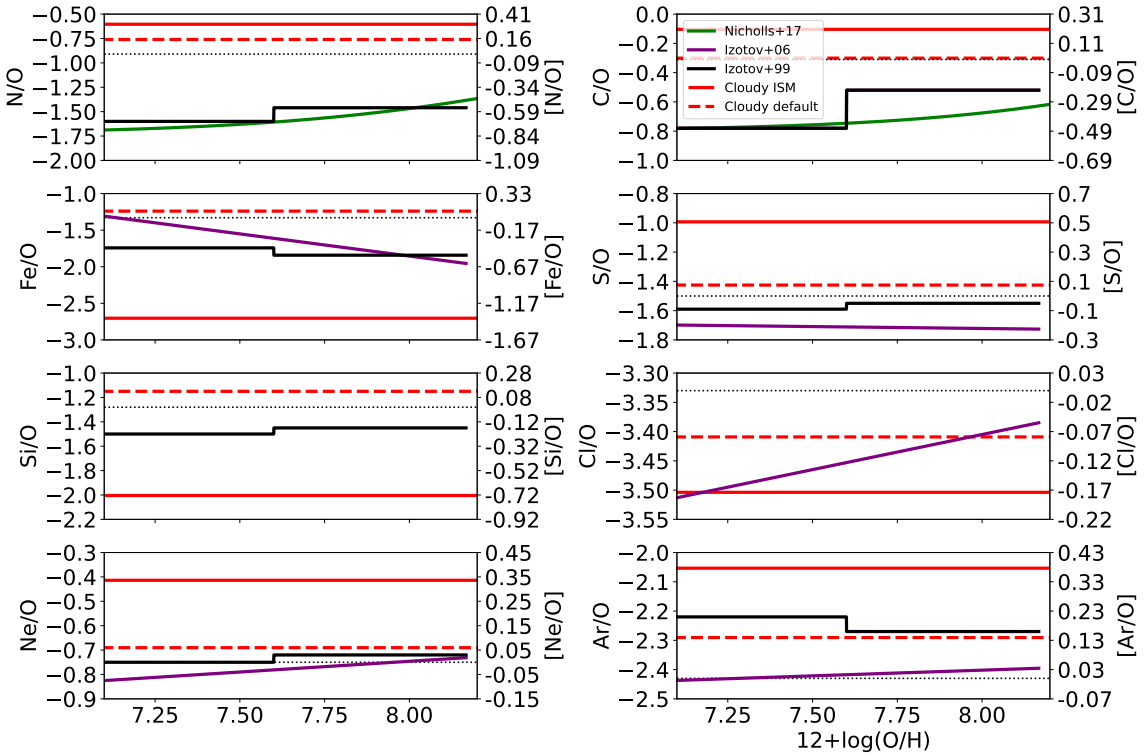

Figure 5.5: Comparison of various abundance profiles, including the `Cloudy` default table and the `Cloudy` ISM table.

Table 5.2: Abundances from Cloudy ISM table.

Element	log(X/H)	12+log(X/H)	Reference
helium	-1.009	10.991	(1),(2)
lithium	-10.268	1.732	(1),(2)
beryllium	-	-	OFF
boron	-10.051	1.949	(1),(2)
carbon	-3.6	8.4	(1),(2)
nitrogen	-4.1	7.9	(3)
oxygen	-3.496	8.504	(3)
fluorine	-7.699	4.301	(4)
neon	-3.91	8.09	(1),(2)
iron	-6.2	5.8	(1),(2)
sodium	-6.5	5.5	(1),(2)
magnesium	-4.9	7.1	(1),(2)
aluminum	-7.1	4.9	(1),(2)
silicon	-5.5	6.5	(1),(2)
phosphorus	-6.796	5.204	(1),(2)
sulfur	-4.489	7.511	(1),(2)
chlorine	-7.0	5.0	(1),(2)
argon	-5.55	6.45	(1),(2)
potassium	-7.959	4.041	(1),(2)

Continued on next page

Table 5.2 – continued from previous page

Element	log(X/H)	12+log(X/H)	Reference
calcium	-9.387	2.613	(1),(2)
titanium	-9.237	2.763	(1),(2)
vanadium	-10.0	2.0	(1),(2)
chromium	-8.0	4.0	(1),(2)
manganese	-7.638	4.362	(1),(2)
cobalt	-8.229	3.771	(5)
nickel	-7.738	4.262	(1),(2)
copper	-8.824	3.176	(1),(2)
zinc	-7.699	4.301	(1),(2)

(1): Cowie & Songaila (1986)
(2): Savage & Sembach (1996)
(3): Meyer et al. (1998)
(4): Snow et al. (2007)
(5): Mullman et al. (1998)

We note that for some of the main elements, the prescription that we use are significantly different from the `Cloudy` default abundances values, and from one gathered in the ISM table (see Figure 5.5). In particular, the profile that we assumed for N/O and C/O are respectively ∼1dex and ∼0.5dex below the default values assumed by `Cloudy` as shown in Figure 5.5. The analytical prescriptions from Nicholls et al. (2017) used for nitrogen and carbon are based on stellar measurements but are compatible with the gas-phase measurements in the BCDs from Izotov & Thuan (1999), as shown in Figure 5.5. This agreement is likely due to the little depletion of carbon and nitrogen in BCDs due to their poor dust content (see discussion in Section 2.2.3).

For other elements, the reference study we used (Izotov & Thuan 1999; Izotov et al. 2006) is empirically calibrated on measurements of the gas-phase metallicity and hence account for possible depletion effects. We also emphasize that the helium abundance is fixed at the value from ISM table, which is not a classical value for galactic studies. Additionally, the abundance for this element is not varying with metallicity in our grid. This limitation can be overcome by adding a corrective scaling based on the helium lines available. This fixing assumes that, as the correction is small, the physical conditions do not change much and the helium line fluxes scale with the helium abundance. This provides a first order correction to match the actual helium abundances by accounting for it as an additional systematic uncertainty.

The metallicity-dependent profiles are important assumptions since they will affect most of the line ratios used in our analysis. In particular, the fact that our code is able to predict metallicities in agreement with optical strong line methods (see Chapter 9), even when no hydrogen recombination line is available, is conditioned to those metallicity-dependent abundance profiles. This assumption is to keep in mind when comparing this grid of models to other samples of objects. In particular, the N/O and C/O profiles might not be adapted to dust-rich galaxies in which depletion on dust grains may be larger. The abundance profiles that were used are specifically tailored to be applied to BCDs galaxies and provided a chemical template adapted to study local low-metallicity galaxies. Comparison to other samples with different physical and chemical conditions should be taken carefully. In particular, considering the unknown physics of primordial environments, possible applications to high-redshift

galaxies should be treated with caution (see Chapter 11.2.1).

Finally, our code could also be applied to mock galaxies resulting from, e.g., post-processed simulations as long as they simulate low-metallicity galaxies and not too high redshifts at which the cosmic microwave background becomes important. Such preliminary tests were performed on outputs from EAGLE simulations in order to constrain the metallicity based on mock observations of [C II]158μm, [N II]122,205μm, [O III]88μm, [O I]145μm, and [N III]57μm (Ramos Padilla, private communication). Because the list of lines does not include any hydrogen recombination line, the metallicity constraints derived from our code are based solely on the abundance profiles used for N/O and C/O. We found that are code was unable to reproduce the wide range of metallicities of the EAGLE simulations, and always found a metallicity near solar. A more careful examination have shown that this result is in fact expected, since the post-processing was performed using constant C/O and N/O ratios. This simple but instructive test strongly emphasizes the need to make sure that the abundance patterns used in the post-processing match the ones which are used in the grid. It also shows that the predictive power of our code, in terms of metallicity, strongly relies on the metallicity-dependent abundance profiles that we assumed.

Dust-to-gas mass ratio

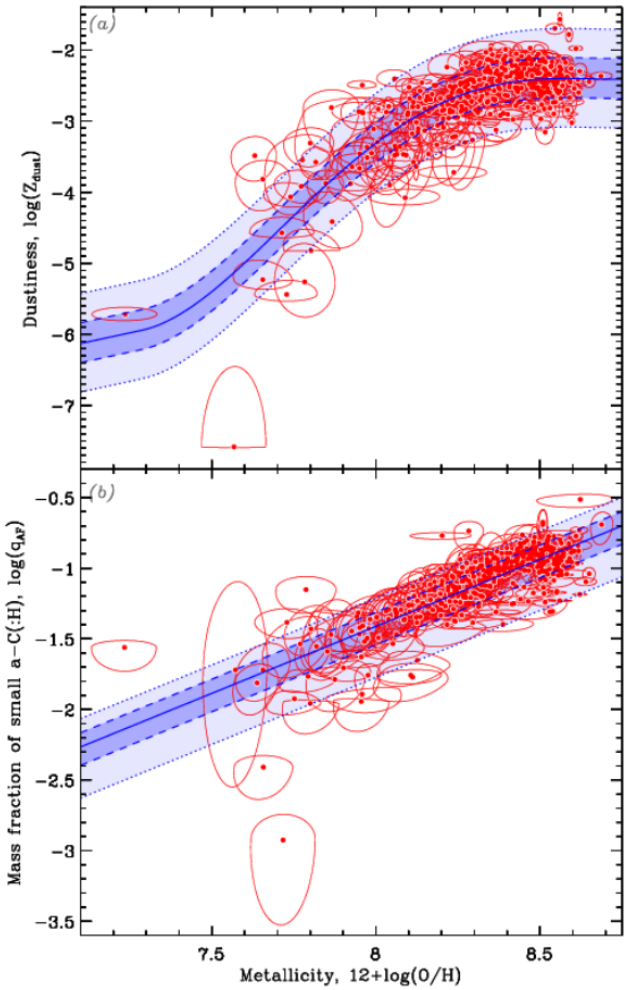

Figure 5.6: Evolution of Z_{dust} and q_{AF} vs. 12+log(O/H). Taken from Galliano et al. (2021).

The evolution of the dust-to-gas mass ratio (or dustiness parameter in Galliano et al. 2021; Z_{dust}) with metallicity follows a non-linear trend, resulting from a change between dominant dust production and destruction processes (see Section 2.2.3). While in the low-metallicity regime, dust production is dominated by condensation in type II SN ejecta and AGB stars, grain growth becomes dominant above a critical metallicity of 12+log(O/H)=8, yielding much higher Z_{dust} (e.g., Rémy-Ruyer et al. 2014; Galliano et al. 2021). Although our grid is not designed to specifically study the continuum emission and dust spectral features, using realistic assumptions regarding the dust content in our modelled galaxies is important as photoelectric effect heating on dust grains is an important heating mechanism in PDRs. Dust abundances also have a strong impact on the formation of H_2 and the shielding of molecules from UV photons.

In the SFGX grid, the dust-to-gas mass ratio and abundance of PAH are computed following the metallicity-dependent prescription of Galliano et al. (2021) based on Bayesian dust SED fits of 798 galaxies (including our sample) using the code `HerBIE` (Galliano 2018) with the `THEMIS` grain properties (Jones et al. 2017). We report here the equations 8 and 9 from Galliano et al. (2021) in which Z_{dust} follows a 4^{th} degree polynomial relation at metallicity above 12+log(O/H) = 7.3 and scales linearly with metallicity below this threshold:

$$\begin{cases} \log Z_{\rm dust} = 11471.808 - 5669.5959x + 1045.9713x^2 - 85.434332x^3 + 2.6078774x^4, \text{ for } x \geq 7.3 \\ \log Z_{\rm dust} = 13.230 + x, \text{ for } x < 7.3 \end{cases} \tag{5.1}$$

where x=12+log(O/H). To account for the intrinsic scatter of the relation, we used two additional dust-to-gas mass ratio points per metallicity bins corresponding to the upper and lower limits of the envelope encompassing 95% of the sample, which correspond to adding [0.68, +0.70] to the fit of log Z_{dust}. The fit and 95% envelop from Galliano et al. (2021) are represented in the upper panel from Figure 5.6.

PAH

Our code includes predictions for PAH emission bands which may be used as additional constraints in future works. In any case, including PAHs is crucial in order to consider the right abundance of free electrons in the ISM with potentially important impact on the population of, e.g., carbon species (e.g., C^+, C^0). Galliano et al. (2021) also provide an analytical prediction for the mass fraction of aromatic features emitting grains ($q_{\rm AF}$), assuming those features are carried by small a-C(:H) grains (see bottom panel from Figure 5.6). We adopt a single value of $q_{\rm AF}$ per metallicity bin, corresponding to the analytical fit from Galliano et al. (2021) which expresses $q_{\rm AF}$, a 1^{st} degree polynomial as a function of 12 + log(O/H):

$$\log q_{\rm AF} = -9.001 + 0.9486x \tag{5.2}$$

In practice, we assume in this study that the aromatic features are carried by PAHs instead a-C(:H) grains and estimate the mass fraction of PAHs ($q_{\rm PAH}$= $M_{\rm PAH}/M_{\rm dust}$) as $q_{\rm PAH} \sim q_{\rm AF}$ / 2.2 (Galliano et al. 2021). Hence, Equation 5.3 becomes:

$$\log q_{\mathrm{PAH}} = -9.343 + 0.9486x \tag{5.3}$$

The abundance of PAHs is known to strongly vary under different physical conditions: they are especially sensitive to metallicity effects and to the strength of the interstellar radiation field (ISRF). Galliano et al. (2021) find that q_{AF} is primarily driven by metallicity effects, while correlations with ISRF indicators are weaker.

We note that this prescription sets the integrated PAH abundance in our `Cloudy` models while the abundance profile follows the default one in `Cloudy`, i.e., scaling as H^0/H. This results in a profile where PAHs are mostly present in the PDR but completely destroyed in the H II regions and nearly absent of the molecular zone. This assumption is consistent with observational studies showing that PAHs are absent in the ionized region (e.g., Relaño et al. 2018; Chastenet et al. 2019), however PAHs can also be present in the molecular phase (e.g., Chastenet et al. 2019). Considering that all the PAHs reside in the PDRs might in turn result in a slight overestimation of the PDR heating by photoelectric effect.

5.2.3 Geometry of H II regions

Gas density radial profile

As most photoionization and photodissociation codes (see Section 3.1.2), `Cloudy` models are stationary and do not account for any dynamical effects due to, for example, expanding H II regions. Thus, the density or pressure profile has to be specified manually. Several density/pressure profiles could be adopted to describe the different phases of the ISM. In our case, the problem is made more complex because we need to find a density law which is representative of the three phases we are considering: H II region, PDR, and molecular zone. Simple density profiles (e.g., constant density used in BOND Asari et al. 2016) are most often used in H II regions studies, as the variations in density yields only minor variations for the lines emitted in the H II regions. However, this choice will have a much stronger effect on the lines emitted near the ionization front and in the PDR/molecular gas.

To consistently describe the density profile throughout the H II region, the PDR, and the molecular gas, we adopted the same density law as in Cormier et al. (2019) in which n_{H} is nearly constant in the H II region and scales with the total hydrogen column density above 10^{21} cm^{-2}. This law provides a simple first order prescription of a smoothly varying density, which can describe both the density profile expected in the PDR of confined, dynamically-expanding H II regions (Hosokawa & Inutsuka 2005, 2006) and in the interior of turbulent molecular clouds (Wolfire et al. 2010). As shown in Figure 5.7, the adopted hydrogen density provides a much smoother variation around the ionization front than assuming a constant pressure profile.

We note that although this prescription is physically-motivated, it remains somewhat arbitrary. Several refinements are possible that may yield a more realistic density distribution. Among the possible improvements, accounting for dynamical effects such as advection due to photo-evaporation of PDR fronts (see Section 1.2.1) would result in different profiles. For example, using the hydrodynamical PDR code *Hydra*, Bron et al. (2018) find that the photo-evaporation processes result in a clear

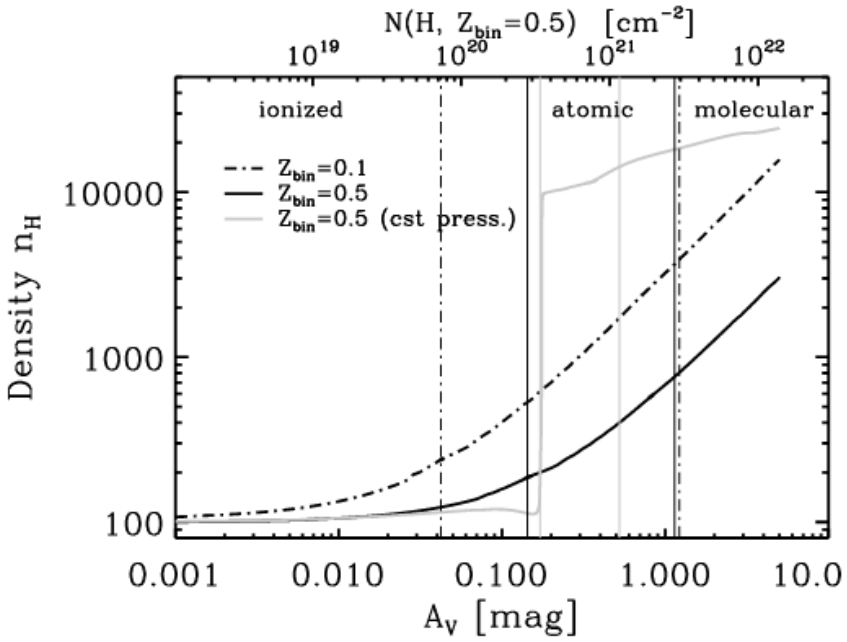

Figure 5.7: Density profile from Cormier et al. (2019) from Cormier et al. (2019). Hydrogen density profile in the default grid of models as a function of A_V, shown for different bins of gas-phase metallicity $Z_{\rm bin}$ = 0.5 (solid line) and $Z_{\rm bin}$ = 0.1 (dot-dash line). They also show the density profile for the constant pressure model (gray solid line). Model parameters are set to $n_{\rm H}$ = 10^2 cm^3 and log U = -2. The phase transitions from H^+-dominated to H I-dominated and from H^0-dominated to H_2-dominated are indicated with vertical lines. We note that $A_V \sim 1$ corresponds to a hydrogen column density that is 5 times larger at $Z_{\rm bin}$ than at $Z_{\rm bin}$ = 0.5.

difference of pressures between the H II region on the one hand and compressed atomic and molecular layer on the other hand. This pressure profile differs from the case of dynamically expanding confined H II regions (Hosokawa & Inutsuka 2005, 2006), in which the compressed PDRs are found to be at pressure equilibrium with the H II regions. Subsequently, a more realistic density profile could be achieved by considering a two-pressure model with a smooth variation in between. A similar density profile was for example adopted in Péquignot (2008) and Lebouteiller et al. (2017).

Finally, we emphasize that the assumed density/pressure law used in our models is a strong limitation of our approach. Ideally, one would rather like to parameterize the density law with random variables that can be inferred in a statistical way. For example, the dense molecular medium in a self-gravitating turbulent cloud is well described by a log normal distribution with a power-law tail at high densities (e.g., Burkhart 2018; Burkhart & Mocz 2019). While this approach is difficult to generalize for a multiphase-study of the ISM, numerous efforts are being pursued to use probability distribution functions (PDF) to describe more accurately the gas distribution within molecular clouds (e.g., Glover et al. 2010; Bisbas et al. 2019; Liu et al. 2021), PDRs (Bisbas et al. 2019) and ionized gas (e.g., Richardson et al. 2014, 2016). Possible modeling improvements such as tabulating the pressure law vs. optical depth at 912Åor combining PDF-based approaches with our multi-component models will be discussed in further details in Section 11.1.1.

Ionization parameter

Each of our 1D model consists of a spherical shell of gas around a central ionizing source, whose inner radius is set to match the input ionization parameter which is defined as:

$$U_{\rm in} = \frac{Q(\mathrm{H}^0)}{4\pi n c R_{\rm in}^2} \tag{5.4}$$

where n is the hydrogen density at the inner radius, $U_{\rm in}$ is the input ionization parameter, c is the speed of light, and $Q(\mathrm{H}^0)$ is the total number ionizing photons produced by the central source defined as follows:

$$Q(\mathrm{H}^0) = \int_{\mathrm{E}_\nu=13.6\mathrm{eV}}^{\infty} \frac{L_*(\nu) + L_{\rm X}(\nu)}{\mathrm{h}\nu} \mathrm{d}\nu \tag{5.5}$$

where $L_*(\nu)$ is the stellar luminosity and $L_X(\nu)$ the luminosity of the X-ray source emitted with a given frequency ν. In Table 5.3 we report the tabulated $Q(\mathrm{H}^0)$ which were used to compute the inner radius used in Cloudy.

Table 5.3: Tabulated $Q(\mathrm{H}^0)$ for BPASS SED at Z=$Z_\odot$ and with $L_{\rm bol} = 10^9 L_\odot$.The values for $L_{\rm bol} = 10^7 L_\odot$ are scaled by a factor 1/100.

Age (Myrs)	log $Q(\mathrm{H}^0)(\mathrm{s}^{-1})$
1	52.61
2	52.47
3	52.32
4	52.27
5	52.14
6	52.03
7	51.96
8	51.89
9	51.82
10	51.76

Table 5.4: Tabulated $Q(\mathrm{H}^0)$ for the multicolor blackbodies. The X-ray-to-stellar luminosity ratios are calculated for a stellar source with $L_{\rm bol} = 10^9 L_\odot$. The values for $L_{\rm bol} = 10^7 L_\odot$ are scaled by a factor 1/100.

X-ray-to-stellar luminosity (%)	log T (K)	log $Q(\mathrm{H}^0)(\mathrm{s}^{-1})$
0.005	5.0	50.45
0.005	6.0	50.04
0.005	7.0	48.91
0.01	5.0	50.76
0.01	6.0	50.35
0.01	7.0	49.21
0.1	5.0	51.76
0.1	6.0	51.35
Continued on next page		

Table 5.4 – continued from previous page

X-ray-to-stellar luminosity (%)	log T (K)	$Q(H^0)(s^{-1})$
0.1	7.0	50.21

We note that this definition of U_{in} only constrains the ionization parameter at the inner irradiated edge and that the volume-averaged ionization parameter could be significantly different depending on the model parameters. We further explore the impact of changing the definition of the ionization parameter, and how it impacts the observed trends of U with various galactic parameters in Section 8.2.3.

Filled sphere vs. thin shell

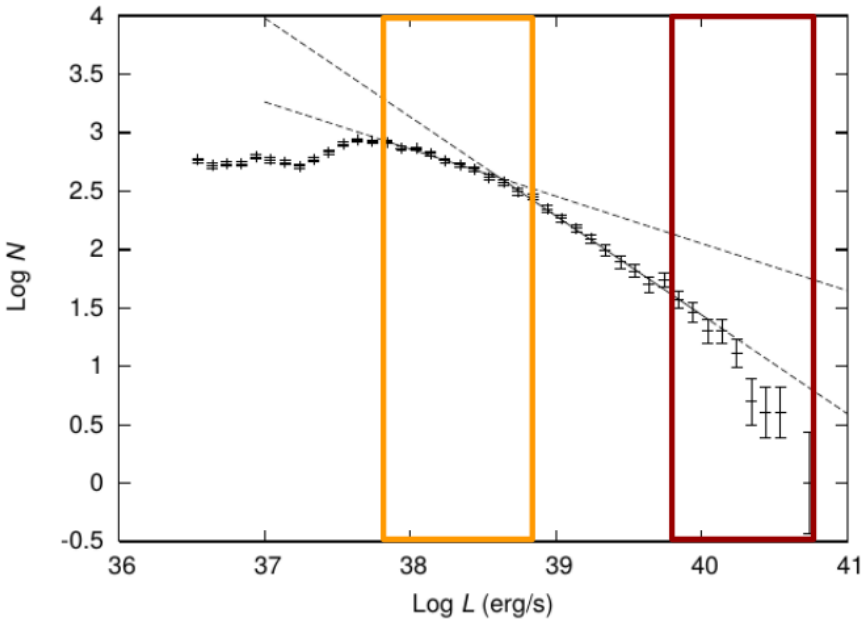

Figure 5.8: Hα luminosity function from Bradley et al. (2006) described by a broken power-law with a break at 38.6±0.1 dex [erg.s^{-1}]. The orange and red square respectively represent our L_{bol}=10^7 $L_\odot$ case (37.8 $\leq$ log Hα (erg s^{-1}) $\leq$ 38.7) and L_{bol}=10^9 $L_\odot$ case (39.8 $\leq$ log Hα (erg s^{-1}) $\leq$ 40.7).

Since the inner radius is set automatically to match a given pair of input ionization parameter and gas density, the only possibility to change the geometry of the region is to change the incident luminosity. At fixed U_{in}, a large cluster luminosity will result in a more shell-like geometry, while a lower cluster luminosity will result in a more filled-sphere geometry in which the gas lies closer to stars. Stasińska et al. (2015) showed that such geometrical effects can affect the low-ionization to high-ionization line ratios (e.g., [O I]/[O III]) as the emission from outer regions is boosted in a compact configuration. Although this geometrical effect is only secondary for most lines arising from the H II regions, it has a much stronger effect on lines emitted near the ionization front and in the PDR/molecular zone because the integrated volume of the zone in which they emit is much more sensitive to geometrical effects.

Because of those geometrical effects, considering the sum of, for example, 100 H II regions with

a luminosity of $10^7\,L_\odot$ is not equivalent to computing a model with a luminosity of $10^9\,L_\odot$. The bolometric luminosity of $10^9\,L_\odot$ chosen by Cormier et al. (2019) corresponds to the case of a galaxy dominated by one single to a few giant H II regions. For the range of ionization parameters and stellar ages covered in our grid, this corresponds to a "thick shell" geometry with log Hα between 39.8 and 40.7 (in erg s^{-1}). Such super-giant H II region with Hα luminosities above 10^{39} erg s^{-1} are numerous among BCD and starburst galaxies, but not always present (Youngblood & Hunter 1999). As shown in Figure 5.8, they populate the upper end of the observed Hα luminosity function derived from resolved galaxies (Bradley et al. 2006) and are likely to be optically thin (Pellegrini et al. 2012). We also consider a lower luminosity $L_{\rm bol}=10^7\,L_\odot$ ($37.8 \leq \log {\rm H}\alpha\,({\rm erg\,s^{-1}}) \leq 38.7$) which corresponds to a typical H II region, similar to those that dominate the Hα luminosity function just before the cut-off value ($\log {\rm H}\alpha$=38.6 erg s^{-1}) from Bradley et al. (2006). This case mimics a bursty star formation where a few hundred to a thousand compact clusters are responsible for the total emission. The photon production rates of such regions ($49.6\,{\rm s^{-1}} \leq \log {\rm N}_c^* \leq 50.6\,{\rm s^{-1}}$) are also compatible with the most luminous ultra-compact H II regions in de la Fuente et al. (2020).

Physical depth of the illuminated clouds

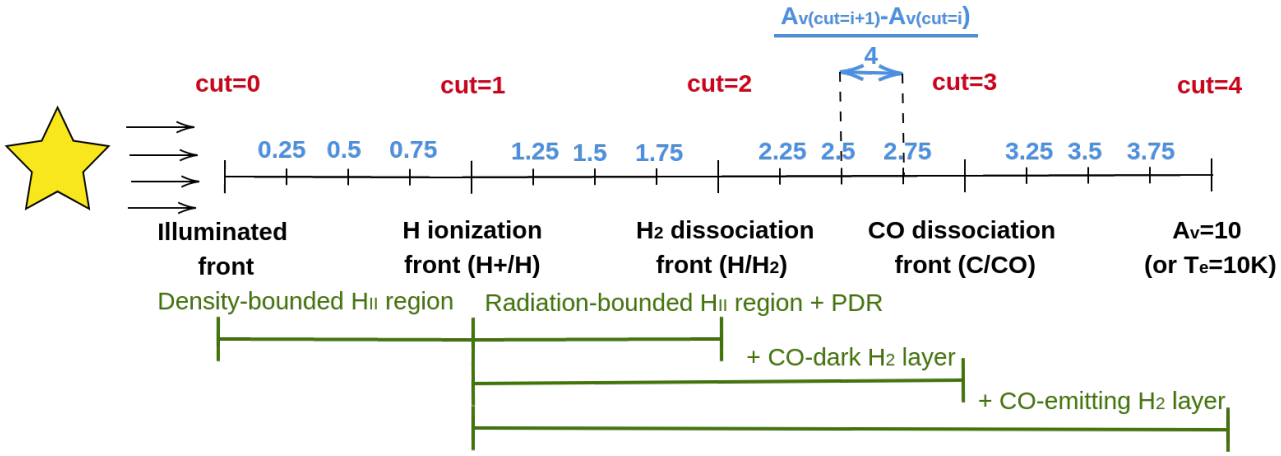

Figure 5.9: Schematic view of the 17 cuts used to create sub-models.

One major change as compared to Cormier et al. (2019) is that luminosities are compiled in a "cumulative" fashion, meaning that we can access the intrinsic luminosity of each line at any given depth in the `Cloudy` models instead of considering only the total resulting luminosity, which obviously depends on the stopping criterion. This is a crucial step to study the escape fraction of ionizing photons, as this parameter is sensitive to the stopping depth of the model, especially near the ionization front.

Defining a "cut" parameter to control the depth is not straightforward. In our case, the choice was motivated by the need to define a parameter allowing to compare models with different metallicities. In addition, one might want to ensure that the physical quantities of interest are sufficiently well sampled with the adopted definition. The adopted "cut" definition is here motivated by the need to sample, in particular, the escape fraction of ionizing photons and will be further detailed in Section 9.2.4.

As illustrated in Figure 5.9, each initial `Cloudy` model is used to create 17 sub-models stopping at different A_V controlled by the "cut" parameter. The original model is successively cut at the inner radius (cut=0), ionization front (cut=1), H_2 dissociation front (cut=2), CO dissociation front (cut=3), and outer radius (cut=4). To sample the different phases (H II region, PDR, CO-dark H_2 region, and CO-emitting H_2 region) defined by those cuts, three additional cuts are added between each integer i (cut=i+0.25, i+0.5, i+0.75), equally spaced in A_V between cut=i and cut=i+1. We stress that stopping the model at a given cut and truncating it a posteriori are not strictly equivalent but our tests have shown that this is a secondary effect.

Our "cut" parameter is analogous to other parameters that were used to describe density-bounded models, such as the Hβ fraction used in Stasińska et al. (2015) and Ramambason et al. (2020) or the zero-age optical depth to LyC photons (τ_λ or τ_c) used in Lebouteiller et al. (2017) and Plat et al. (2019). Although the cut parameter does not have a physical meaning, it allows us to ensure a good sampling of the region near the ionization and provides a simple characterization of density-bounded regions (cut<1) vs. ionization-bounded models (cut>1). Additionally, the definition remains identical regardless of the metallicity (as opposed to A_V).

Since each Cloudy model is cut a posteriori into sub-models, the stopping criterion is not crucial in this study. However, to incorporate the emission from different phases, models should be deep enough to include the neutral and molecular zone for each set of parameters. This transition is metallicity-dependent and requires reaching deep enough A_V for the lowest metallicity models. Most of our Cloudy models are computed until they reach a maximum A_V=10. Only the densest models cannot reach A_V=10 because their electronic temperature drops below 10 K before reaching this optical depth.

5.3 Secondary parameters

While we define primary parameters as the ones used to compute a single `Cloudy` model, we also extract secondary parameters which can be calculated from the `Cloudy` models for each individual sectors. Because the topological models presented in Section 3.2 rely on the combination of several `Cloudy` models considered as components of a global model, those secondary parameters must be described by analytical combinations. Because of the MCMC sampling used in our code (which will be described in Section 6.2), those secondary parameters are can be easily computed from the inferred set of primary parameters. Those local quantities can be easily integrated on the depth parameter to obtain predictions for global quantities such as: the total masses of a given specie, the total number of photons reaching a given depth, or the escape fractions within a given energy range.

5.3.1 Extensive parameters vs. intensive parameters

We distinguish *extensive* quantities, which scale with the total luminosity and can be expressed as a linear combination using the mixing weights of each sector/cluster (see Section 3.2) or *intensive* if the quantity can be expressed as a ratio of extensive parameters. The value $P_{\rm ext}$ corresponding to the combined model can be expressed as a function of the $P_{\rm i,j}$ values of the multiple sectors arranged around multiple clusters as follows:

$$P_{\rm ext} = \sum_{\rm j=1}^{N_{\rm clusters}} \sum_{\rm i=1}^{N_{\rm sectors,j}} w_{\rm i} w_{\rm j} P_{\rm i,j} \tag{5.6}$$

where $w_{\rm i}$ and $w_{\rm j}$ are the mixing weights and $L_{\rm i,j}$ the predicted luminosity of a given line in the i$^{\rm th}$ sector of the j$^{\rm th}$ star cluster. For an intensive parameter, a corrective scaling factor should be added to take into account the relative contribution of each cluster depending on the inferred total bolometric luminosity. This can be expressed as follows:

$$\begin{cases} P_{\rm int} = \sum_{\rm j=1}^{N_{\rm clusters}} s_j \sum_{\rm i=1}^{N_{\rm sectors,j}} w_{\rm i} w_{\rm j} P_{\rm i,j} \\ s_j = \frac{L_{\rm bol,j}}{\sum_{k=0}^{N_{\rm clusters}} L_{\rm bol,k}} \end{cases} \tag{5.7}$$

where $L_{\rm bol,j}$ is the total bolometric luminosity of the j$^{\rm th}$ cluster.

5.3.2 Important secondary parameters

Integrated luminosities

We save both the *intrinsic* and *emergent* luminosities from `Cloudy` models, which are both *extensive* quantities. The intrinsic spectrum includes all the physics of line formation, including destruction by local background opacities such as grains or photoelectric absorption, but does not include the effects of absorption and scattering of the line from regions away from that where the lines form. It corresponds to the observed spectrum after correction for reddening or line-of-sight extinction.

In the following applications, we use only the *intrinsic* line luminosities to avoid considering the pre-defined dust-geometry used in `Cloudy` to compute the emergent luminosities. The luminosities used to constrain our code should, in principle, be corrected *a priori* for dust attenuation effect. Nevertheless, we also include in our code the possibility to use uncorrected measurements and account for an attenuation law adjusted directly within the Bayesian framework (see Section 7.2).

The luminosities are stored in a cumulative mode, meaning that they can be retrieved at each depth in the cloud. The predictions are available for all the lines which were stored in our `Cloudy models`, regardless of whether they are being used as constraints or not. Hence, our code allows us to provide predictions for both observed and unobserved tracers, provided that they are in the list of 516 lines available in the SFGX grid (see Table A1). In addition to that list, we also save the total ionizing luminosities $L(1-\infty)$, $L(1-1.8)$, $L(1.8-4)$, and $L(4-20)$, and the photons production rates $Q(\rm H^0)$=$Q(1-\infty)$, $Q(1-1.8)$, $Q(1.8-4)$, and $Q(4-20)$, where the ranges correspond to the energy in Ryd. The energy ranges correspond respectively to the hydrogen ionization potential (1 Ryd) and helium first (1.8 Ryd), and second (4 Ryd) ionization potentials. They will be used later to examine the escape fractions (see Section 9.4.1).

SFR estimates

Among the lines stored in our grid, some of them (e.g., Hα, Hβ) are well-known tracers of the SFR and have been extensively studied to calibrate some predictive relations. One interesting feature of our code is also that it allows us to obtain the predicted the *intrinsic* value of line fluxes (e.g., the intrinsic Hα) which can directly be linked to the SFR without additional correction for dust attenuation. Hence, the PDFs obtained for well-known SFR tracers provide a simple way to derive PDFs for the predicted SFR. The SFR is also an *extensive* quantity. We present here a few calibrations taken from Calzetti et al. (2012) which are used in the consistency checks presented in Chapter 8. We emphasize, however, that other calibrations could be used as long as they rely on tracers available in the grid. We convert the total number of ionizing photons $Q(H^0)$ into an SFR estimate using the formula given in, e.g., Calzetti et al. (2012):

$$\mathrm{SFR}(Q(H^0)) = 7.4\,10^{-54}\,Q(H^0) \tag{5.8}$$

where $\mathrm{SFR}(Q(H^0))$ is in $M_\odot\,yr^{-1}$ and $Q(H^0)$ in s^{-1}. This analytical prescription was calculated using a Kroupa IMF with masses between 0.1 and 100 $M_\odot$. Assuming that all the ionizing photons are reprocessed by the gas to produced Hα and a case B recombination coefficient for hydrogen lines (with n_e=100cm^{-3} and T_e=10^{4}K), the latter formula also yields a calibration based on Hα:

$$\mathrm{SFR}(H\alpha) = 5.5 \times 10^{-42}\,L(H\alpha) \tag{5.9}$$

This calibration can be refined to account for an additional contribution from obscured star-forming region that are not seen in Hα. This can be done by using empirical calibrations based on mixed tracers (e.g., Hα and L_{TIR}; Kennicutt et al. 2009). This diagnostic corresponds to a SFR conversion assuming a Kroupa IMF[2] and with coefficients calibrated on the *Spitzer*-SINGS sample:

$$\mathrm{SFR}(H\alpha + \mathrm{TIR}) = 5.5 \times 10^{-42}\,L(H\alpha) + 0.0020\,L_{\mathrm{TIR}} \tag{5.10}$$

As detailed in Calzetti et al. (2012), the use of mixed tracers is especially interesting to account for both dust-obscured and dust-unobscured star formation.

Gas and dust masses

Because gas masses scale, on first order, with the luminosity, we also compute them as *extensive* quantities. We extract the gas masses associated with different phases of the ISM. In particular, we extract the gas masses associated to a given phase of the ISM as traced by the density profile of a given specie. Specifically, we define:

- $M(H^+)$, the total gas mass associated to the ionized reservoir as traced by H^+ profile.

[2]We applied a corrective factor of 0.945 (assuming that all the luminosity comes from stars with masses above 1$M_\odot$ and that all the mass comes from stars with masses below 8$M_\odot$) to our predicted SFR to compare them to measurements which assumed a Kroupa IMF.

- $M(H^0)$, the total gas mass associated to the neutral hydrogen reservoir as traced by H^0 profile.
- $M(H_2)$, the total gas mass associated to the molecular gas reservoir as traced by H_2 profile.
- $M_{\rm dust}$ the total mass of dust.

For the gas masses, the mass is calculated as follows:

$$\begin{cases} M_{\rm H,i} = m_{\rm H,i} \int_{R_{\rm in}}^{R_{\rm out}} n_{\rm H,i}(r) V(r) dV \\ V(r)dV = \frac{4}{3}\pi((r+dr)^3 - r^3) \\ n_{\rm H,i}(r) = x_{\rm H,i}(r) n_H(r) \end{cases} \tag{5.11}$$

where $m_{\rm H,i}$ is the mass of the hydrogen ion or molecule (H^+, H^0, or H_2) used as tracer, $n_{\rm H,i}$ the density of this tracer, $x_{\rm H,i}$ the relative abundance of the tracer with respect to total hydrogen, and $n_{\rm H}$ the density of hydrogen. We note that the relative abundance profiles of H^+, H^0 and H_2 verify $x(H^+)+x(H^0)+x(H_2)=1$.

For the dust mass, we use the dust-to-gas mass ratio as defined in Section 5.2.2. The total dust mass is then defined as:

$$M_{\rm dust} = Z_{\rm dust} \int_{R_{\rm in}}^{R_{\rm out}} n_H(r) V(r) dV \tag{5.12}$$

where $Z_{\rm dust}$ is the dust-to-gas mass ratio and n_H the hydrogen density.

Fraction of M(H_2) in the C^0, C^+, and CO zones

The gas mass definition introduced in Equation 5.11 relies on the use of an abundance profile of hydrogen under different states. This definition is flexible and can easily be adapted to trace, among others, the gas reservoir of ionized, neutral, or molecular gas associated with other elements (e.g., carbon). These observables are extensive quantities which quantify the gas masses associated with specific tracers. An application which uses this definition to quantify the amount of "CO-dark" H_2 gas will be presented in Chapter 10.

We define the gas reservoir associated with a given tracer as follows:

$$\begin{cases} M_{\rm H,i}^{\rm X,j} = m_{\rm H,i} \int_{R_{\rm in}}^{R_{\rm out}} n_{\rm X,j} n_{\rm H,i}(r) V(r) dV \\ V(r)dV = \frac{4}{3}\pi((r+dr)^3 - r^3) \\ n_{\rm H,i}(r) = x_{\rm H,i}(r) n_{\rm H}(r) \\ n_{\rm X,j}(r) = x_{\rm X,j}(r) n_{\rm X}(r) \end{cases} \tag{5.13}$$

where $n_{\rm X,j}$ is the density of the element X in an ionization state (or molecular form) j and $x_{\rm X,j}$ is the relative abundance of this tracer with respect to the total density of the element X. By design, we note that $\sum_j x_{X,j} = 1$. We apply this formula to the carbon element to derive $M(H_2)_{C^+}$, $M(H_2)_{\rm C}$, $M(H_2)_{\rm CO}$ for which $x(C^+)+ x(C^0)+ x({\rm CO}) \approx 1$ in most models.

Fraction of [C II] emission traced by H^0 and H^+

We also define the fraction of emission arising in a given ISM phase for any emission line. This fraction is then an intensive secondary parameter, which is derived from extensive quantities (line luminosities). We use the following equations:

$$\begin{cases} f L_{H,i} = \frac{L_{H,i}}{L} \\ L_{H,i} = \int_{R_{\rm in}}^{R_{\rm out}} x_{H,i}(r) L(r) dr \end{cases} \tag{5.14}$$

where $L_{H,i}$ is the emission associated with the ionized, neutral, or molecular hydrogen phase, $x_{\rm H,i}$ the relative abundance of the tracer (H^+, H^0, or H_2) with respect to total hydrogen, and $L(r)$ the line luminosity emitted by the gas slab between r and $r + dr$. This definition is used in Chapter 9 to define the fraction of [C II]158μm arising from the ionized phase ($f[\text{C II}]_{\rm H^+}$).

We emphasize that the use of *relative* abundance profiles which are used as a weighting functions in Equation 5.11, 5.13, and 5.14 allows us to define the ionized, neutral, and molecular phases continuously. It allows us to probe the total gas mass or emission associated to any given ISM phase without having to choose a somewhat arbitrary definition of the ionization and dissociation fronts. For example, our definition is slightly different from the one used in Cormier et al. (2019) to define $f[\text{C II}]_{\rm H^+}$ in which they define strict cuts at the depth where the considered tracer becomes the dominant one at 95%.

Escape fraction of ionizing photons

The escape fraction of ionizing photons is also an intensive parameter which can be derived from extensive quantities. We can use either emission line luminosities (e.g., Hα, Hβ), the total ionizing luminosity, $L(1-\infty)$, $L(1-1.8)$, $L(1.8-4)$, and $L(4-20)$, or the photon production rates, $Q(H^0) = Q(1-\infty)$, $Q(1-1.8)$, $Q(1.8-4)$, and $Q(4-20)$, where the ranges correspond to energy in Ryd. We refer the reader to Section 9.2.2 for a discussion on those different definitions.

Part III

Tools

Chapter 6

Bayesian statistics and ISM studies

In the previous chapters, we introduced our modeling approach that relies on topological models (see Chapter 4) in which galaxies are modeled as combinations of components drawn from a pre-computed grid of `Cloudy` models, described in Chapter 5. Although multicomponent models present interesting advantages, they also complicate the exploration of their parameter space, which becomes extremely large. Frequentist methods, such as the χ^2 method, are not well adapted to exploit the results from topological models (see Section 3.2.2).

To overcome these limitations, we decided to use Bayesian statistics. Its wide variety of algorithms is well adapted for applications to the physics of the ISM. In Section 6.1, we give a general overview of the existing codes that use Bayesian statistics to study the ISM. In Section 6.2, we specifically focus on sampling methods, such as MCMC algorithms, to explore the large parameter space associated with topological models.

6.1 Bayesian framework

The Bayesian approach is a formalism allowing to deal with various problems that can be expressed in a probabilistic way. While it was first formulated in the 18$^{\text{th}}$ century (Bayes 1763), its applications were severely limited by the difficulty to evaluate analytical formulae involving complex, multidimensional integrals and requiring large computing power. The improvement of numerical techniques such as sampling algorithms has allowed the revival of Bayesian analysis in the 19$^{\text{th}}$ century and its application to a wide variety of problems, including astrophysical ones. This approach is especially well suited to probe the underlying physical laws responsible for producing astrophysical observables, while accounting for uncertainties associated with the observations.

In essence, the Bayesian approach allows an observer to convert some knowledge on the Probability Distribution Function (PDF) of a set of data $\overrightarrow{x}$ into knowledge on the PDF of a set of parameters $\overrightarrow{\theta}$. As illustrated in Figure 6.1, this allows to approximate the *posterior* PDF of the parameters $\overrightarrow{\theta}$ by combining *a priori* knowledge on the parameters through the *prior* PDF; $p(\overrightarrow{\theta})$, the PDF of the observed data (i.e., the *evidence*), and the *likelihood* which links the data and parameters by assuming a given model $\mathcal{M}$ that outputs predictions for the observations as a function of the parameters.

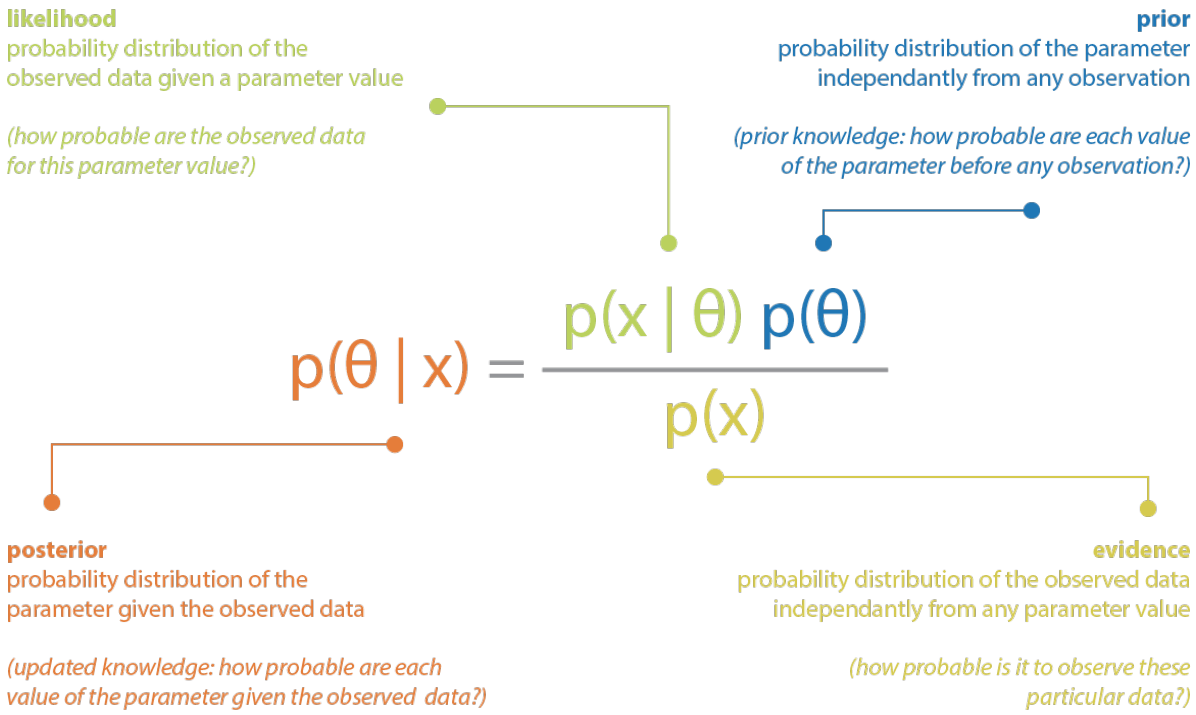

Figure 6.1: Illustration of the Bayes theorem. Credits: https://towardsdatascience.com/bayesian-inference-problem-mcmc-and-variational-inference-25a8aa9bce29.

The main interest of using a Bayesian approach instead of a simple distance minimization routine (such the χ^2 method) is that it outputs the full PDF of a given parameter instead of a single most likely value or a range of values. We note that the likelihood function does not have to be analytical; it can be defined by interpolating values on pre-computed model grids. Such approaches combining a Bayesian framework with large, pre-computed grids of astrophysical models have been used in several studies aiming to constrain physical and chemical parameters of the ISM. The results rely on the exploration of a large, N-dimensional parameter space defined by the grid of models.

Different strategies can be used to explore the parameter space. They all require the definition of a likelihood function. The likelihood provides a metrics to quantify the distance between observations and predictions. A first brute-force approach can be implemented by evaluating this likelihood for each model for each grid point (no sampling, but interpolation is possible). This method has successfully been used in several codes based on multidimensional grids of photoionization models to constrain a few key physical parameters from observed H II region optical emission lines. Among others, recent studies using brute-force approaches have focused on:

- constraining the oxygen abundance (O/H) and ionization parameter with the code `IZI` in Blanc et al. (2014).
- constraining O/H, N/O, and the ionization parameter with the codes `HII-CHI-MISTRY` in Pérez-Montero (2014) and `BOND` in Asari et al. (2016).
- constraining O/H, the effective temperature in H II regions, and the ionization parameter with

`HII-CHI-MISTRY-TEFF` in Pérez-Montero et al. (2019, 2020).

- constraining O/H, the ionization parameter, the pressure, and a parameter controlling the hardness of the radiation field with `NebulaBayes` in Thomas et al. (2018).

Other efforts to include Bayesian-like approaches based on the evaluation of a χ^2-based likelihood function on large grids of models have been developed in state-of-the-art fitting codes such as `CIGALE` (Boquien et al. 2019). Their implementation also relies on a direct evaluation of the likelihood (no sampling) on sub-parts of a large grid of models delimited by recursive algorithms. As opposed to the previously-mentioned studies which have focused on H II region emission lines only, `CIGALE` calculates the energy balance across the full spectrum (from X-ray to IR) and includes predictions for both H II region emission lines and the continuum. It can be used to estimate physical parameters of galaxies such as SFR, stellar mass, dust luminosity, dust attenuation, or the AGN fraction (see, e.g., Villa-Vélez et al. 2021).

One of the main advantages of those brute-force Bayesian approaches, with respect to the χ^2 method, is that they allow constraining the mean/median values predicted for physical parameters and deriving associated error bars. Although they do not include any sampling algorithm to explore the PDF, they can provide joint-PDFs on N-dimensional grids of models which can be marginalized by integrating over poorly constrained parameters, considering them as a nuisance variable. While those methods do not "explore" the parameter space with a step-by-step probabilistic method, as done by sampling algorithms, they have the great advantage of being extremely fast to compute. Additionally, because the likelihood is evaluated on the whole parameter space, this approach gets around possible problems of sampling algorithms which might get stuck around a local solution (see discussion in Section 6.2).

Nevertheless, the evaluation of PDFs on pre-computed grids of model is not a flexible process (fixed size and spacing of parameters) and can be memory-demanding for large grids, so it is well adapted for problems with low dimensionality. Combining pre-computed models from a grid or adding nuisance variables can only be achieved by adding pre-tabulated columns to the grid of models, which can quickly become much too large to be handled with the brute-force approach presented in this section. An alternative option is to use sampling methods, which we now present.

6.2 Interest of sampling methods

To make the most of the flexibility provided by our topological models (see Section 3.2) and explore a parameter space having numerous free parameters, the brute-force approach is not suitable. In particular, we want to be able to add continuous random variables (e.g., scaling associated to systematic errors or mixing weights used to combine components) and sample them simultaneously. This approach avoids having to store large grids of models in memory while allowing on-the-fly interpolation. Although sampling methods require a careful examination of the convergence process, there is much to gain in combining the Bayesian framework describe in this section with sampling algorithms. Specifically, Monte Carlo Markov Chain (MCMC) methods[1], which can efficiently explore

[1]This section was redacted based on several courses, including:
https://brilliant.org/wiki/markov-chains/

multidimensional, complex parameter space with limited memory requirements, are well suited for our purposes.

Sampling methods are powerful tools which allow approximating continuous functions based on numerous discrete samples. Specifically, MCMC methods are a class of algorithms that provide numerical approximations of multidimensional integrals. A Markov process is a dynamically evolving mathematical system which transitions from one state to another by following specific probabilistic rules. A Markov chain is a Markov process with discrete time and discrete state space. The main property of a Markov chain is to be a "memory-less" process, meaning that no matter how the process has arrived at its present state, the possible future states are fixed. This "Markov property" can be mathematically expressed as follows. For any positive integer n and possible states i_0, i_1,..., i_n of the random variable X:

$$p(X_n = i_n | X_{n1} = i_{n1}) = p(X_n = i_n | X_0 = i_0, X_1 = i_1, \ldots, X_{n1} = i_{n1}) \tag{6.1}$$

In other words, the probability distribution of the current state is determined only by the knowledge of the previous state. The general principle of MCMC algorithms is to create a series of correlated samples (or chains) that will converge in distribution to the target distribution. Each chain draws many discrete samples from the continuous target distribution. In practice, MCMC algorithms draw an initial arbitrary set of points sufficiently distant from each other. Each position is then updated at each step of the code in a stochastic way, using various algorithms that look for the places which contribute the most to the integral and assign them a higher probability for the next step. The Markov chains hence proceed by local exploration of the parameter space until all the regions of interest have been uncovered.

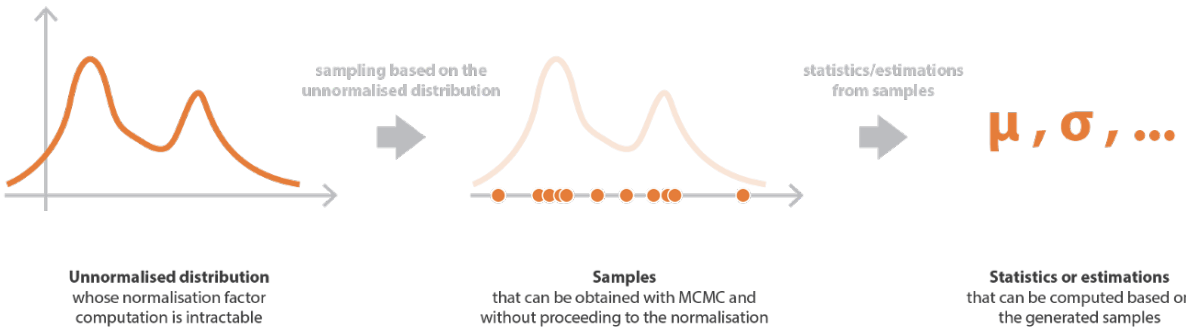

Figure 6.2: Illustration of the MCMC principle. Credits: https://towardsdatascience.com/. The samples obtained at step 2 come from the MCMC algorithm which selects a sequence of draws while exploring the parameter space.

The samples are then used to evaluate integrals over specific variables (as illustrated in Fig. 6.2). This provides a simple way to calculate the expected mean/median value or variance of any random variable, or to approximate the distribution by Kernel Density Estimation (KDE; see Section 7.4.4). The first values drawn by the Markov Chains are highly dependent on the initial conditions, and are

https://towardsdatascience.com/

usually removed as part of the "burn-in" phase, in which the sampler may not have yet reached the regions of high probability. Nevertheless, the Markov property (see Equation 6.1) verified by the chains ensures that the temporal mean when, trajectory becomes infinitely long, is equal to the spatial mean. This property, called *ergodicity*, ensures that, at the limit, the early behavior (including the burn-in phase) of the trajectory becomes negligible and only the long run stationary behavior really matters when computing the temporal mean. A few hand-picked MCMC algorithms, going from simple random walks up to more refined methods, will be presented in the Section 7.3. Although MCMC methods are sometimes less efficient (e.g., in terms of speed) than deterministic methods, they are more widely applicable, especially in the cases of high-dimensional parameter space.

When focusing on physical quantities that can be constrained by models, there is much to gain in combining sampling methods with a Bayesian framework to efficiently explore the parameter space spanned by large grids of models. This approach (Bayesian framework + MCMC applied to model-based likelihood function) was for example implemented in Chevallard et al. (2013) in which they used a Metropolis-Hasting sampler combined with a large library of RT models. Using constraints from broad-band photometry and the Hα/Hβ line ratio, they were able to constrain the dust content and stellar ages in a sample of star-forming galaxies.

This method was further refined to develop the Bayesian SED fitting code `BEAGLE` which uses state-of-the-art models of galaxies to perform spectroscopic and photometric SED fitting, while accounting for numerous instrumental effects (Chevallard & Charlot 2016). The MCMC method used in `BEAGLE` is based on Nested Sampling algorithms incorporated in the `MULTINEST` (Feroz & Hobson 2008; Feroz et al. 2009) code, which is specifically designed to deal with multimodal distributions in N-dimensional parameter spaces. Another approach was used to develop `HerBIE` (Galliano 2018; Galliano et al. 2021), a Bayesian code designed to constrain dust properties based on IR photometric data. The code uses a Gibbs MCMC sampler to explore large grids of dust models using a hierarchical Bayesian framework. The interest of adding a hierarchical structure will be briefly discussed in Section 11.1.2.

Finally, an example closer to our analysis (i.e., exploiting emission lines from photoionization models) can be found in Mingozzi et al. (2020) in which they combine the code `IZI` (Blanc et al. 2014, see Section 6.1) with a simple MCMC sampler (`emcee`; Foreman-Mackey et al. 2013) to explore a grid of photoionization models (`MAPPINGS-IV`; Allen et al. 2008; Dopita et al. 2013) and derive PDFs of the oxygen abundance, the ionization parameter, and the attenuation E(B-V) while accounting for various prior distributions for the ionization parameter based on physically-motivated constraints from the [S III] $\lambda\lambda$9068,9532Å/ [S II] $\lambda\lambda$6716,6731Å emission line ratio. Our approach is very similar to their work but includes an additional layer of complexity by introducing the topological models (see Section 3.2) and a specific sampling algorithm (SMC, see Section 7.3) to deal with multipeaked posteriors. Furthermore, we use combination of `Cloudy` models instead of `MAPPINGS-IV` models used in Mingozzi et al. (2020). This is motivated by the fact that we are looking at the integrated emissions from unresolved galaxies and that we include emission lines arising from various gas phases, while their study focused on spatially-resolved ($\sim$ 1 kpc) maps and considered only the ionized gas. Our Bayesian framework and MCMC sampler are defined directly within our code, `MULTIGRIS`, which we now present.

Chapter 7

MULTIGRIS: A NEW BAYESIAN MULTIDIMENSIONAL GRID SEARCH

In Chapter 6, we reviewed existing codes that exploit the advantages of Bayesian statistics to study the ISM. We also motivated the use of sampling methods in combination with the topological framework presented in Chapter 4.

Those considerations led to the development of our own code, based on MCMC algorithms which performs comparisons between observed emission lines and predictions from topological models, within a Bayesian framework. The current chapter describes the general structure of the code in Section 7.1 and the Bayesian modeling set-up in Section 7.2. The choice of the sampling algorithm is described in Section 7.3 and the processing steps necessary to analyze the results in Section 7.4. Finally, we discuss different ways to quantitatively compare models run using our code with different set-ups in Section 7.5.

7.1 General structure

`MULTIGRIS` (Lebouteiller & Ramambason 2022a,b) is a new open source code designed to combine components drawn from a large grid of models. It is based on the Python module PyMC (see `https://ascl.net/1506.005`; Salvatier et al. 2016) which allows fitting Bayesian models using a variety of numerical methods, including MCMC and variational inference. The `MULTIGRIS` code and its associated documentation are available here: `https://ascl.net/2207.001`. It can be downloaded from gitlab: `https://gitlab.com/multigris`.

An overview of the code structure is provided in Figure 7.1. The Context directory contains the grid of models that will be used. In this thesis, all the applications are based on the mgris_sfgx context associated with the grid described in Chapter 5. However, the code can be used in combination with any grid that is provided as a table with the parameters and observables stored as columns. In particular, our code is provided with the BOND database (Asari et al. 2016) and has also been tested with a grid of LOC (Richardson et al. 2014, 2016) models. Possible applications using those two grids will be described in Section 11.2.1. Other context folders can be created in order to use other grid of models as needed.

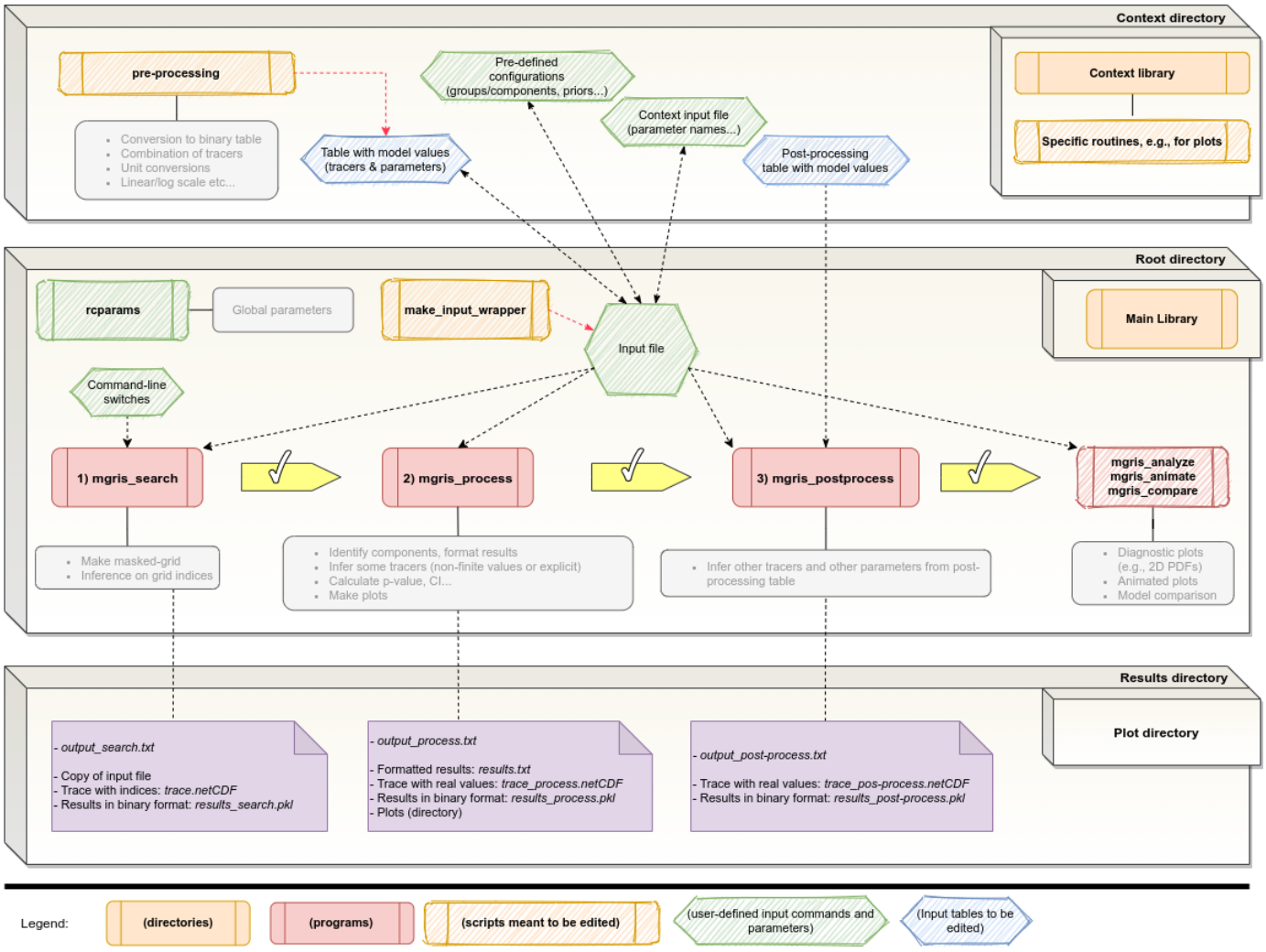

Figure 7.1: Workflow within MULTIGRIS.

The code is divided in three main programs:

- **mgris_search**: performs the inference on the primary parameters (i.e., defining a set of predicted observations) using a list of observed values and their error bars as constraints. This script encapsulates the definition of the Bayesian model (described in Section 7.2) and the MCMC sampling (described in Section 7.3).

- **mgris_process**: distinguishes components a posteriori (see Subsection 7.4.1), outputs predictions for the primary parameters (see Section 5.2) and observables used as constraints, formats the results, and produces some diagnostic plots. The results are stored in a text format (*results.txt*) and in a binary using the pickle format (*results.pkl*). The MCMC draws are stored using a netCDF machine-independent data format (*trace_process.netCDF*)

- **mgris_post_process**: predicts secondary parameters (see Section 5.3) from the context post-processing table. The results are stored in text (*results_pp.txt*) and in binary using the pickle format (*results_pp.pkl*). The draws from the MCMC sampling are stored in a netCDF format (*trace_post_process.netCDF*)

7.2 Model definition

In MULTIGRIS, we define a model $\mathcal{M}$ associated with a given configuration representing a galaxy or a H II region, as defined in Section 4.1. Each configuration is associated with a set of parameters $\overrightarrow{\theta}$. Like in any Bayesian code, we consider a set of data $\overrightarrow{O}$, which are dependent on the parameters of our model. Before any data is observed, both the parameters and the data are *uncertain*. We only know some prior information on the parameters $\overrightarrow{\theta}$, which are expressed as priors $p(\overrightarrow{\theta}|\mathcal{M})$. Knowing that we observed a given set of data $\overrightarrow{O}$ (with the associated uncertainties $\overrightarrow{U}$), we aim to *reduce the uncertainty* on the parameters of our model. This can be done by updating our prior knowledge with the additional information coming from the observed data. Using the Bayes theorem, the posterior PDF of the parameters can be expressed as:

$$p(\overrightarrow{\theta}|\overrightarrow{O},\mathcal{M}) = \frac{p(\overrightarrow{\theta}|\mathcal{M}) \times \mathcal{L}}{p(\overrightarrow{O}|\mathcal{M})} \tag{7.1}$$

with $\mathcal{L}$= $p(\overrightarrow{O}|\overrightarrow{\theta}\,\mathcal{M})$ the likelihood of the data, $p(\overrightarrow{\theta}|\mathcal{M})$ the prior probability, and $p(\overrightarrow{O}|\mathcal{M})$ the evidence or marginal likelihood which integrates all parameter combinations:

$$p(\overrightarrow{O}|\mathcal{M}) = \int_{\theta} p(\overrightarrow{O}|\theta,\mathcal{M})p(\overrightarrow{\theta}|\mathcal{M})d\theta \tag{7.2}$$

As seen in the previous Section 6.2, those integrals can be calculated using MCMC sampling methods. We now describe the core of our code, which consists in defining the likelihood to describe how the data and parameters are linked.

7.2.1 Lines and upper limits

MULTIGRIS is designed to deal with suites of observables (in our case emission lines) which we represent as independent random variables (RV). Each RV is described as a Normal distribution centered on the measured value and with a σ corresponding to the uncertainty of the observation. The likelihood can be expressed as:

$$\mathcal{L} = p(\overrightarrow{O}|\overrightarrow{\theta},\mathcal{M}) = \prod_{i=0}^{N_{\rm obs}} \mathcal{N}(\mu = O_{\rm i}, \sigma^2 = U_{\rm i}^2), \tag{7.3}$$

where $\mathcal{N}(\mu,\sigma^2)$ describes a Gaussian distribution of mean μ and of full width half maximum σ, $N_{\rm obs}$ is the number of emission lines with observed fluxes $O_{\rm i}$ and uncertainties $U_{\rm i}$. Asymmetric error bars are also accounted for by considering two half-Normal distributions. For undetected lines with instrumental upper limits, the Normal distribution is replaced by a half-Normal.

7.2.2 Scaling factor

By default, the model is scaled globally to match the set of observed values, making the scaling factor one of the inferred parameters. This scaling is introduced in order to match the observed *total*

luminosity of the observed galaxies. As described in Section 5.2, our models are computed with a fixed bolometric luminosity of $10^7\,L_\odot$ or $10^9\,L_\odot$ (see discussion on the geometry of H II regions, Section 5.2.3). In essence, we make the assumption that the models are not directly comparable to observations but rather represent *components* of a given topology.

The code determines the most likely configuration based on the line fluxes given as inputs. To define a global scaling factor, we assumed that, on first order, a galaxy can be described as a linear combination of its H II regions (see extended discussion from Section 3.2). This is equivalent to using line ratios instead of absolute line fluxes as constraint. Nevertheless, we partly account for second order effects caused by the diversity of H II region geometries (filled bubble vs. thin shell) by accounting for two possible geometries (associated with the two luminosity values).

We also make the implicit assumption that all sectors contribute to a sizeable amount of emission in at least one tracer. In the case of a single cluster configuration (described in Section 4.2), this results in considering only models for which the stellar clusters are completely surrounded by gas (i.e., the solid angles covered by each sector sum up to 4π). This assumption is motivated by practical reasons to avoid degeneracies between the covering angle and the depth of each sector. Forcing configurations with a covering angle of unity rules out solutions with, for instance, very deep and thin sectors associated to large scaling factors. It also imposes that ionizing photons leakage is always associated with a density-bounded sector (even a very thin or diffuse one) but does not arise from completely empty sightlines.

Eventually, the tracers used to define the global scaling factor may impact the configurations which are favored. By default, we define the scaling factor as a Normal RV applied as a multiplicative factor to all the predicted line fluxes. The mean value is initialized with the distances between predictions and observations, calculated on the full grid and averaged for all the tracers used as constraints. It is also possible to use one tracer (e.g., Hα, Hβ, L_{TIR}) or a combination of tracers (e.g., lines emitted in H II regions only) to estimate the prior on the scaling factor.

7.2.3 Extinction law

As described in Section 5.3.2, we preferentially use the predicted *intrinsic* emission lines to create the input grids, in order to avoid using the predefined dust-geometry from `Cloudy` to correct for attenuation. The intrinsic spectrum includes all the physics of line formation, including destruction by local background opacities such as grains or photoelectric absorption, but does not include the effects of absorption and scattering of the line from regions away from that where the lines form. It corresponds to the observed spectra after correction for reddening or line-of-sight extinction.

Nevertheless, our code can be used both with reddened or dereddened observations. In the first case, dereddened observations for which a correction has been applied to account for internal and Milky Way extinction can be directly compared to the intrinsic fluxes from Cloudy. In the second case, one possible option is to use `MULTIGRIS` with uncorrected luminosities and apply an extinction law within the code to modify the predicted observation (see below for details).

Both approaches have their own advantages; while corrections of the observations are usually based on continuum fitting, the extinction law for models has to be defined a priori. Nevertheless,

correcting the predicted fluxes directly within MULTIGRIS allows us to consider the parameters controlling the extinction law as additional RVs, which can be inferred in a Bayesian way. This method also allows accounting for the uncertainties associated with the extinction correction, while this is most often not the case when extinction correction is applied directly on observations. Finally, it enables considering different laws for each sector of the topological models, which can be used to model anisotropic dust geometries. Inhomogeneous dust geometries (such as, e.g., the "picket-fence" models; Zackrisson et al. 2013; Gazagnes et al. 2018) might have important effects on the absorption of UV photons and the associated emission lines.

Finally, even when using deredenned observations, one might want to include an additional RV representing the differential extinction between IR and optical lines in order to account for possibly different zones of emission. This can, for example, be used to account for optically thick emission within a cloud. Indeed, IR observations can probe much larger depths within the cloud, while optical lines most often trace external layers only. To first order, this effect can be accounted for by applying an additional extinction correction to a selection of lines (e.g., optical).

The extinction laws in MULTIGRIS are computed by using the dust_extinction Python module. This library provides many extinction laws which can be easily implemented in MULTIGRIS. The default law is computed from Weingartner & Draine (2001b) which assumes a Galactic reddening parameterized by the ratio of total-to-selective extinction R_V, which we fixed at the classical value of 3.1. The choice of the law defines a reddening curve $k(\lambda)$ which is applied to redden the prediction of lines that falls within a given validity wavelength range using:

$$\begin{cases} f_{\text{reddened}}(\lambda) = f_{\text{pred}}(\lambda) \times 10^{-0.4 A_\lambda} \\ A_\lambda = k(\lambda) E(B-V) = \frac{k(\lambda) A_V}{R_V} \end{cases} \tag{7.4}$$

where A_V is the visual extinction in V band and $E(B-V)$ the color excess. We define $E(B-V)$ as a RV that follows a Half-Normal distribution with a standard deviation of 1. The value of $E(B-V)$ can be inferred as a global extinction for the whole model or independently for each sector.

We note that the extinction laws we consider only take into account a dust-screen geometry. To account for more complex dust geometries, one can instead use varying extinction laws for each sector or global attenuation laws that include the effects of having multiple sources extinguished by different columns of dust and the scattering of photons into the observation beam. Attenuation laws can be implemented through the dust_attenuation Python module by defining not one but several RVs corresponding to the parameters controlling the law. To what extent the assumptions made regarding gas and dust geometries used to define attenuation laws are consistent with our topological models remains to be investigated in order to make a self-consistent model.

7.2.4 Priors

By default, we use weakly informative priors for all the parameter as a Normal distribution centered on the mean of the grid and with a full width half maximum σ equal to the entire parameter range. These priors slightly reduce the probability to be located at the edges of the parameter range. In other words, we design the grid such that the extreme parameter values are considered to be somewhat

less likely. It is hence important that our grid spans a large enough parameter range. Other priors can be added on the primary parameters (see Section 5.2). The code allows us to include physically motivated priors by linking some parameters by logical equations. The supported operations include:

- Comparing the parameter values between sectors. This can be done using the classical operations equal (=), greater than (>), or low than (<). The corresponding syntax is:

$$\text{parameter\_name (i) operation(j)} \tag{7.5}$$

  with i and j being the number corresponding to the two sectors and operation being either =, < or >.

- Setting a value for a parameter:

$$\text{parameter\_name (i) } = \text{ value} \tag{7.6}$$

- Setting a normal distribution for a parameter:

$$\text{parameter\_name (i) } = \text{ value sdev} \tag{7.7}$$

- Using a literal expression binding a parameter to others:

$$\text{parameter\_name (i) } == \text{ equation} \tag{7.8}$$

- Setting all parameters to the same value for all sectors. In this case, the first sector will be used as reference. This is the command used to set all the parameters related to the incident radiation sources in the single-cluster, multisector case presented in Section 4.2.

$$\text{parameter\_name } (*) \ = \tag{7.9}$$

More details can be found about the syntax in the ReadMe file file associated with each Context folder.

7.2.5 Systematic uncertainties

Systematic uncertainties are treated as nuisance variables encompassing a set of one or several observable(s), distributed by default as a Normal distribution. They can be used to account for uncertainties such as:

- Instrumental calibration uncertainties: a common scaling factor applied to all the lines observed with the same instrument.

- Elemental abundance uncertainties: a common scaling factor is applied to all the lines corresponding to the same element. We make the assumption that fluxes scale, to first order, with the total elemental abundance, regardless of the ionization stage. This simplification remains valid for small variations of the physical/chemical conditions, as long as they do not strongly affect the chemical reactions or physical conditions associated with line emissions.

- Uncertainties associated with the modeling: a common scaling factor can be applied to mimic the effect of an unaccounted for gas phase (e.g., contamination by DIG, incomplete PDR covering factor).

An example including systematic uncertainties is presented in LR22 in which a scaling of on the helium line fluxes that approximates a scaling of the total helium abundance has been included (see previous discussion in Section 5.2.2).

7.3 MCMC sampling

7.3.1 Choice of the sampler

The sampling methods used by `MULTIGRIS` are encoded in the Python module PyMC which provides a large panel of MCMC algorithms. Most of the algorithms presented here are random walkers that start at one arbitrary position in the parameter space and update it sequentially to explore a N-dimensional space. We also describe the Sequential Monte-Carlo (SMC) method, which simultaneously uses several independent random walkers in a parallel way. The following algorithms were tested to benchmark our code (LR22):

- The Metropolis-Hastings algorithm (e.g., Robert 2015) provides a generic way to construct a Markov chain on a given parameter space that converges in distribution to a "target" distribution. This sampler is one of the most classically used among MCMC algorithms due to its simplicity and versatility. However, this algorithm rarely achieves optimal rates of convergence (Mengersen & Tweedie 1996) and may get into convergence difficulties if improperly calibrated. Its random exploration of the parameter space requires numerous samples, and its implementation can be limited by time requirements in the case of large datasets.

- Adaptive Differential Evolution Metropolis sampling step that uses the past chain values to inform and generate jumps (Vrugt et al. 2008).

- No-U-Turn sampler (NUTS) for continuous variables based on Hamiltonian mechanics (Hoffman & Gelman 2011). This algorithm avoids the random walk behavior and sensitivity to correlated parameters by taking a series of steps informed by first-order gradient information. It is designed to converge much more efficiently than simple methods such as Metropolis-Hasting.

- The Sequential Monte-Carlo (SMC) sampler is a method that progresses by a series of successive annealed sequences from the prior to the posterior. The PyMC implementation is based on Ching & Chen (2007) and Mison et al. (2013). We further describe this algorithm in the next Section 7.3.2.

We refer to LR22 for more details on the benchmark comparing sampler methods. The main result of this comparison is that random walkers and gradient-based algorithms are not performing well in the case of multipeaked solutions, which are often present in the astrophysical applications envisioned with `MULTIGRIS`. Among the four samplers that were tested, the best results were obtained with SMC.

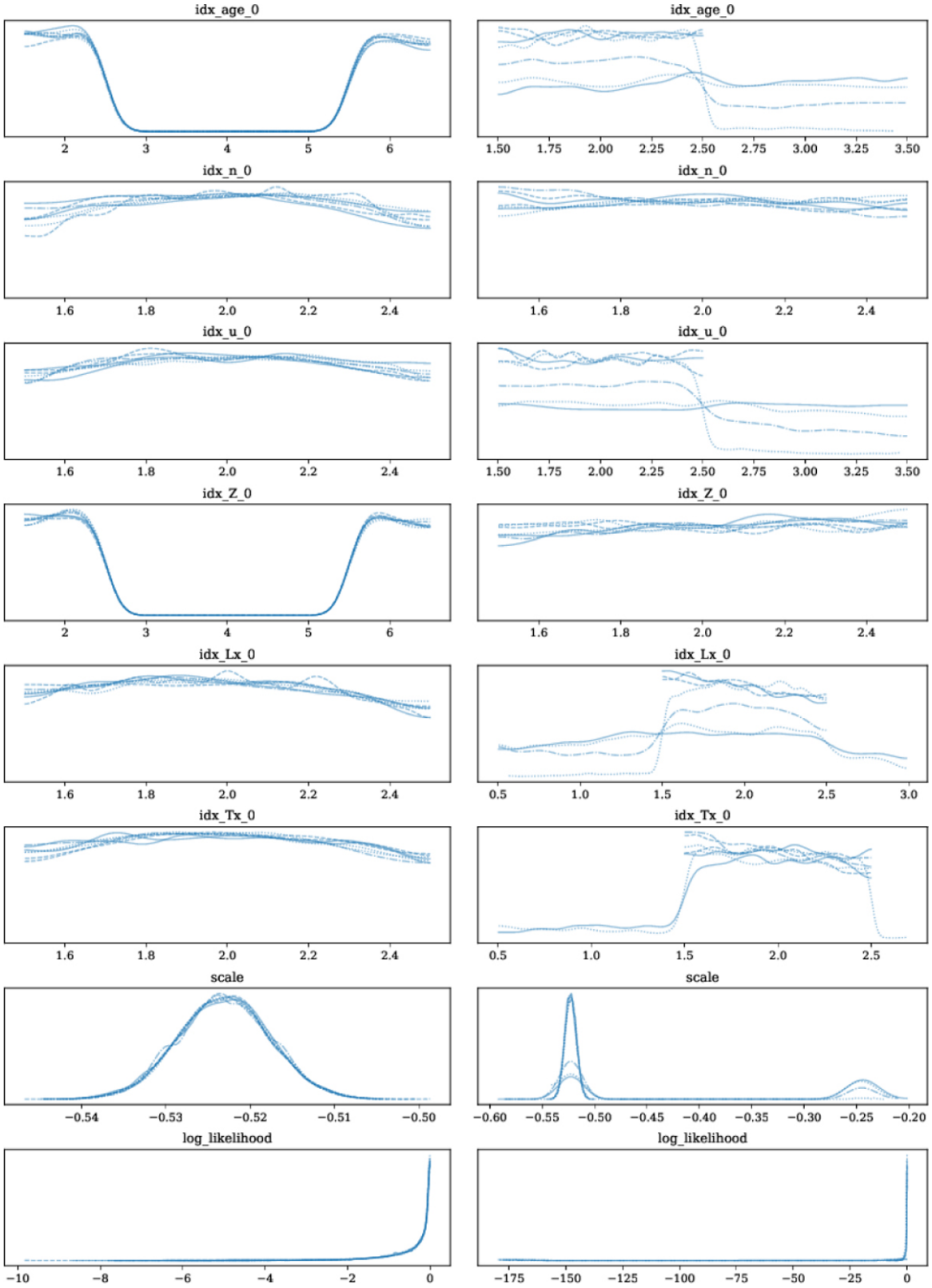

Figure 7.2: Comparison SMC/NUTS samplers from LR22: Benchmark model run with a degeneracy artificially introduced for the age and metallicity, with the SMC step method on the left and NUTS on the right. Each individual line corresponds to a given SMC job, while for NUTS it corresponds to a different chain. Note the different x-axes range for the two columns.

In particular, as shown in Figure 7.2, the SMC sampler is performing much better at recovering multi-peaks introduced by degenerated parameters while the NUTS sampler is likely to get stuck around a local solution. This is explained by the fact that NUTS needs long enough chains to authorize excursions either to or from the high-probability region. In comparison, SMC simultaneously identifies the high-likelihood regions by running in parallel numerous chains starting from random values sampled from the prior distribution. Another advantages of SMC is that it does not have a burn-in phase. We hence selected this sampling method which is adapted to our scientific objectives and further describe this algorithm in the following.

7.3.2 Sequential Monte Carlo sampler (SMC)

SMC is part of the general family of particle filtering techniques. It uses a combination of various techniques, with in particular tempered likelihood and importance sampling. The importance sampling is a variance reduction technique that can be used in the Monte Carlo method. It assesses the "important" parameters that impact more the estimates and sample them more frequently. The notion of tempered likelihood is used in tempering methods, which come from the statistical physics formalism and are designed to find the lowest free energy state of a system of many interacting particles at low temperature.

The posterior is updated iteratively, starting from a "hot" state following the prior distribution (with $T = \infty$ and $\beta = 0$ the inverse temperature index) to a "cold" state, which effectively corresponds to the posterior distribution (β=1). This β index is then used as an exponential index. The tempered posterior is proportional to the product of the tempered likelihood and the prior. For a model $\mathcal{M}$, this is expressed as:

$$p(\overrightarrow{\theta}|\overrightarrow{O}, \mathcal{M})_\beta \propto p(\overrightarrow{O}|\overrightarrow{\theta}, \mathcal{M})^\beta p(\theta). \tag{7.10}$$

SMC first samples the prior ($\beta = 0$), calculates a new β value to match a pre-defined effective sample size (by default half the number of draws), computes importance weights using the tempered likelihood $p(\overrightarrow{O}|\overrightarrow{\theta}, \mathcal{M})^\beta$, computes a new set of samples by resampling according to importance weights, computes the mean/covariance of the proposed distribution, runs relatively short (on the order of ~10) independent Metropolis-Hastings chains from the proposed distribution to explore the tempered posterior.

The importance weights increasingly account for the likelihood as the stage number increases. The stage number continues to increase until the true posterior is sampled ($\beta \geq 1$). The overall process is illustrated in Figure 7.3. Since SMC samples numerous chains exploring the entire parameter space instead of a single chain, it is important to have enough samples to probe the prior parameter space. Furthermore, since each individual chain walks with independent Metropolis-Hastings, it is in principle possible to have only few samples per parameter. However, better results are obtained when using many chains per parameter combination to either obtain an average PDF if the chains have significantly different distributions and to evaluate whether all chains have converged to the same distribution.

While the number of samples may seem appropriate for the considered prior distribution, they

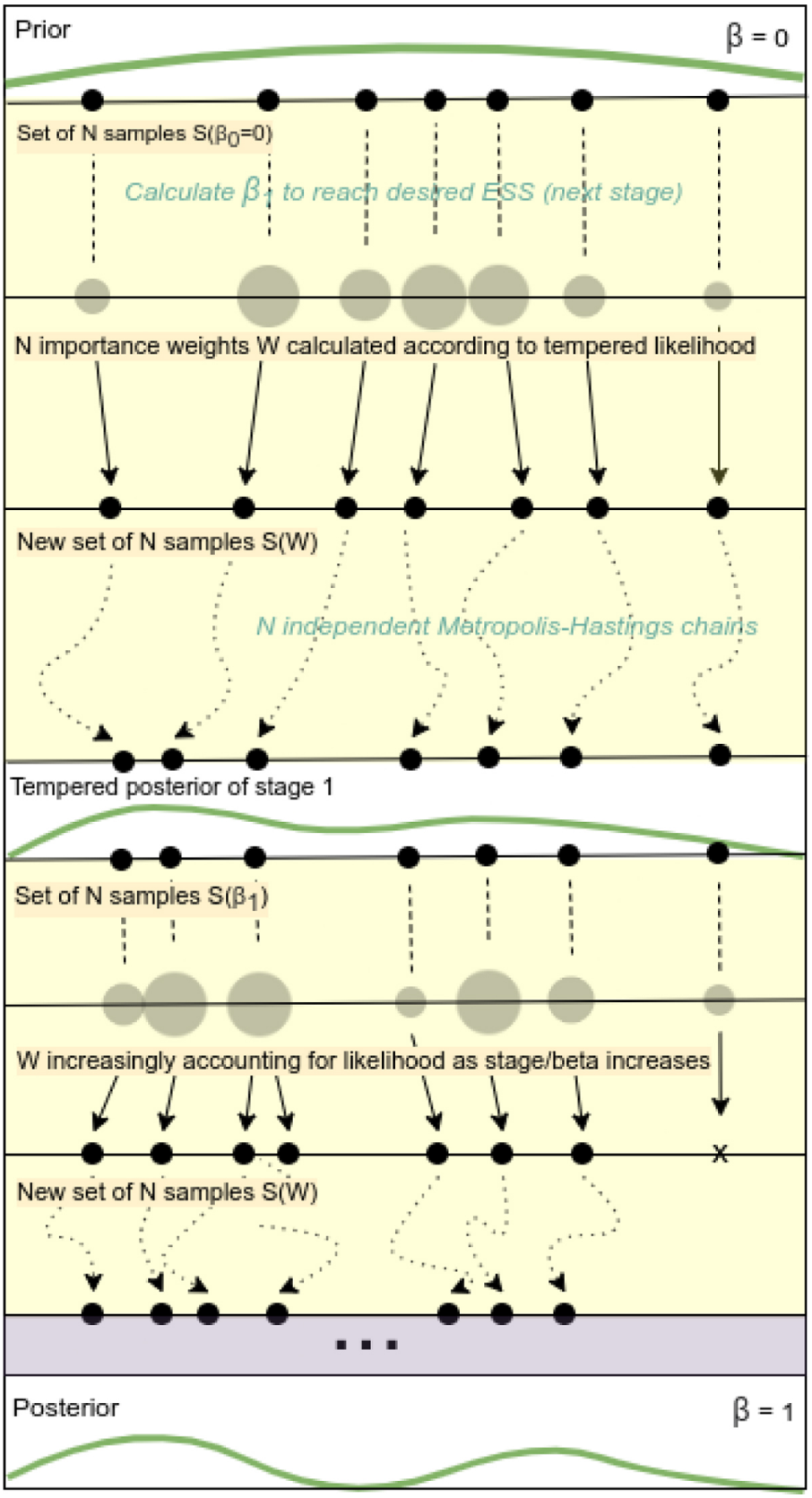

Figure 7.3: Illustration of SMC principle, taken from LR22. From top to bottom, the distribution is sampled from the prior, reweighted and resampled according to the tempered posterior density. Surviving samples seed new Markov chains and walk through a number of Metropolis steps.

also need to be numerous enough to probe the posterior distribution, which may show multiple peaks. If only a few samples are considered, this may lead to significant stochasticity between independent model runs. SMC runs several jobs in parallel and, in the followings, we always show the combined model results. The inference result contains the posterior distribution of the parameters, which are then processed for further analysis.

7.4 Processing the inference results

7.4.1 Component identification

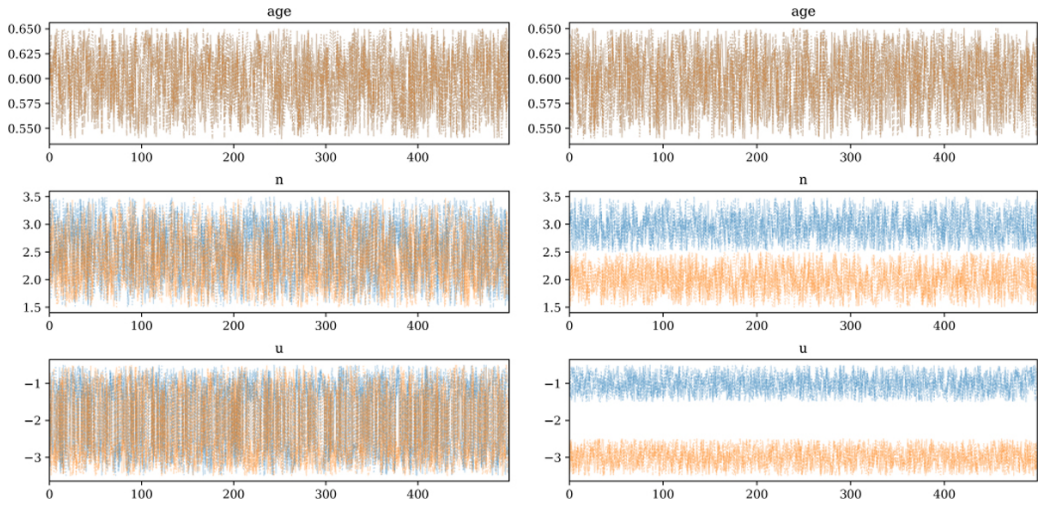

Figure 7.4: Illustration of the component identification step from LR22. Posterior distribution of parameters for a model with two components. Only some parameters have been allowed to have different values for each component (here u and n). The raw posterior distribution is shown on the left, for which the two components may switch during inference. The final result is shown on the right, for which the individual component parameter values have been isolated a posteriori. For visual purposes, only a subset of the parameters are shown

An important post-sampling step is to identify the components. By default, drawn values of primary parameters (used for the inference) are not assigned to any specific component (i.e., components are not explicitly identified or tied to any given parameter or parameter set). This is meant to assert whether inferred parameters have indeed significantly different values between each component. If values are not significantly different between components, the components are thus allowed to switch back and forth while the solution should remain stable (specifically, for the NUTS/Metropolis samplers this means that draws may switch with time, while for the SMC sampler this means that each draw at any given stage may probe any component). If the main objective is to derive probability density functions of secondary parameters (see Section 5.3), there is no need to identify explicitly each component. Nevertheless, if the parameter value for each component is difficult to disentangle, some samples may be wasted to switch from one component to the other and the probability distributions may not be as smooth as possible.

In some cases, it may be important to examine the posterior distribution of primary parameters for each component individually. `MULTIGRIS` performs a processing step to identify and characterize

individual components from the chain, by redistributing components through a minimization of the standard deviation for all parameters for individual chains and components (see Figure 7.4). This is an iterative process, re-distributing components for each draw along the entire chain several times. This re-distribution bears some uncertainties and the individual component properties may still have some level of degeneracy. The global solution (combination of the components) is, however, not affected by the re-distribution of components since all parameters are switched together. Depending on the solution, it may be difficult to isolate components a posteriori, which may indicate a need for more constraints. For this reason, it may be useful to force the identification of components a priori. In MULTIGRIS, components can be sorted using a prior constraint on any given parameter. Sorting with a prior constraint will obviously affect the inference, and the result will be different depending on which parameter is used for sorting.

7.4.2 A posteriori checks

Convergence and agreement between chains

After running a model, one wants to check whether the sampler has converged. This can be done by verifying that the independent chains are in agreement with one another. To do so, several indicators are provided in PyMC:

- the $\hat{R}$ diagnostic (Vehtari et al. 2019) indicates if the chains mix well for a given parameter. Ideally $\hat{R}$ should be below 1.1 for all parameters. In practice, while $\hat{R} < 1.1$ are obtained for all parameters for single-sector models, this criterion is hard to satisfy for models with several components. Nevertheless, they should remain as close as possible to unity, since large $\hat{R}$ indicates that the chains have converged toward distinct probability distributions.
- the Effective Sample Size (ESS) corresponds to the effective number of uncorrelated draws. It indicates whether the chains are long enough or if the number of samples should be increased. The *results.txt* output includes the ESS calculated for the "bulk" of the samples, related to efficiency of mean and median estimates, as well as for the "tail", corresponding to the 5% and 95% quantiles.
- the Markov Chain Standard Error (MCSE) estimates the error introduced by sampling and thus the level of precision of our estimates.
- the auto-correlation of individual parameters, which quantifies the relationship between a variable's current value and its past values. An autocorrelation of +1 represents a perfect positive correlation, while an autocorrelation of −1 represents a perfect negative correlation. One way to check that the sampler has sufficiently explored the entire parameter space and that it is not stuck in the burn-in phase is to ensure that the auto-correlation is close to 0.

Agreement with the observations

One might also want to check how well the model performs at reproducing the observations. Several indicators can be used to quantify the agreement, which are provided for each emission lines in the

result.txt output file:

- the distance Δ corresponding to the difference between the observed and the mean predicted value (in logarithmic scale) and its translation in terms of $n-\sigma$ value (corresponding to $\Delta/\sigma_{\rm obs}$).
- the upper and lower bounds for prediction corresponding to the High Density Interval (HDI) at 94% (i.e., the parameter falls within 94% probability within this interval).
- the p-value informs on how likely it is that the agreement between observation and prediction would have occurred by random chance (i.e., that the null hypothesis is true). The p-value (or 1 minus the p-value, depending on whether the prediction underpredicts or overpredicts the data) should ideally be small (e.g., below 0.05 for a classic significance threshold) but extremely small values (e.g., <0.025) can also arise due to over-fitting or bad fit. In the code outputs, the lines corresponding to p-values <0.025 or >0.975 are hence flagged. On the other hand, p-values which are close to 0.5 may indicate large HDI.

Finally, global values averaged on all the parameters (primary and secondary) are reported at the end of the *results.txt* file:

- a global p-value.
- the percentage of posterior draws that agree with the observation within 1, 2, and 3 σ.

As a general rule, a good model should have a global p-value as close as possible to 1 according to our definition (but below 0.975, corresponding to over-fitted line fluxes) and should also have the largest percentage of posterior draws in agreement with the observations. Although several metrics are provided, the thresholds used to define a model which is "sufficiently good" remain somewhat arbitrary.

Post-processing

Primary parameters (see Section 5.2) are used for inference, they are the ones that are used as variables to construct the input grid. From the posterior distribution of primary parameters, we can therefore obtain the corresponding distribution for any observable or parameter in the grid. The post-processing script uses the PDF of primary parameters in order to deduce new observables that were not used as constraints, or secondary parameters (see Section 5.3) that completely depend on the primary parameters. In our case, any output from the `Cloudy` models can be extracted to be inferred in the post-processing step.

7.4.3 Visualizing PDFs for individual objects

We now present some important diagnostic plots which are produced by the *mgris_process* and *mgris_post_process* routines. The following plots are taken from runs for the galaxy He 2-10, for which we used 22 emission lines and 6 upper limits to constrain a 3-sector configuration. This model corresponds to the best model selected for this galaxy in Ramambason et al. (2022).

Visualizing the parameters

Figure 7.5 shows the 1D posterior PDF of the primary parameters. On the right-hand side, we show the excursions of the sampler around the solution as a function of the number of steps. This representation is useful for visual inspection to identify potential issues in the sampling process. The discrete sampling is then converted into a 1D continuous distribution, shown of the left-hand side. Figure 7.6 shows another representation of the PDF of the primary parameters under the form of a violin plot representing the HDI at 94% level. The left-hand side panel of Figure 7.7 shows a representation of 2D-PDFs for pairs of primary parameters taking the form of a corner plot. This visualization allows to easily identify degeneracies and correlations between parameters.

Visualizing the line flux agreements

In order to visually compare the predicted line fluxes with the observations, we provide several plots shown in Figure 7.8 and 7.9. In Figure 7.8, we show a mosaic for all the lines which are used as constraint. The observed data accounting for noise is simulated to match the values given as detection and error bars and assuming a fixed arbitrary spectral resolution Δv. The red Gaussians are obtained by sampling a posteriori the PDFs derived by MULTIGRIS for each line flux. On the left-hand side of Figure 7.9, we show the relative agreement between the observation and the mean predictions as y=(obs-mod)/obs. The color code allows visualizing how well the detection as well as the upper limits are reproduced.

Another interesting plot is the correlation map shown on the right-hand side of Figure 7.9. The correlations are calculated using the MCMC draws which are sampled simultaneously for all lines, hence it allows the user to see if some lines are tracing the same parameters. For example, on Figure 7.9, one can see that, according to the input grid, most H lines trace the same parameters around the solution (the same applies to He or [S II] lines).

Visualizing the post-processed quantities

In Figures 7.10 and 7.11 we show the 1D-PDFs of the quantities which were derived by the post-processing step described in Section 7.4.2. The first column shows the PDFs of lines which were used as constraints for the inference process, while the third and fifth columns correspond to predictions for unobserved lines and secondary parameters of interest.

7.4.4 Visualizing PDFs for a sample of objects

Representing multidimensional PDFs is not as straightforward as plotting a single value with error bars. Most of the time, the results we present will be marginalized 2D-PDFs for pairs of parameters. By nature, the MCMC draws provide a *discrete* representation of the PDF. Although this discrete sampling makes the approximation of integrals much easier (replaced by discrete sums), the visualization of N-dimensional clouds of points can be difficult. We present in this section tools allowing to represent 2D-PDFs in a simple way. They will be used in the applications described in Chapter 9 and

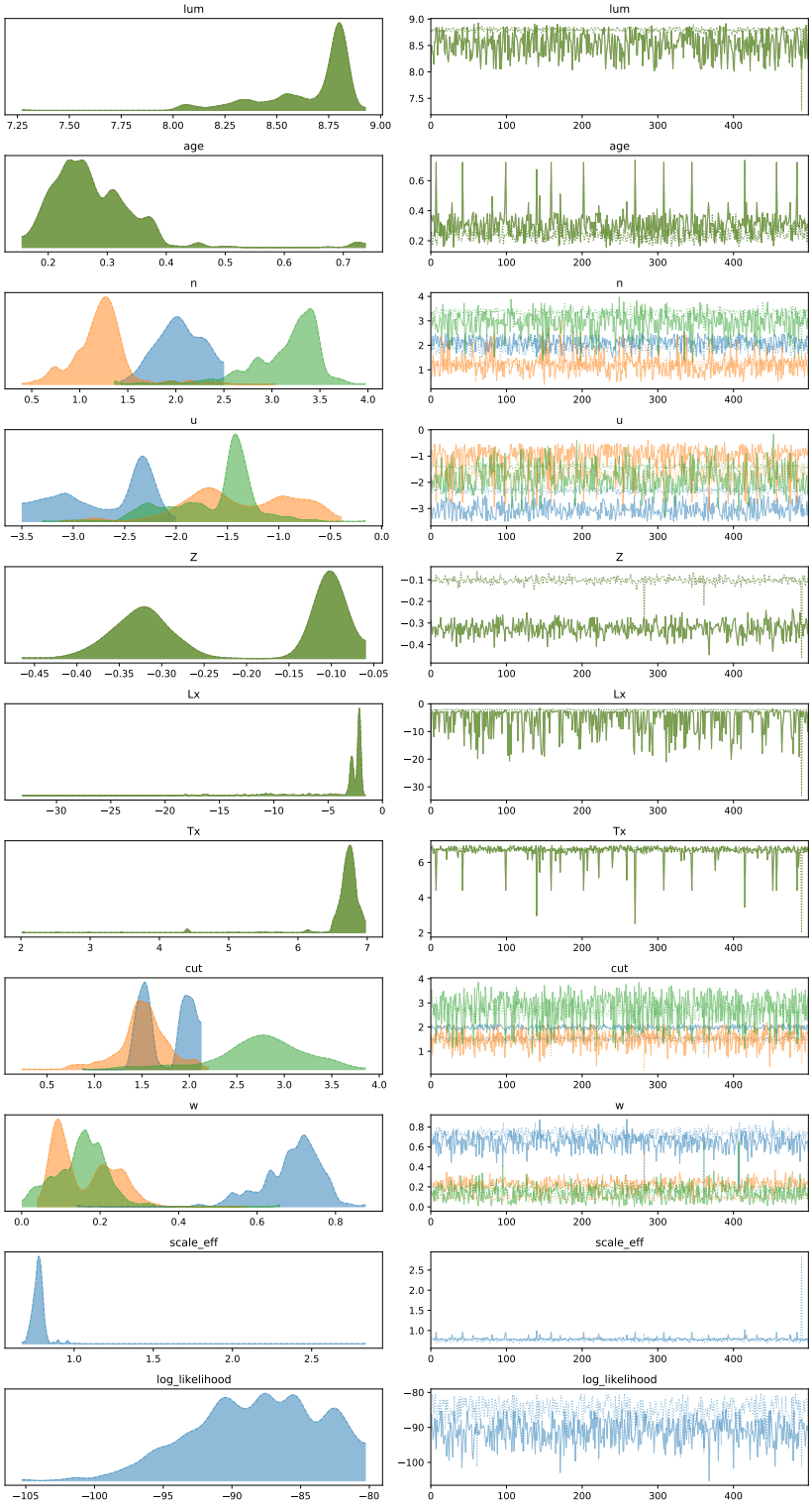

Figure 7.5: Diagnostic plots for primary parameters in a 3-sector model. The colors represent the PDFs obtained for each of the sectors. Some parameters are tied together by priors, hence a single value is assumed for the following parameters controlling the ionizing sources: the total cluster luminosity (1rst row), the stellar age (2nd row), the metallicity (5th row), the X-ray luminosity and temperature (6th and 7th rows). The marginal likelihood and effective scaling factor (2 bottom rows) are also defined uniquely for each model.

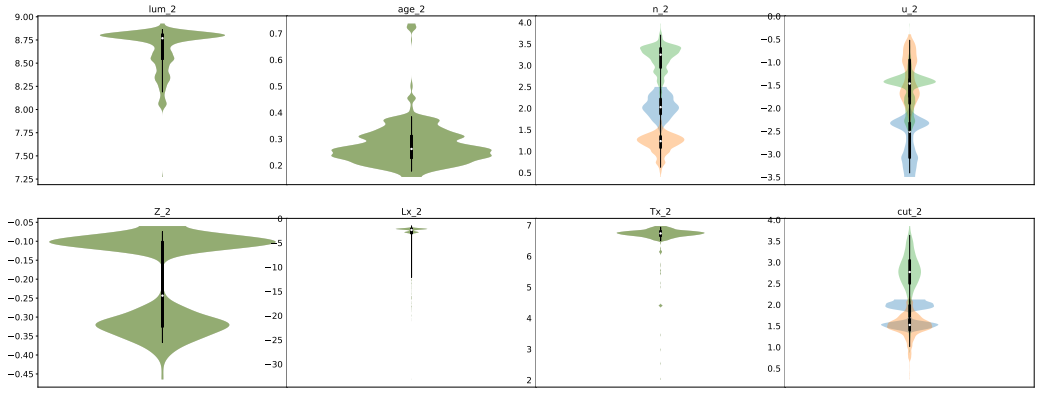

Figure 7.6: Violin-plots for primary parameters showing the median and the HDI at 94% for a 3-sector model. The colors represent the PDFs obtained for each of the sectors

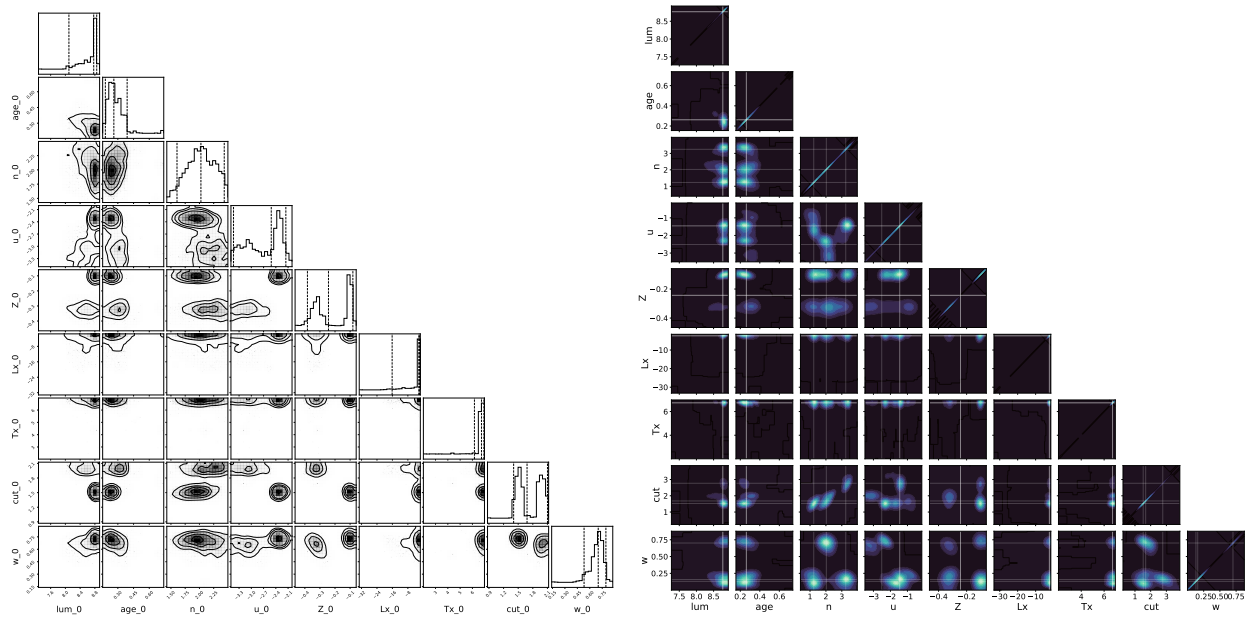

Figure 7.7: **Left**: Corner plot showing 2D-PDFs for pairs of primary parameters for 1 of the 3 components. This visualization tools focuses on the main peak only, hence the axis do not span the whole parameter space. **Right**: Corner plot showing smoothed KDE for pairs of primary parameters for the 3 components.

10.

Kernel Density Estimate

The Kernel Density Estimate (KDE) provides a smoothed approximation of discrete functions by performing a convolution with a specific kernel and a bandwidth. Several kernels can be used to perform the convolution. Using a convolution kernel allows us to obtain a smooth and continuous representation of any distribution. Additionally, the KDE estimates can easily be performed with high dimensionality, which is not the case for histograms. The KDE can be used to visualize trends

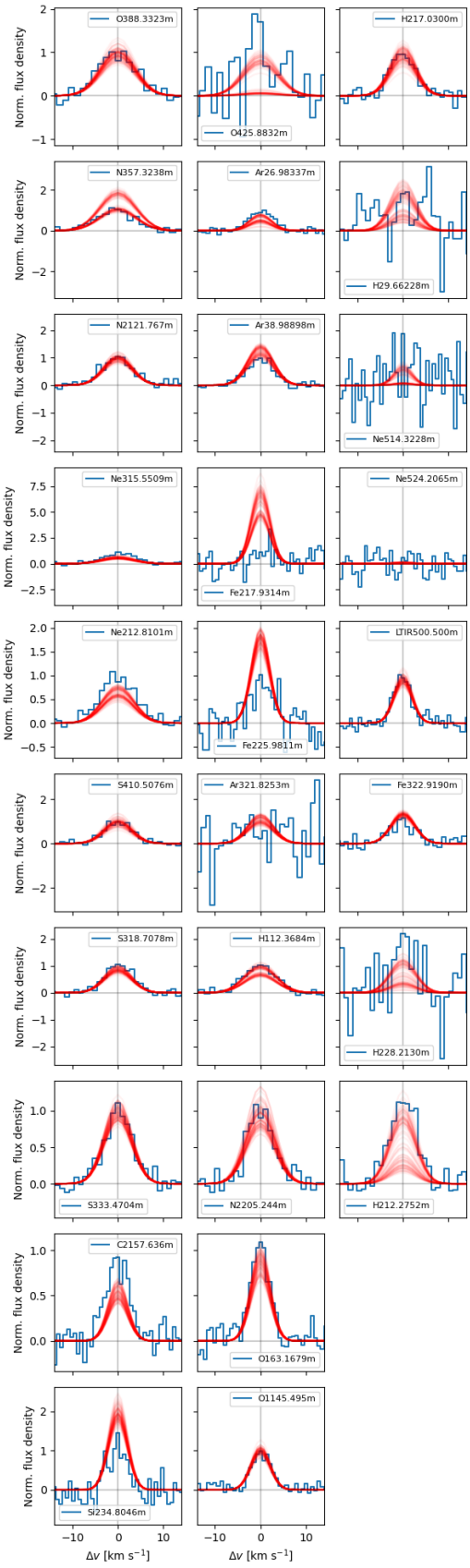

Figure 7.8: Mosaic showing the observations in blue with a noise level corresponding to the observed error bar and the predictions in red.

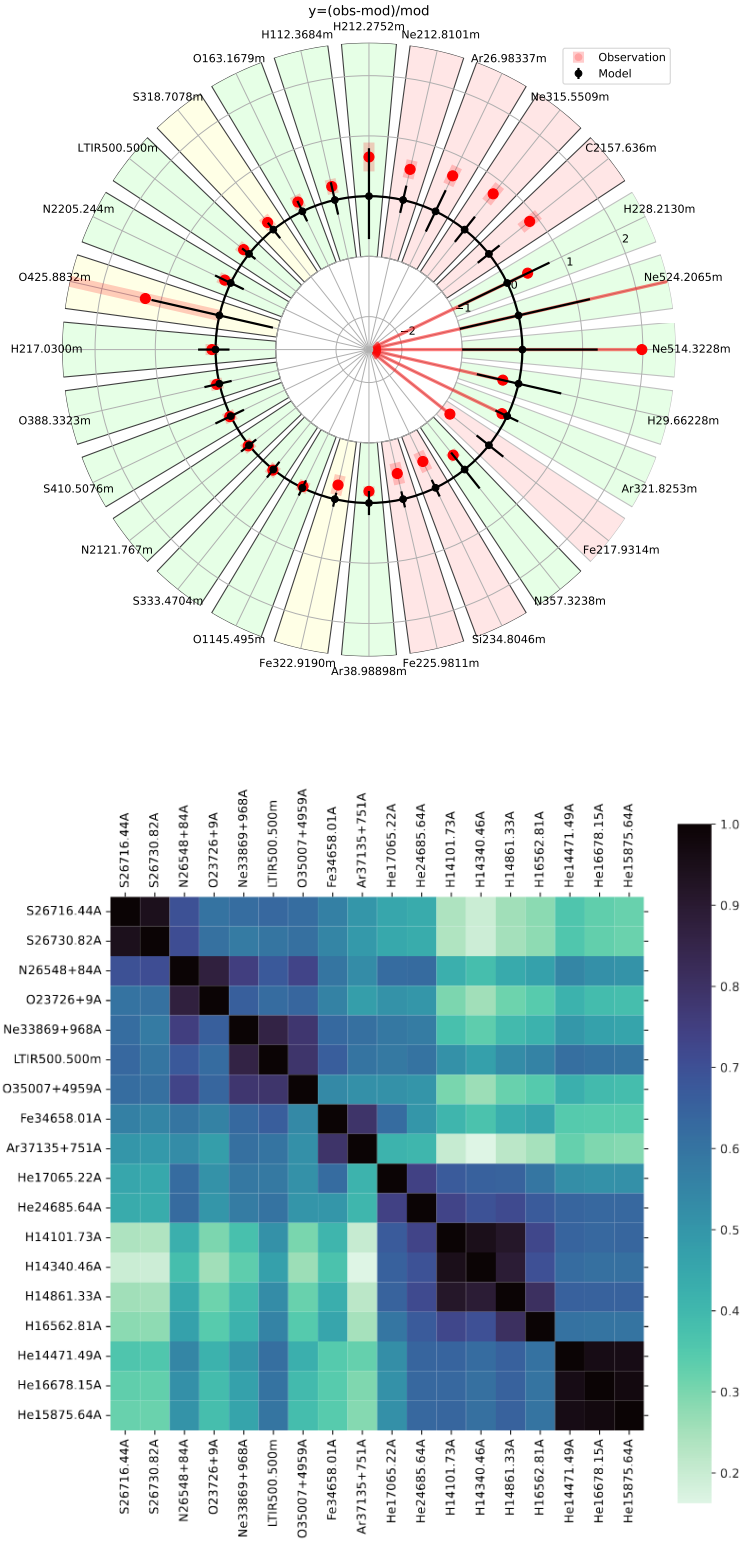

Figure 7.9: **Top**: Relative agreement between observations and predictions; the dark line shows the observed value and associated errorbars while the red dots correspond to model prediction. The color code indicates whether the two values included within each other error bars (green), compatible within error bars (yellow), or not reproduced within error bars (red). **Bottom**: Map of the correlation coefficients between predicted observations for a benchmark run from LR22.

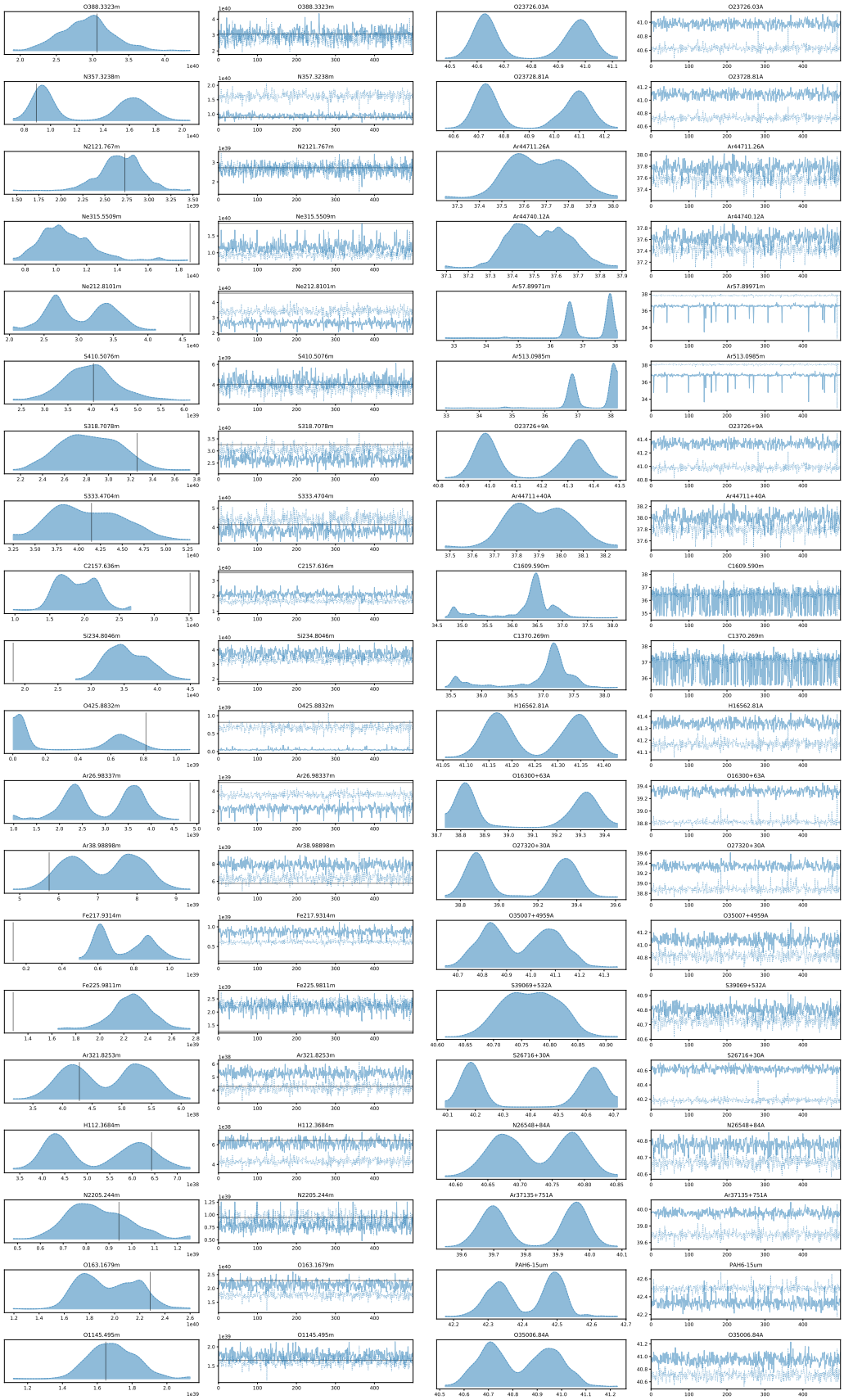

Figure 7.10: Posterior PDFs of post-processed quantities for the emission lines used as constraint (columns 1-2), the predicted line fluxes (columns 3-4).

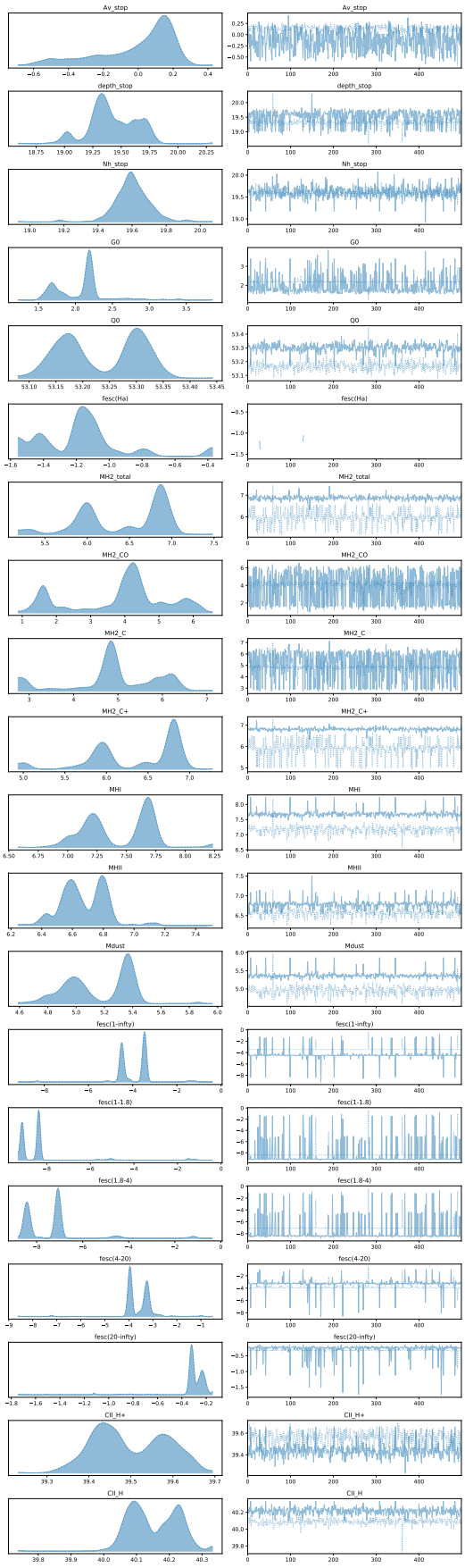

Figure 7.11: Posterior PDFs of post-processed quantities for the secondary parameters.

for both individual objects or for a sample of objects. In the latter case, we concatenate the samples drawn for any given parameter in the whole sample. This allows us to examine the global variations of one parameter with respect to another by plotting their KDE for the whole sample.

Skewed Uncertainty Ellipse

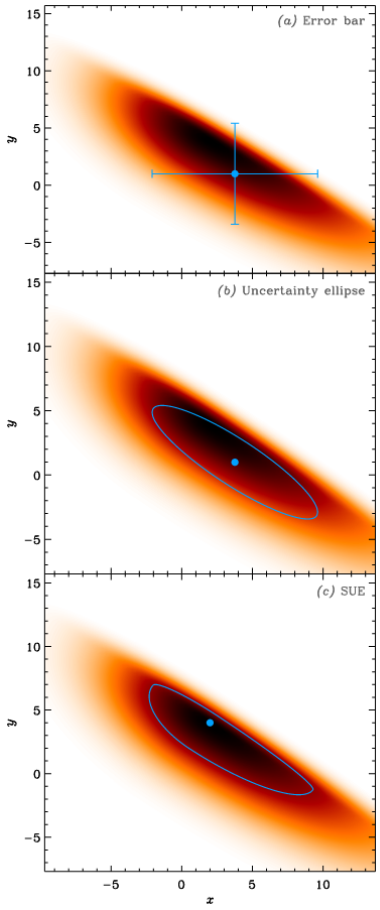

Figure 7.12: Illustration of Skewed Uncertainty Ellipse from (Galliano et al. 2021). The orange density contours in the three panels represent an arbitrary bivariate PDF of two variables, x and y. Panel (a) shows the corresponding traditional error bar: the dot is the mean and the bars show the $\pm 1\sigma$ extent along both axes. Panel (b) shows the widely-used uncertainty ellipse, which can be viewed as the mode and 1σ contour of a bivariate normal distribution having the same means, standard-deviations, and correlation coefficient as the PDF. Panel (c) shows the concept of SUE

.

To represent 2D-PDFs of individual objects, we adopt the Skewed Uncertainty Ellipse (SUE) representation (see e.g., Galliano et al. 2021). A SUE represents the 1σ contour of a 2-dimensional split-normal distribution adjusted to have the same means, variances, skewnesses, and correlation coefficients as the underlying posterior PDF. The center of the SUE marks the location of the robust mean of the 2D-PDF. They are derived based on the calculation of the robust means and variances. We used the detailed procedure described in Galliano et al. (2021).

One minor modification has been added to deal with quantities for which the Median Absolute Deviation (MAD) can be null (e.g., such as the escape fraction). We hence replaced the MAD,

defined as MAD=median(|X-median(X)|), by the 68th percentile of the same quantity |X-median(X)| which ensures that this quantity remains strictly positive. While the representation is convenient to locate the parameter space of highest probability (e.g., 1σ) for a given object, SUEs are not always representative of the underlying PDF, especially when several peaks are present.

In addition to using SUEs which delimit the regions of highest probability around the main peaks of the posterior PDF, one might want to add a point to represent the location of maximum probability. Once again, we stress that the choice of this maximum is somewhat arbitrary. Classical tools include plotting the mean, median, or mode values of the PDF, which can be significantly different for a skewed or asymmetric distribution. Our tests have shown that for the quantities that we discuss in the results Chapters (9 and 10), the mean, median, and mode values are most of the time located within the 1σ contour of the SUEs, centered around the robust mean.

For a few objects, however, those values can lie outside the contours because the PDFs are multipeaked or distorted. In those cases, the SUEs representation is actually not adapted to represent PDFs. In any case, we stress that the SUE representation is only meant to guide the eyes, in a more informative way than just providing a mean value with errors. The analysis of trends and correlations between parameters is always performed using the full MCMC draws without discarding any sample that might lie away from the main probability peak.

7.5 Model comparison

7.5.1 Predictive power

One interest of the flexible configuration scheme presented in Section 4.1 lies in the possibility to run several models assuming different configurations and compare them a posteriori. To do so, one needs to choose a metric to quantify how well a given configuration is performing at reproducing the data. In particular, we have no means to know a priori how many components are required in our multi-components models. While we have motivated in Section 3.2 the interest of using more than one single component, defining how many components are needed – and sufficient – to accurately model the ISM, based on a given set of observable, is still an open question.

Conveniently, the use of a Bayesian framework allows us to use metrics which are classically used to assess how well a model performs at reproducing the data. Among the metrics available in PyMC, we explored the following:

- The Widely applicable information criterion (WAIC) described in Watanabe & Opper (2012).
- The Leave-one-out cross-validation (LOO) described in Vehtari et al. (2017).
- The marginal likelihood $p(\vec{O}|\mathcal{M})$ (see Equation 7.2), i.e., the probability of the observed data given the model (where parameters are integrated out). This quantity is, however, computed only for the SMC sampler (because SMC approximates the full prior space if there are enough samples) and might not be easily extracted for other samplers.

We note that the task of comparing models within a Bayesian framework is complex and still in

development. Other metrics may be found in the literature and determining the best one to use is not straightforward. However, using information criteria (such as WAIC or LOO) or the marginal likelihood presents advantages with respect to just comparing the χ^2 value associated to each model. Among the various metrics presented here, we chose to use the marginal likelihood as a selection criterion in the first applications presented in Chapters 9 and 10.

The marginal likelihood is ideally suited to penalize models which consider too many components without significantly improving the agreement with the data. This is an important improvement with respect to the χ^2 method that asymptotically reaches a minimum value when the number of components increases, reflecting the fact that the fit does not improve significantly anymore. While the χ^2 can give information on the minimal number of sectors needed but does not penalize models having too many components which might overfit the data.

On the other hand, the marginal likelihood reaches a maximum and decreases as the number of potential solutions increases through the larger number of components (see details in Lebouteiller & Ramambason 2022b). Even though increasing the number of components increases the number of potential solutions, such models become less and less likely as the overall number of parameter sets increases (spreading the prior), implying that the probability of the data given the model decreases.

Eventually, our model comparison is inspired by the Bayes factor ($R_{\rm bayes}$) which provides a simple way to compare two models $\mathcal{M}_1$ and $\mathcal{M}_2$, and is defined as:

$$\begin{cases} R_{\rm bayes} = \frac{p(\mathcal{M}_1|\overrightarrow{O})}{p(\mathcal{M}_2|\overrightarrow{O})} \\ R_{\rm bayes} \propto \frac{p(\overrightarrow{O}|\mathcal{M}_1)p(\mathcal{M}_1)}{p(\overrightarrow{O}|\mathcal{M}_2)p(\mathcal{M}_2)} \end{cases} \tag{7.11}$$

Hence, we compare models with different priors or models with a different number of components by using the marginal likelihood $p(\overrightarrow{O}|\mathcal{M})$, under the assumption that both models have the same a priori probability $p(\mathcal{M})$. Although this assumption might not always be verified, it is necessary to allow model comparison since we have no means to estimate the prior probabilities associated with each configuration.

Qualitatively, very simple configurations (e.g., single cluster and single sector) may have lower a prior probabilities because they are less likely to reproduce the complexity of the ISM. Model comparison may be used to select one preferred model or to combine models through averaging:

$$p(\overrightarrow{\theta}|\overrightarrow{O}) = \sum_{i=0}^{K} \alpha_i p(\overrightarrow{\theta}|\overrightarrow{O}, \mathcal{M}_i) \tag{7.12}$$

where K is the number of models and α_i is a weighted coefficient defined based on the comparison criterion. For example, α_i can be defined based on the marginal likelihood as follows:

$$\alpha_i = \frac{p(\mathcal{M}_i|\overrightarrow{O})}{\sum_{j=0}^{K} p(\mathcal{M}_j|\overrightarrow{O})} \tag{7.13}$$

An alternative to weighing through the marginal likelihood is to use information criteria methods such as WAIC or LOO with the Akaike weight for a model i as:

$$\alpha_i = \frac{e^{-\frac{1}{2}dIC_i}}{\sum_{j=0}^{K} e^{-\frac{1}{2}dIC_j}} \tag{7.14}$$

with $dIC = 2 \times \text{elpd}_{\text{LOO,WAIC}}$ where elpd is the expected log pointwise predictive density (Gelman 2014) for either LOO or WAIC defined as:

$$elpd = \sum_{i}^{n} log \int p(O_i|\theta) p_{\text{post}}(\theta) d\theta. \tag{7.15}$$

7.5.2 Inferring the optimal number of components

The number of components has to be set explicitly in MULTIGRIS but there are some methods to estimate how many components are necessary, either by running models with a different number of components and compare their likelihood or by running a model with n components and infer whether up to $n-1$ components can be ignored. The likelihood reaches a maximum for the most likely number of components given the data (and obviously the prior). In other words, the data is not sufficient to constrain with high likelihood a higher number of components.

The number of components can thus be considered as a free parameter as well. This could, in theory, provide a better way to test whether the solution converges to parameters with similar values with any number of components. It can be seen as the model solution for a maximum or n components. Nevertheless, the tests performed using this method have shown that the results are not always in agreement with the best models found by using the other metrics previously mentioned and tends to be biased toward a low numbers of component. Because of these drawbacks, we did not select this method, although it provides, in theory, an interesting way to select the optimal number of component in a Bayesian way. This interesting alternative may be further examined in the future.

Part IV

Results

Chapter 8

Application to the ISM

The framework presented in Chapter 7 is quite general and can be used to compare predictions and observations associated with various physical applications. In the context of this thesis, we focus specifically on applying `MULTIGRIS` to ISM studies. In this chapter, we present the first tests that were performed to ensure that our new code is well suited for this science case.

Among the crucial points to be tested, we first made sure that the framework is self-consistent and successfully able to recover sets of parameters associated with mock observations generated with the photoionization and photodissociation code `Cloudy` (see Section 8.1). We then tested our algorithm by predicting simple observables for well-known samples of local galaxies (see Section 8.2). While those tests cannot ensure that our new algorithm outputs results which truly describe the actual structure of the ISM, they ensure that our results are self-consistent and in agreement with previous studies performed in the local universe. Those sanity checks are crucial to build a solid ground before diving in the analysis of more complex quantities and topologies in Chapters 9 and 10 and applying `MULTIGRIS` to other samples of galaxies, including poorly constrained ones (see Section 11.2.1).

8.1 Benchmarks on mock observations

8.1.1 Recovering the right configuration

The first series of tests aimed at recovering the most likely configuration (see Section 4.1) based on mock observations. This is essential to ensure that the various topologies (see Chapter 4.1) can be uniquely identified when enough constraints are given. The tests were performed on single cluster multisector configurations, since they are the ones used in the applications presented in LR22 and Ramambason et al. (2022). As described in Section 7.5, different metrics can be used to compare configurations, which do not always yield the same results. In the top panels from Figure 8.1, we compared the values obtained for different indicators: the likelihood, the marginal likelihood (see Equation 7.2), the LOO, and WAIC. We refer to Section 7.5.1 for the definition of those metrics.

Applying `MULTIGRIS` on mock observations with a number of sectors varying from 1 to 3, and

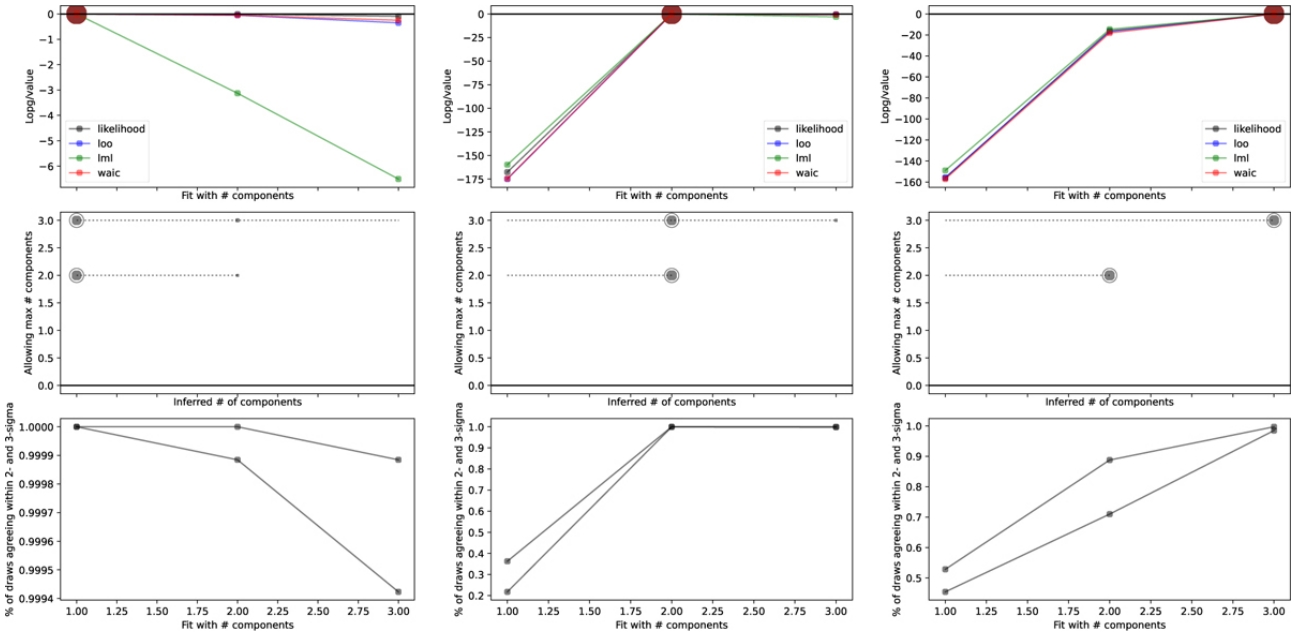

Figure 8.1: Test on mock observations with varying number of sectors from LR22 performed on models with an input of 1, 2, and 3 sectors from left to right. In the top panel we show the log values of LOO, WAIC, and the SMC marginal likelihood ("lml") for MCMC runs with a number of sectors from 1 to 3. In the middle panel, we show the likelihood weights of each sector when the number of sectors is allowed to vary. Each horizontal line corresponds to a model with the number of sectors allowed to vary up to 2 or 3. For each horizontal line, the filled circles show the inferred weight and the circled one highlights the number of sectors with the largest weight. In the bottom panel, we show the fraction of draws that agree with the observed value within 2- and 3-σ

using as constraint a complete list of 13 emission lines[1] and considering a fixed uncertainty of 0.2 dex for all lines, we find that all those metrics are in agreement and successfully select the correct number of sectors. The agreement between the 4 metrics is shown in the first row of Figure 8.1. In the first column, we note that the metric which penalizes the most models considering too many components (e.g., considering 2 or 3 sectors to reproduce single-component models) is the marginal likelihood ("lml"). This behavior is due to the marginalization which leads to drastically increasing the marginal likelihood when the number of free parameters increases (see discussion in Section 7.5.1).

The second row in Figure 8.1 shows the optimal number of sector found using the inference method (i.e., considering the number of components itself as a free parameter) presented in Section 7.5.2. We show the results of two runs with respectively 3 and 2 sectors by default, for which a binary random variable allows ignoring the contribution of one or two sectors. In this test, we find that the number of ignored sectors is always in good agreement with the true number of sectors.

Finally, the bottom panel shows the fraction of draws that agree with the observed value within 2- and 3-σ (see Section 7.4.2). As expected, this fraction is low when not enough sectors are considered (e.g., only one or 2 sectors to reproduce a 3-sector model). Furthermore, this fraction reaches a maximum (equal to unity is the ideal case of mock observations) when the correct number of sector is considered. However, when too many sectors are considered, this indicator remains very close

[1] [C II]157μm, [O I]63μm, [N III]57 μm, [Ar II]26μm, [Ar III]8μm, [O III]88μm, [N II]205μm, H_2 17μm, [Si II]34μm, [Ne II]12μm, [Ne III]15μm, [O I]145μm, and [Ne V]14μm.

to one since the addition of sectors does not significantly decrease the agreement with observation. Although this metric is informative, it does not allow distinguishing between configurations as long as they have a sufficient number of sectors. This issue is reminiscent of the limitations listed for the χ^2 method (see Section 3.2.2) and motivates the use of other metrics such as the marginal likelihood.

8.1.2 Recovering parameter values

We have seen that in the ideal case of a mock observation for which we consider a large suite of lines, our code is able to uniquely identify the best configuration. However, when applying the code to real observations, several uncertainties might complicate the identification of a favored configuration such as, e.g., the uncertainties on the emission lines, the uncertainties on the models themselves, the use of fixed abundances and density profiles, or the insufficient number of available constraints.

Due to all those uncertainties, the configuration favored by the code might not correspond to the configuration that would best match the galaxy in an ideal, perfectly unbiased and well-constrained observation. Hence, we want to quantify the effect of different configurations on the values inferred for the most important parameters in our study. This test is useful to understand which quantities are the most sensitive to the topology and what uncertainties are associated with the choice of the configuration itself.

We show on Figure 8.2 the variations of the inferred parameters and associated error bars (quantiles) when forcing a given number of sector to reproduce a fixed suite of lines. The tests were performed separately on radiation-bounded configuration and density-bounded configurations. When the true solution corresponds to a single sector, using a model with two or three sectors does not change significantly the parameter values, except for the escape fraction of ionizing photons. However, such models become more and more unlikely as the number of sectors increases. Similarly, using three sectors for a true solution with two sectors leads to a good agreement with expected results, this time even for the escape fraction.

More generally, using a larger than necessary number of sectors usually does not impact significantly the results except for the escape fraction in some cases. On the other hand, using a lower than necessary number of sectors may result in significant differences for all parameters, but keeping in mind that the model is likely unsatisfactory (see Figure 8.1). We find that the parameters which depend the most on the topology are the ones which depend on the depth of each sector, that is, the escape fraction of ionizing photons but also the mass of H_2.

8.1.3 Discussion on the optimal number of constraints

In Figures 8.3 and 8.4 we show a benchmark test analyzing the evolution of the solution with an increasing number of lines. For this illustration, the series of lines we consider is sorted based on their detection or potential detection in high-z sources with telescopes such as ALMA or IRAM/NOEMA. The first set of three lines includes [C II], [O III]88 μm, and [O I]63 μm. We then add [N II]205 μm and [O I]145 μm, then [Ne II]12.8 μm, [Ne III]15.6 μm, [N III]57 μm and [Ar II]7 μm, then [Ar III]9 μm, H_2 17 μm, [Si II]34 μm, and [Ne V]14 μm. As the figures show, the solution naturally improves with the number of lines considered. In the radiation-bounded case (eight parameters; Fig. 8.3), a satisfactory

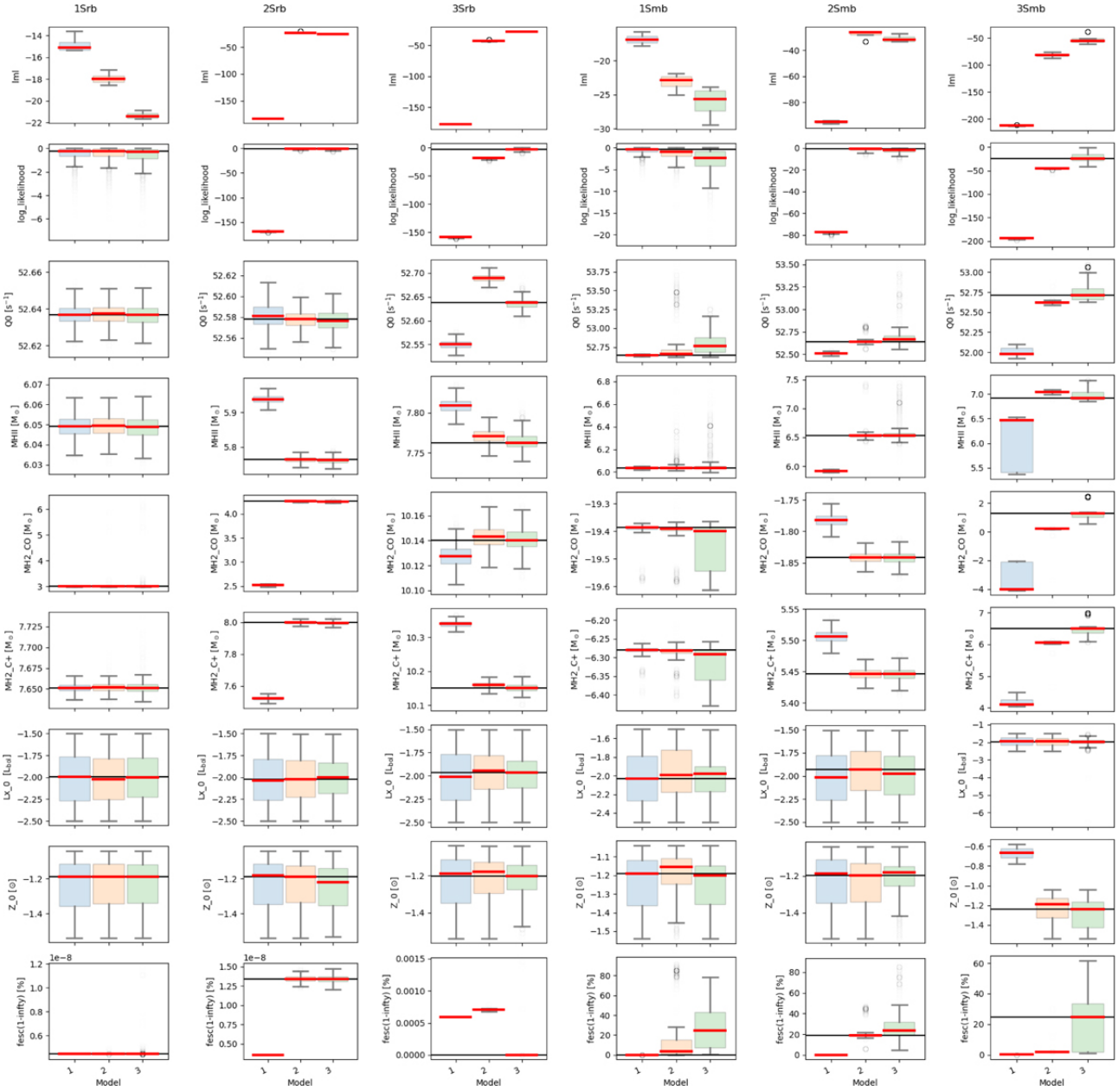

Figure 8.2: Benchmark to estimate the number of sectors required, for models with an input of 1, 2, and 3 sectors and for radiation-bounded ("rb") or density-bounded (or matter-bounded, "mb") template models. Each plot shows a standard box plot configuration with the median value in red, with the box spanning from the first to third quartile, and the horizontal gray bars extending to 1.5 times the interquartile range. Faint open circles show the data that extend beyond the horizontal gray bars. In each panel we show the results forcing the number of components from 1 to 3. The two top panels show the marginal likelihood ("lml") and the likelihood while the other panels below show several secondary parameters. The parameters examined here are from top to bottom the number of ionizing photons, the mass of H^+, the mass of H_2 traced by CO, the mass of H_2 traced by C^+, the X-ray luminosity, the metallicity, and the escape fraction of ionizing photons. The horizontal lines show the expected value, except for the likelihood (second panel) where it marks the maximum value.

solution is found even with three lines. For the density-bounded case (ten parameters; Fig. 8.4), for which there is an added degeneracy due to the inclusion of the depth parameter, a satisfactory solution

is found with nine lines.

While this test provides a general idea of the number of lines necessary to constrain radiation-bounded and density-bounded two-sector configurations, the situation becomes much more complex when considering more complex configurations and applying the code to real observations and their associated uncertainties. As a general rule, increasing the number of constraints available should in theory help to constrain more parameters and lift degeneracies. However, the situation may be more complex if outliers or observations with abnormally small error bars are present in the dataset. Hence, choosing the optimal set of lines to constrain a model is not a trivial issue. We refer to LR22 for a more extended discussion on this topic and report here the main steps that are proposed to select a suitable set of lines:

- Include all available observations for a first run.
- Identify observations that are not well reproduced in order to consider potential missing systematic uncertainties and/or incorrect model assumptions (e.g., number of parameters and specific configuration).
- Check if some observations effectively constrain at least one parameter, and inversely, if every parameter is constrained by at least one observation.
- Check if some observations constrain the same parameter(s), in which case redundant observations may be removed to prevent the likelihood from being controlled only by few parameters without changing significantly the solution.

8.2 Comparison with empirical constraints

We now compare the prediction obtained by `MULTIGRIS` for important physical quantities with estimates based on empirical calibrations. In this section, we summarize and extend the main results that are presented in the consistency checks from Ramambason et al. (2022). We refer the reader to Section 9.3 for further details. While the results in this section are only shown for multisector configurations, we also performed tests using more complex configurations based on distributions of parameters (see Sections 9.4.2 and 10.3.5). Such level of complexity is not needed to reproduce the observables studied in this section, but we made sure that they provide values in good agreement with the results we present here for metallicity, SFR, and mean ionization parameter.

8.2.1 Metallicity

We used `MULTIGRIS` to estimate the metallicities of a sample of 39 local galaxies from the Dwarf Galaxy Survey (described in Section 2.3.1). One specificity of our code compared to previous works on this sample is that it uses the information provided by all the lines available to derive a posterior PDFs. The metallicity is a primary parameter (see Section 5.2) in our grid of models on which we performed a linear interpolation.

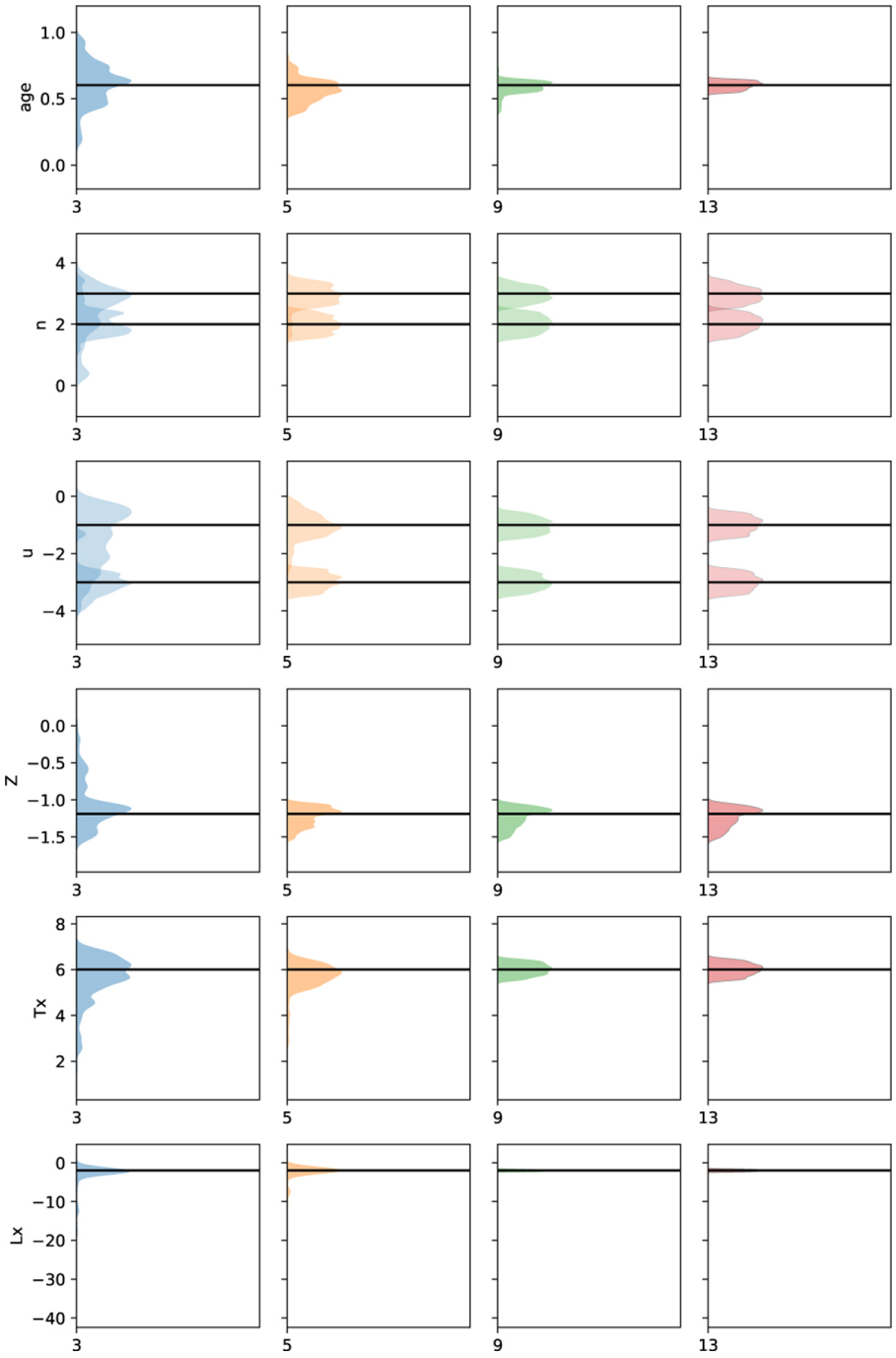

Figure 8.3: Evolution of the solution with the number of lines considered for a benchmark run from LR22. Each row corresponds to a parameter, and each column represents a different number of lines from three to 13 ([C II]+ [O III]88 μm+[O I]63 μm, +[N II]205 μm+[O I]145 μm, +[Ne II]12.8 μm+[Ne III]15.6 μm+[N III]57 μm+[Ar II]7 μm, +[Ar III]9 μm+H_2 17 μm+[Si II]34 μm+[Ne V]14 μm. The horizontal black lines indicate the expected solution. The model used here is radiation-bounded

Our predictions are derived using a weakly informative prior on metallicity. We use a Gaussian prior centered on the measured metallicity with a large width (σ = 0.1+ $\sigma_{\rm mes}$ in logarithmic scale, where $\sigma_{\rm mes}$ is the uncertainty associated with the measured metallicity). The prior is used to lift potential degeneracies between stellar age and metallicity, as described in Ramambason et al. (2022). It prevents the code from considering solutions which are completely offset from the metallicity found with optical lines. However, since no hard constraint is imposed on the metallicity, the posterior PDF

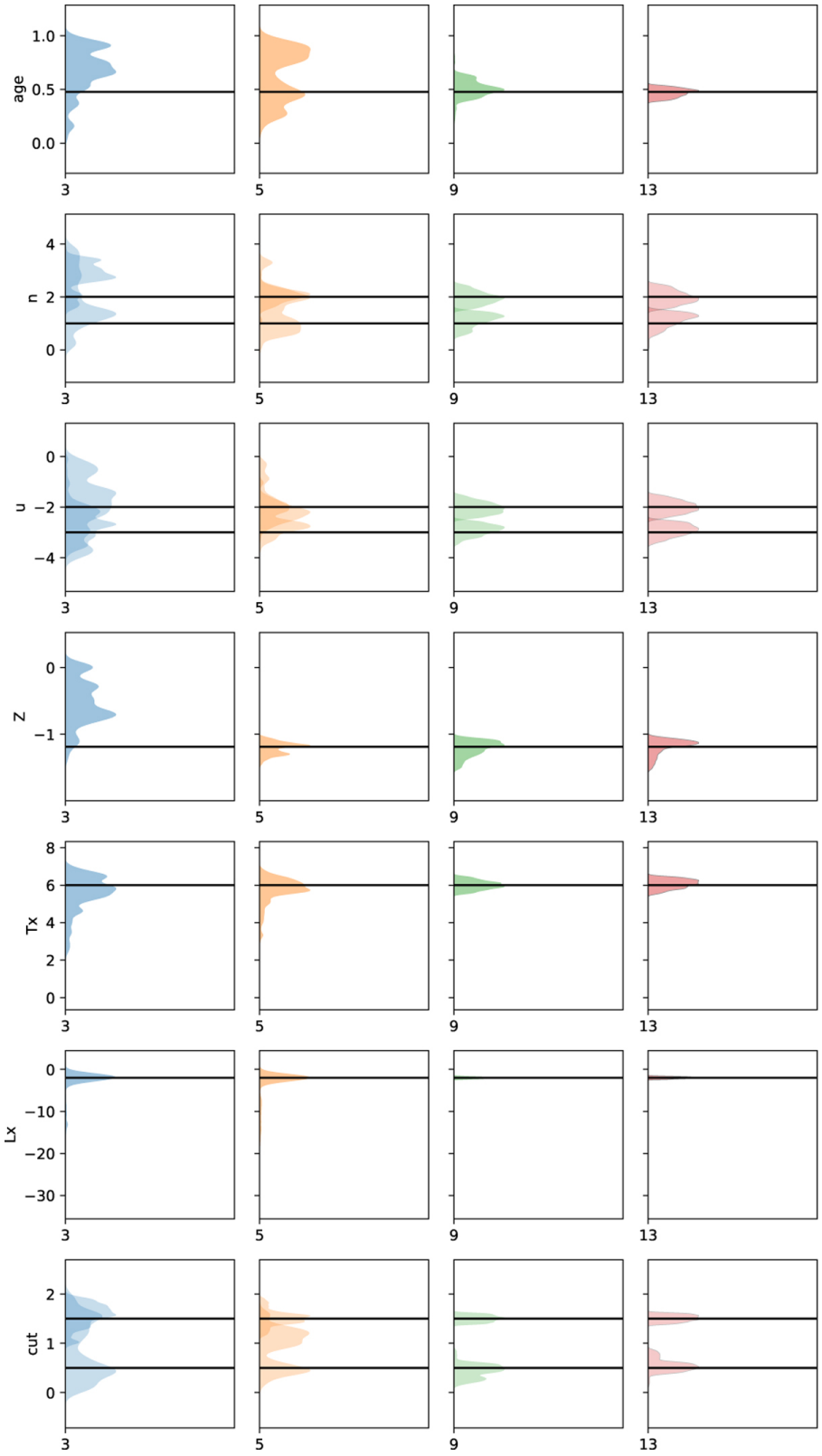

Figure 8.4: Same as Figure 8.3 but for a density-bounded model.

remains free to diverge from the prior distribution, if the information given by the emission lines is not compatible with it.

In the left-hand side panel of Figure 8.5, we compared our measurements to predictions based on the optical oxygen lines ratio (R23-P) method from Pilyugin & Thuan (2005). This method is

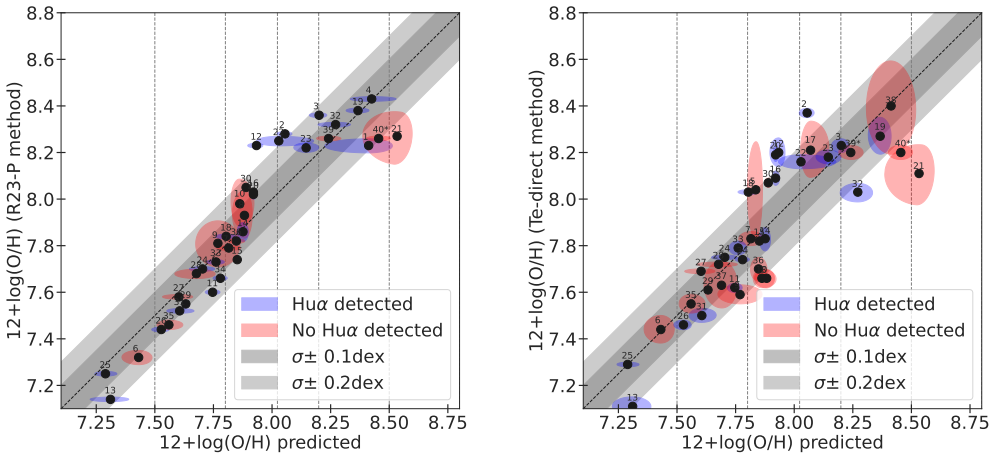

Figure 8.5: **Left**: Predicted metallicity as compared to metallicity estimate from the R23-P method (Pilyugin & Thuan 2005). No measurements are available for 4 galaxies (HS0017+1055, HS0052+2536, HS2352+2733 and Mrk1089). **Right**: Predicted metallicity as compared to metallicity estimate from the T_e-direct method (Izotov et al. 2006). No measurements are available for 2 galaxies (Haro11 and He2-10). The 1σ contours of the SUEs are calculated with a dispersion corresponding to the measured uncertainty for the y-axis.

applicable to all galaxies in our sample except 4 (HS 0017+1055, HS 0052+2536, HS 2352+2733 and Mrk 1089). Their method is based on a two-parameter calibration involving the R23 ratio and an excitation parameter P, which corrects the estimates by accounting for the physical conditions in the H II regions. We found metallicities in agreement within 0.1 dex for 31 out of 36 galaxies and within 0.2 dex for 35 out of 36 galaxies.

In the right-hand side panel from Figure 8.5, we compared our measurements to predictions based on the T_e-direct method from Izotov et al. (2006). This method relies on the detection of the auroral line [O III]4363Å which is detected in all galaxies but two in our sample (Haro 11 and He 2-10). We find a somewhat larger scatter when comparing with the T_e-direct method but our predictions remain in agreement within 0.1 dex for 25 galaxies out of 37 and within 0.2 dex for 36 out of 37 galaxies.

We note that 3 galaxies stand out in both plots:

- For NGC 1705 (galaxy #21), we systematically overpredict the metallicity with respect to the R23-P method and T_e-direct method. We note that this is the only galaxy in our sample in which a metallicity gradient has been reported (Annibali et al. 2015). The latter study reports a significant negative gradient in oxygen abundance of -0.24 ± 0.08 dex kpc^{-1}. This galaxy stands as an exception since other studies of dwarf galaxies in the nearby universe have reported nearly flat metallicity gradients (Kobulnicky & Skillman 1997; Croxall et al. 2009; Haurberg et al. 2013; Lagos & Papaderos 2013). Thus, we be biased by the use of IR tracers which may trace mostly the central part of the galaxy, where the metallicity is higher.

- For II Zw 40 (#12), we systematically underpredict the metallicity with respect to the R23-P method and T_e-direct method. While the reason for this is not obvious, we discuss here some specific features which may influence our results. First, II Zw 40 is a known merging system

composed of two small interacting galaxies (Baldwin et al. 1982). Additionally, other studies have reported hints of an inhomogeneous gas and dust structure for II Zw 40, with localized dust-enrichment possibly associated with massive star clusters (Consiglio et al. 2016).

- For Haro 3 (#2), we also systematically underpredict the metallicity. Several studies have hinted at the presence of a massive dust lane close to the stellar cluster, which might significantly contributed to the IR emission (Johnson et al. 2004; Hunt et al. 2006).

Hence, for those two latter galaxies, evidence suggests that the dust is not well-mixed with the gas as it is assumed in our `Cloudy` models. The discrepancy we observed between our results based on IR lines and empirical methods based on optical tracers might be linked to attenuation effects related to their specific dust geometries, which may affect the oxygen lines used in the optical-based diagnostics of metallicity.

8.2.2 Star formation rates

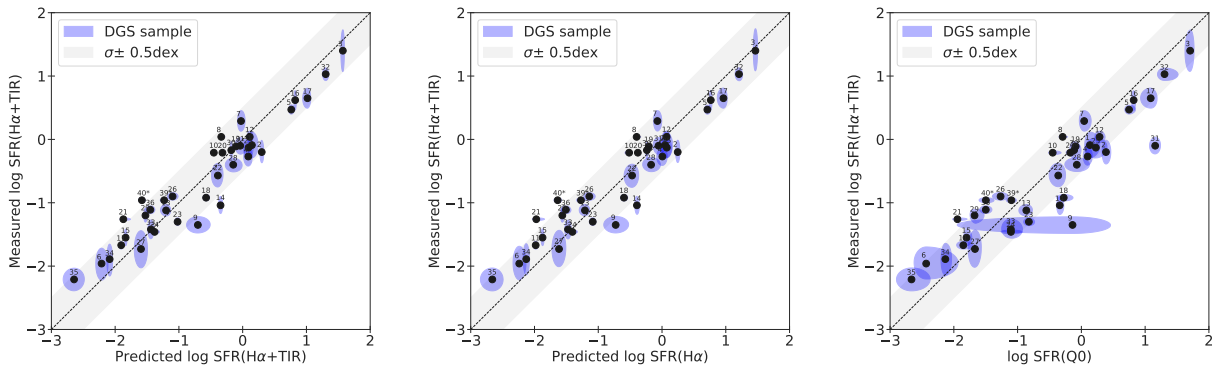

Figure 8.6: From left to right: Predicted SFR(Hα+TIR), SFR(Hα), and SFR(Q(H^0)) vs. empirical estimates from Rémy-Ruyer et al. (2015) based on Hα and TIR. The 1σ contours of the SUEs are calculated with a dispersion corresponding to the measured uncertainty for the y-axis. For 4 galaxies (HS 0017+1055, HS 0052+2536, HS 1319+3224 and HS 2352+2733) no Hα measurements are available. For 2 of those galaxies (#5: HS 0052+2536 and #9: HS 1319+3224), the measured SFR is derived in using diagnostics based on FUV.

While star formation rates are not directly evaluated by our code, we can provide predictions based on classical star formation rate tracers (see Section 5.3.2). In Ramambason et al. (2022), we predicted the intrinsic Hα intrinsic luminosities, the total infrared luminosities ($L_{\rm TIR}$), and the total number of ionizing photons Q(H^0) based on the observed emission lines in the DGS. We emphasize that those predictions were derived using exclusively emission lines in the infrared domain. Those quantities were then used as proxies to derive SFR estimates in this sample of star-forming galaxies. We use three different approaches to estimate the SFR based on calibrations from Calzetti et al. (2012) and Kennicutt et al. (2009) and derive SFR(Q(H^0)), SFR(Hα), and SFR(Hα+TIR) using respectively Equations 5.8, 5.9, and 5.10 presented in Section 5.3.2.

In Figure 8.6, we compare the SFR predicted by `MULTIGRIS` based on those three proxies to SFR estimates provided in Rémy-Ruyer et al. (2015) which used the prescription from Kennicutt

et al. (2009) based on Hα and $L_{\rm TIR}$. We find a globally good agreement between our SFR predictions with the measurements from Rémy-Ruyer et al. (2015). More specifically, the values derived using SFR(Hα+TIR) and SFR(Hα) are in agreement within 0.5 dex for all galaxies except one (#14: Mrk 153). The discrepancy for this galaxy is linked to the fact that we overestimate Hα for this galaxy, which may due to extinction effects. We observe a larger scatter for our prediction based on SFR(Q(H^0)) which tends to be slightly higher than the prediction we obtain for SFR(Hα) and SFR(Hα+TIR). This effect is linked to the fact that SFR(Q(H^0)) includes photons that may have been absorbed if the medium was radiation-bounded. We refer to Ramambason et al. (2022) for further discussion on this topic, in which we derive a correction to apply to SFR in order to account for escaping photons.

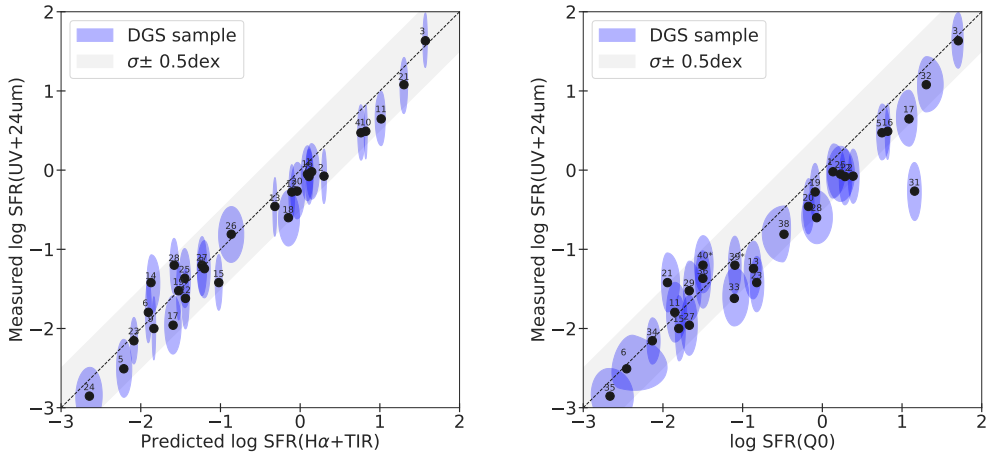

Figure 8.7: Predicted SFR(Hα+TIR and SFR(Q(H^0)) vs. empirical estimates from Rémy-Ruyer et al. (2015) based on mixed tracers FUV and 24μm. The 1σ contours of the SUEs are calculated with a dispersion corresponding to the measured uncertainty for the y-axis.

Finally, we tried to compare our SFR prediction to estimates based on UV proxies, since the latter are sensitive to star formation over larger timescales than the Hα and TIR proxies. In theory, UV proxies might be of interest to get information on the star formation history. The UV fluxes from GALEX were available for 28/40 galaxies in our sample. SFR estimates corrected for attenuation effects were provided in De Looze et al. (2014) using calibrations from Hao et al. (2011) and Kennicutt et al. (2009) based on GALEX FUV and MIPS 24μm.

In Figure 8.7, we provide two plots showing that our predictions are in globally good agreement (below 0.5 dex except for #31: Tol 1214-277, which is poorly constrained with only one sector) with the estimated SFR(UV+24μm). We compare the SFR(UV+24um) to the predicted SFR(Hα+TIR) and SFR(Q(H^0)) and find that in both cases we tend to slightly overpredict SFR with respect to SFR(UV+24μm), especially among the highest SFR values in our sample (i.e., log SFR > -1) which also correspond to the lowest metallicities. This effect was already pointed out in De Looze et al. (2014) who compared various SFR calibrations based on FUV or Hα for the DGS sample and found a discrepancy between the two tracers in the low-metallicity regime. The authors caution the use of FUV as a SFR tracer in local metal-poor dwarf galaxies likely dominated by one or a few bursts of star formation, for which the hypothesis of constant star formation might no longer hold. They state that the use of FUV as a tracer of SFR might lead to underestimating the true level of star formation

in the case of bursty SFH.

8.2.3 Ionization parameter

The ionization parameter is a key parameter controlling the line emission. However, disentangling the effect of the ionization parameter from other galactic parameters (e.g., metallicity, stellar age, f_{esc}) is not an easy task. We examine in this section two main issues that make the interpretation of the ionization parameter values difficult:

- First, the definition of a global averaged ionization parameter integrated on galactic scales is not straightforward and the adopted conventions vary in the literature. Since the ionization parameter is intrinsically a *local* quantity which varies with the distance from ionizing sources, hydrogen density and filling factor of the gas, the trends which are derived in various studies strongly depend on the adopted definition for the average ionization parameter.

- Second, the ionization parameter has a strong dependency on other galactic parameters, making it difficult to isolate its effect on emission line fluxes.

In this section, we provide a physically-motivated definition of an average ionization parameter $<U>$ and examine its dependencies on other galactic parameters.

Defining $<U>$

The input ionization parameter U_{in} in the SFGX database is defined at the illuminated edge corresponding to the inner radius of H II regions (see Equation 5.4). Nevertheless, this definition is not representative of the physical conditions throughout the whole H II region and does not account for its geometry. In particular, this definition is not appropriate for thick shells of gas for which the inner ionization parameter (U_{in}) defined at the illuminated front can be very different from the ionization parameter at the ionization front U_{out}. One commonly used workaround (e.g., Stasińska et al. 2015) is to define a volume-average ionization parameter as follows:

$$< U >= \frac{1}{V} \int_{R_{in}}^{R_{out}} U(r) dV \tag{8.1}$$

where V is the total volume of the H II region and $U(r)$ is the local ionization parameter defined at radius r. In our `Cloudy` models, this can be calculated by summing over the different zones of each model as follows:

$$< U >= \frac{3}{4\pi R_{out}^3} \sum_{i=0}^{N_{zones}} \frac{Q(H^0)}{4\pi n_i c r_i^2} dV_i \tag{8.2}$$

We note that this definition can be used both for radiation-bounded models (i.e., R_{out}=R_S, the Strömgren radius) and for density-bounded model (i.e., $R_{out} < R_S$).

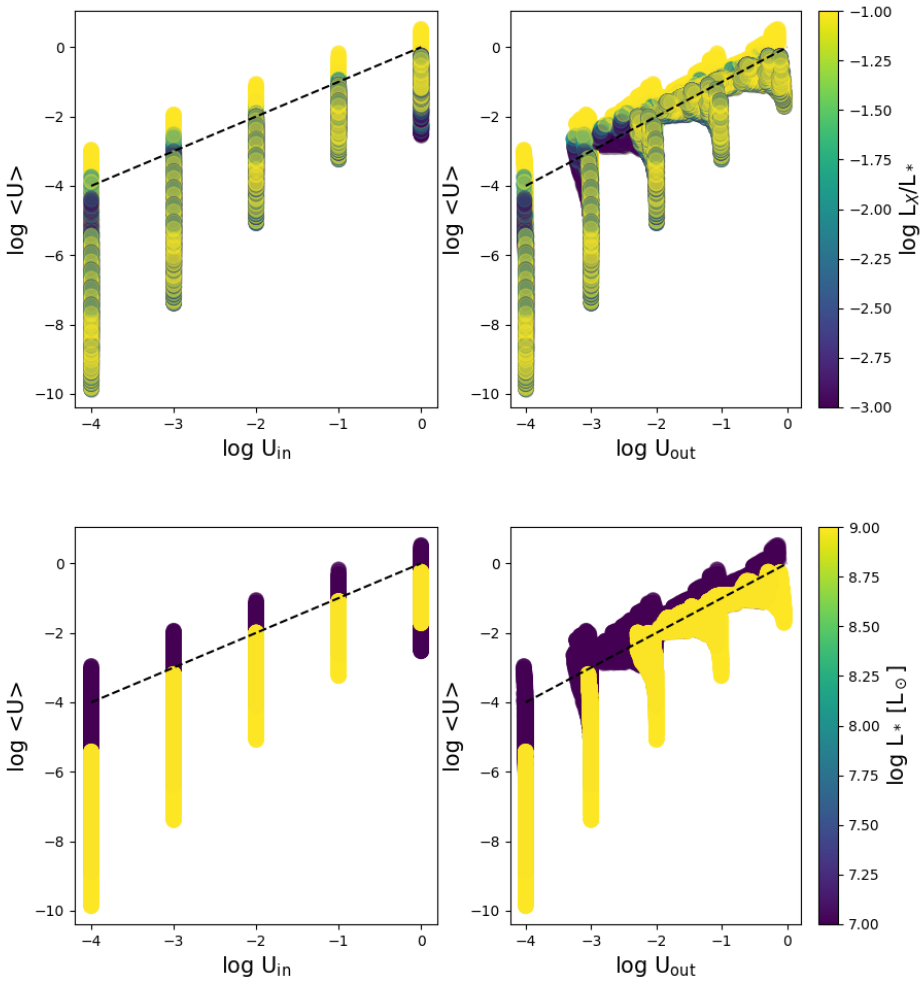

Figure 8.8: Evolution of the volume-averaged ionization parameter $<U>$ as a function of $U_{\rm in}$ and $U_{\rm out}$ in the model grid. The color code represents the X-ray luminosity ratio L_X/L_* (first panel) and the stellar luminosity L_* (second panel). We show only radiation-bounded models (i.e., cut=1). The dotted line shows the 1:1 relation.

$<U>$ vs. $U_{\rm in}$ and $U_{\rm out}$

In Figure 8.8, we show the relation between $U_{\rm in}$, $U_{\rm out}$ and the volume-averaged $<U>$ color coded with parameters controlling the ionizing sources (stellar and X-ray). As expected, when no X-ray source is present, $U_{\rm in}$ is always an upper limit to the volume-average $<U>$. We note, however, that when we add a contribution from an X-ray source, $<U>$ can sometimes reach values slightly higher than the $U_{\rm in}$ for the highest X-ray contribution in our grid (10%), as shown in the first panel. This is only the case for models having a total stellar luminosity of $10^7\,L_\odot$ for which the gas is set closer to the ionizing source for a same $U_{\rm in}$. In those models, the hard photons produced by the X-ray sources are impinging thick shells of gas placed close to the ionizing source. In this specific configuration, the two-photon radiation[2] (Osterbrock & Ferland 2006) becomes non-negligible, hence acting as an additional local source in terms of number of ionizing photons produced by the nebular gas.

[2]The two-photon radiation is a de-excitation process in which two photons are emitted, their energy being equal to the energy of the transition. The probability of 2-photon emission is small, but it can become the main channel for deexcitation if collisions are negligible.

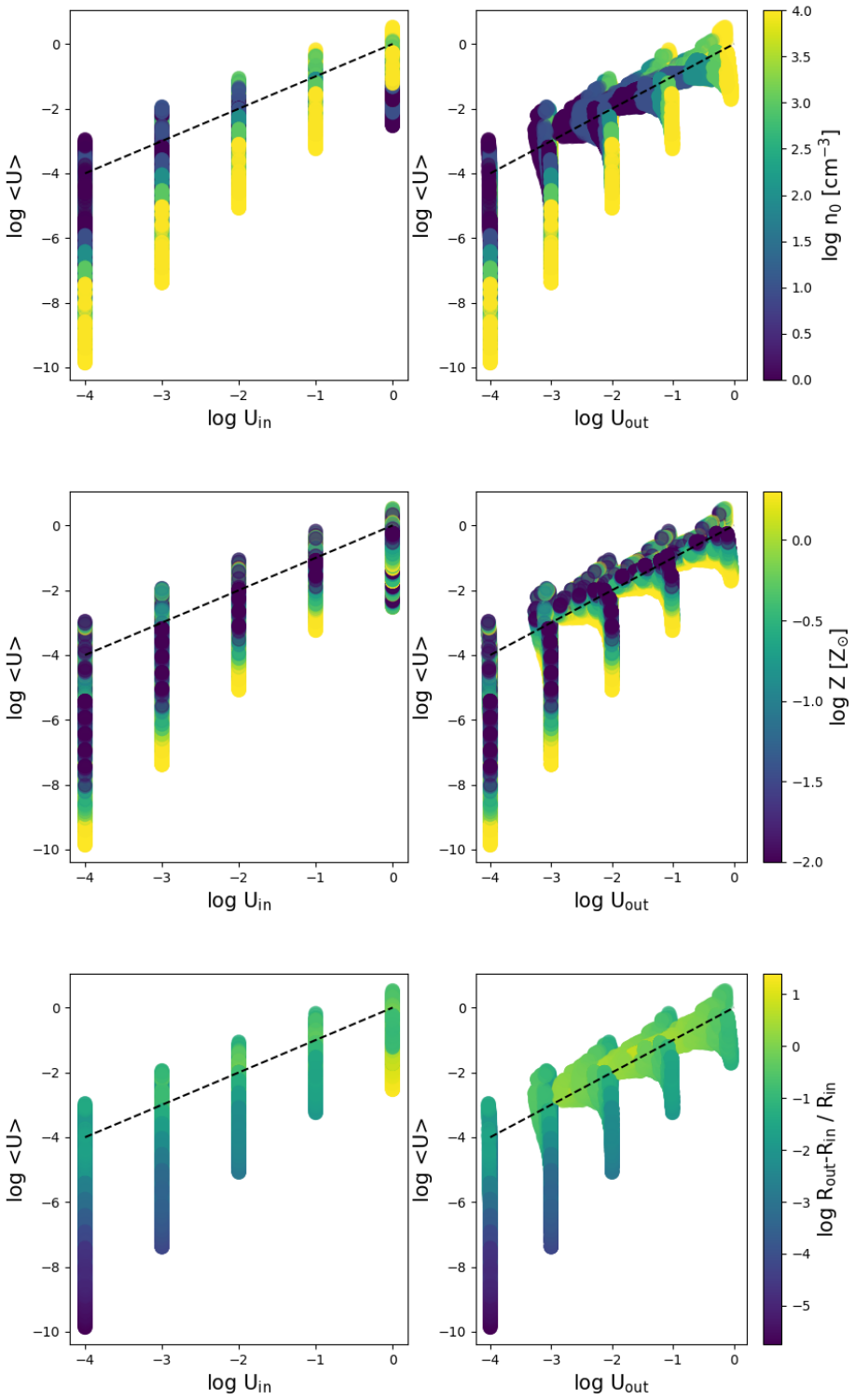

Figure 8.9: Evolution of the volume-averaged ionization parameter $<U>$ as a function of $U_{\rm in}$ and $U_{\rm out}$ in the model grid. The color code represents the initial hydrogen density n_0 (roughly constant throughout the H II region; first panel), the metallicity Z (second panel), and the thickness of the models relative to their inner radius (third panel). We show only radiation-bounded models (i.e., cut=1). The dotted line shows the 1:1 relation.

Among the primary parameters of our grid (See Section 5.2), we find that $<U>$ is mostly sensitive

to the density and metallicity of the H II regions. The two first panels from Figure 8.9 show that variations in terms of metallicity and initial hydrogen density. Both parameters yield variations of the H II region geometries, which translates into a large range of volume-averaged $<U>$ at a fixed $U_{\rm in}$. For $U_{\rm in}$ below 10^{-1}, we find that $<U>$ tends to increase with decreasing n_0 and decreasing metallicity. This can be explained by the fact that models corresponding to lower hydrogen density and lowest metallicity results in thicker H II regions which cover larger volumes. Indeed, we find that the main parameter controlling the $<U>$ vs. $U_{\rm in}$ and $<U>$ vs. $U_{\rm out}$ relations is the thickness of the models relative to the inner radius, as shown on the third panel from Figure 8.9. At fixed $U_{\rm in}$, we find that $<U>$ tends to increases with the ratio of $(R_{\rm out} - R_{\rm in})/R_{\rm in}$.

Interestingly, we note that those trends do not hold for the highest ionization parameter (log $U_{\rm in}$=0) in our grid. For such models, the higher $<U>$ are instead reached for the densest models (n = 10^4 cm^{-3}) corresponding to thick dense shells of gas, while more diffuse models which reach thickness values larger than their inner radius have a lower $<U>$ for the same $U_{\rm in}$.

Dependencies of $<U>$

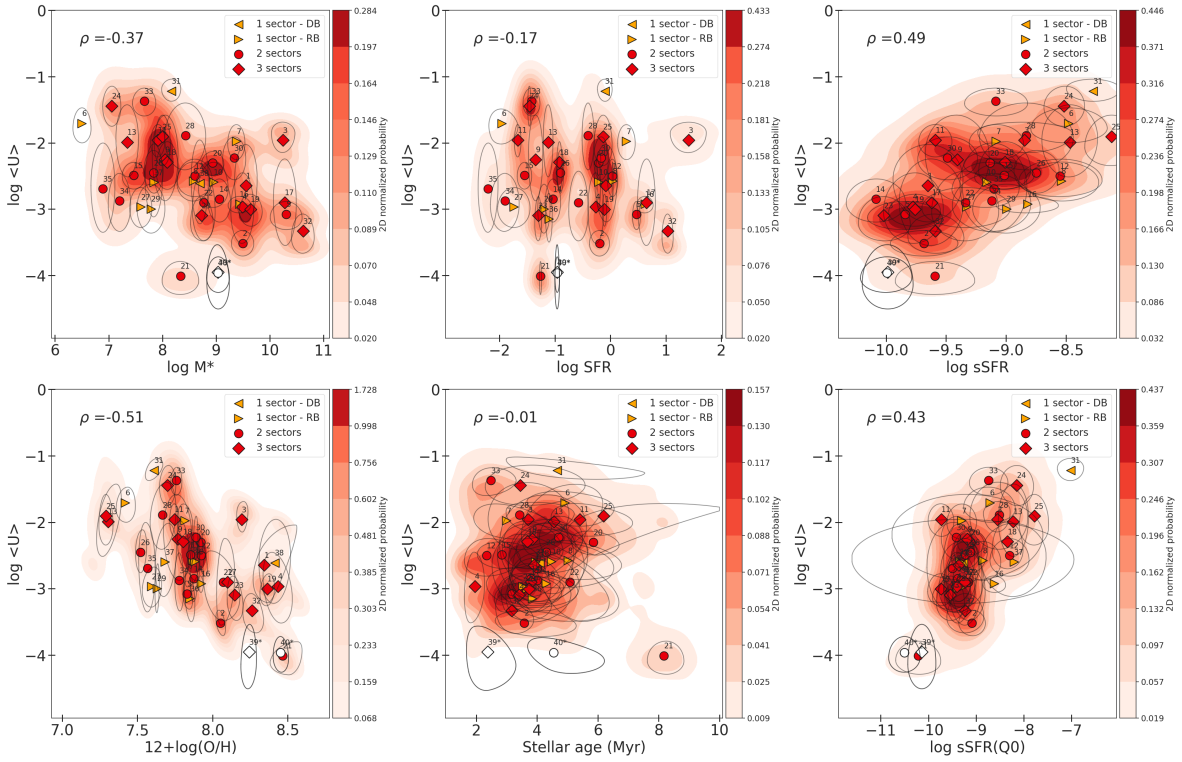

Figure 8.10: Dependencies of $<U>$ with various galactic parameters in the DGS. The first row shows the evolution of $<U>$ with galactic parameters taken from Rémy-Ruyer et al. (2015): M_*, SFR, and sSFR. The second row shows the evolution of $<U>$ with parameters estimated by `MULTIGRIS`: metallicity, stellar age and sSFR(Q(H^0)) as defined in Section 5.3.

We now examine the dependencies of the volume-averaged $<U>$ with galactic parameters in the DGS sample. The global $<U>$ for combined models is calculated as an *intensive* secondary parameter (see Section 5.3.1).

In Figure 8.10 we show the dependencies of $<U>$ with several galactic parameters. We find

an anti-correlation of $<U>$ with the stellar mass and the metallicity. The anti-correlation between ionization parameter and metallicity has been observed in several studies focusing on samples of H II regions and star-forming galaxies (e.g., Pérez-Montero 2014; Morisset et al. 2016; Thomas et al. 2019). Nevertheless, this relation is largely debated with several other studies reporting no correlation, or even a positive correlation with metallicity (see Ji & Yan 2022, and references therein). The positive correlation was, however, mostly observed in resolved studies of H II regions and starburst galaxies, hence this tension might result from spatial resolution effect (Kewley et al. 2019).

The claimed $U - Z$ positive correlation found in the latter studies has also been tentatively explained by the relation between U and SFR (e.g., in Dopita et al. 2014) which we do not observe. On the other hand, we do find a clear correlation of $<U>$ with the sSFR. This correlation is observed both with the measured sSFR from Rémy-Ruyer et al. (2015) and with the predicted sSFR($Q(H^0)$) derived using our estimates for the total number of ionizing photons $Q(H^0)$ (see Section 5.3). This finding is in line with the trend observed in Mingozzi et al. (2020) which find that the correlation between U and sSFR is mostly driven by stellar masses. Additionally, we find no correlation with the age of the burst estimated from our models.

Using $<U>$ as a primary parameter

We have seen in the previous section that the values we derive for the volume-averaged $<U>$ can differ from U_{in} and U_{out} by several orders of magnitude. Since the volume-averaged ionization is more representative of the whole H II region, it could have been an alternative to use $<U>$ instead of U_{in} as an input parameter. Although the code is flexible enough to change easily the primary parameters on which the inference process is performed, replacing U_{in} by $<U>$ creates an irregular and incomplete grid, which introduce converging issues for the MCMC sampler. Hence, the current grid cannot easily be used with an interpolation on $<U>$, without re-running additional models.

In practice, considering a grid that is sampled in U_{in} creates numerous duplicates in $<U>$ ($\sim$ 15% of the grid, once we removed models which do not stop at A_V=10). Those duplicates are predominantly found among models combining the lowest initial densities with strong sources of irradiation (i.e., presence of a luminous X-ray source or gas located near the stellar cluster associated with high ionization parameters). For such duplicated models, the relative variations in terms of predicted emission lines remain limited: it introduces an uncertainty on the flux value below 10% for lines arising from H II regions (e.g., Hα and [O III]), while slightly larger variations, below 20%, are observed for lines arising from neutral gas (e.g., [C II], [O I] and CO).

Most importantly, the variations for the escape fraction of ionizing photons (main observable of the study presented in Chapter 9) is of ~6% (median value for all the set of duplicates in the grid). Hence, we conclude that the presence of duplicates in terms of $<U>$ in our grid introduces an additional uncertainty on our predictions, with a somewhat limited impact on the interpretation. Although we find relatively minor variations of our predictions (below 20% uncertainties), we emphasize that future grids should ideally be sampled using $<U>$ rather than U_{in}.

Chapter 9

Estimating the escape fractions of ionizing photons

In Chapter 8 we presented the first applications to our new code (see Chapter 7) to ISM studies. After testing the code on mock observations directly drawn from combinations of models from the SFGX grid (presented in Chapter 5), we applied `MULTIGRIS` to galaxies from the Dwarf Galaxy Survey (see Section 2.3.1) and derived predictions for observables which are well constrained by observational constraints: the metallicity, the star formation rates, and the ionization parameter. We find that those quantities are not very dependent on the inferred topology and that our results agree well with empirical prescriptions. On the other hand, the escape fraction of ionizing photons (introduced in Section 2.1.3) has multiple dependencies. In particular, it is sensitive to variations in terms of topology (as discussed in Section 8.1) and strongly depends on physical observables that we examined in Section 8.2.

In the current chapter, we now intend to constrain a complex quantity: the Lyman continuum (LyC) escape fraction ($f_{\rm esc}$) from galaxies. In Section 9.1, we present the direct and indirect observational proxies to measure $f_{\rm esc}$ and discuss their possible biases. In Section 9.2, we present the preliminary analysis done to derive a physically meaningful definition $f_{\rm esc}$ in our models. In Section 9.3, we present the paper that gathers our main results, which was accepted for publication in A&A. Finally, we provide a few additional developments on $f_{\rm esc}$ calculated in different energy ranges in Section 9.4.1.

9.1 Context

Direct LyC detection

Measuring the LyC in the UV domain provides a direct estimate of ionizing radiation which escapes along lines of sight toward an ionizing source. The direct detection of LyC-leakage is, however, extremely challenging and has so far been limited to specific ranges of redshift ($z \sim 0.1 - 4$). In particular, this method is extremely sensitive to the absorption of UV photons by the surrounding neutral intergalactic medium (IGM). Direct detections of leaking LyC radiation are not possible – or

extremely unlikely – above $z \sim 4$ due to the absorption by neutral hydrogen in the IGM, preventing any direct observation of potentially LyC-leaking galaxies directly within the epoch of reionization. Another important caveat is the dependency on galactic orientation; while direct detections accurately predict the number of photons escaping along a given line of sight, the value may not be representative of the global galactic value, especially if the Interstellar Medium (ISM) is inhomogeneous. Important variations are expected for different orientations (e.g., Mauerhofer et al. 2021), which complicates the interpretation.

Regardless of the uncertainties about the exact value of f_{esc}, direct measurements have allowed to unambiguously identify samples of LyC-leaking galaxies at $z \lesssim 0.5$ (e.g., Bergvall et al. 2006; Leitet et al. 2013; Borthakur et al. 2014; Leitherer et al. 2016; Izotov et al. 2016a,b, 2018a,b; Wang et al. 2019; Izotov et al. 2021c; Wang et al. 2021; Flury et al. 2022a,b), and at $z \sim 2-3$ (e.g., Vanzella et al. 2015, 2016, 2018, 2020; Shapley et al. 2016; de Barros et al. 2016; Steidel et al. 2018; Bian et al. 2017; Fletcher et al. 2019; Nakajima et al. 2020; Rivera-Thorsen et al. 2019; Pahl et al. 2021, 2022; Begley et al. 2022) with observed f_{esc}(LyC) up to 72%. Between $z \approx 0.5$ and 2, only one LyC-leaking galaxy was detected at $z \approx 1.4$ with the Astrosat satellite (Saha et al. 2020).

As mentioned earlier, such detections are impossible to perform at high-redshift ($z \gtrsim 4$) due to IGM absorption. They are also very challenging at low redshifts, despite the higher photon rates, because of instrument (e.g., HST/COS) sensitivity in the rest-frame UV. The few detections available correspond to remarkably small f_{esc} values in the local universe and only one galaxy in our sample has a LyC detection (Haro 11, f_{esc}=3.2%, Leitet et al. 2013). Similarly, other observations of a few galaxies in the Local Universe detect little to no LyC leakage (e.g., Bergvall et al. 2006; Leitet et al. 2013; Borthakur et al. 2014; Leitherer et al. 2016).

When no direct LyC detection is possible, indirect proxies allow estimating f_{esc} based on different methods. All those methods consist in probing the surrounding gas to assess its porosity to photons. Possible proxies include the analysis of the ISM structure, either in emission or absorption, based on continuum measurements and emission/absorption lines. They may also rely on the analysis of the gas kinematics. The indirect proxies differ in the case of resolved observations in the local universe (see Section 9.1) or in the case of unresolved observation near and far (see Section 9.1).

Indirect proxies for resolved observations

Indirect methods, whether applied to local objects based on the morphological classification of H II regions (Pellegrini et al. 2012), the mapping of the ionization parameter (e.g., Zastrow et al. 2011, 2013), on the estimation of the intrinsic ionizing photon production rate from resolved stars (Choi et al. 2020), or on the ionized gas kinematics (Eggen et al. 2021), are all suggestive of qualitatively large escape fractions of ionizing photons from H II regions.

While ionization mapping techniques and gas kinematics allow disentangling different components of the ISM (spatially or in velocity), the morphological classification of H II regions and analysis of resolved stars can serve to derive estimates of the photon production rates and compare to what is absorbed. The latter methods have hence provided quantitative estimates on the number of escaping photons, while the other have remained qualitative.

Additionally, results from Polles et al. (2019) on the local starbursting galaxy IC 10 indicate that

the porosity depends on the spatial scale, with most clouds being density-bounded at small scales (~60 to 200pc), while larger regions become more and more radiation-bounded when reaching galactic scales. This result highlights the complexity of the ISM in which the energy produced by feedback is deposited at various spatial scales and over different dynamical timescales, hence producing a highly inhomogeneous internal structure. While indirect methods relying on spatially resolved information can be applied in the local universe, they are not applicable in more distant unresolved galaxies.

Indirect proxies for unresolved observations

When no spatial information is available, several other proxies have been successfully applied to estimate global escape fractions of ionizing photons. First, far-UV absorption lines can serve as a promising proxy to set an upper limit on the amount of escaping photons (e.g., Reddy et al. 2016; Gazagnes et al. 2018, 2020; Chisholm et al. 2018; Saldana-Lopez et al. 2022). This method relies on saturated (i.e., optical depth $\tau >> 1$ in the line center) absorption lines to derive a covering fraction of H I gas, which tightly constrain the maximum number of ionizing photons unabsorbed by the neutral gas. The use of absorption lines remains somewhat sensitive to line-of-sight variations but has the advantage of accounting for the total H I gas reservoir, which is not always seen in emission (e.g., Gazagnes et al. 2020).

While the previously mentioned studies have performed absorption lines studies along lines of sight oriented toward stellar ionizing sources, recent studies have used similar techniques to exploit the afterglows that follow Gamma Ray Burst (GRB) emission in order to probe the properties of the surrounding ISM (Tanvir et al. 2019; Vielfaure et al. 2020). This method is especially interesting to derive a direct determination of the $(f_{1500}/f_{900})^{\rm int}$ ratio due to the intrinsically featureless (power-law) spectra of GRB afterglows. GRBs are hence promising tools to search for LyC-leakage in very faint galaxies at $z \sim 2-4$.

Another approach consists in looking at far-UV color selection diagrams (e.g., Vanzella et al. 2015; Naidu et al. 2017). This method overcomes the lack of spectroscopic detection of LyC by using photometric data from the *Hubble* Deep UV imaging survey. They used filters that cover the redshifted range of LyC in galaxies at $z \geq 2$, hence providing direct images of the LyC photons. These UV data were combined with archival HST images at longer wavelengths and spectroscopic redshifts from the literature to search for potential LyC candidates over the full redshift range $z \sim 1.9-4$.

Finally, emission lines arising from different phases of the ISM can also serve as useful probes of the global porosity to ionizing photons. In the UV domain, the Lyman alpha (Lyα) line profile has proven to be an interesting proxy, with the presence of double-peaked or triple-peaked profiles being associated with LyC and Lyα leakage (e.g., Verhamme et al. 2015; Henry et al. 2015; Verhamme et al. 2017; Izotov et al. 2020; Maji et al. 2022). In particular, the velocity separation of the blue peak and red peak has been shown to strongly anti-correlate with the measured escape fraction of Lyα photons. Recent studies have also highlighted the interest of using proxies such as weak helium lines (Izotov et al. 2017) or the Mg II $\lambda\lambda$2796,2803Å doublet (e.g., Henry et al. 2018; Chisholm et al. 2020; Xu et al. 2022; Katz et al. 2022) to probe the low-density lines of sight necessary for LyC-photons to escape.

In the optical range, several proxies can serve as indicators of LyC-leakage. In particular, line

ratios involving ions with different ionization potentials, produced at different depths, have been proposed as indicators to discriminate between radiation-bounded and density-bounded H II regions (defined in Section 1.2.1). Radiation-bounded regions correspond to ionized spheres delimited by their Strömgren radius set by the equilibrium between production of photons by stars and ionization of the surrounding gas. Density-bounded regions are instead delimited by the lack of matter, which sets their outer radius before the Strömgren radius (see Section 1.2.1). Hence, they allow part of the LyC-photons produced by stars to escape from H II regions. The oxygen line ratio [O III]λ 5007Å/[O II] $\lambda\lambda$3726, 3728Å (O32) proposed by Jaskot & Oey (2013) and Nakajima & Ouchi (2014) was successfully used to select LyC-leaking candidates, but no strong correlation was found with the measured $f_{\rm esc}$ values (see Izotov et al. 2018b; Naidu et al. 2018; Bassett et al. 2019; Nakajima et al. 2020, and discussions therein). Based on a similar idea, the lack of emission from ions with low ionization potentials like [S II] $\lambda\lambda$6716,6731Å has also been proposed to target LyC-leaking candidates (Wang et al. 2019, 2021).

Using integrated emission lines in the optical and infrared domain palliates viewing angle dependencies which limit the use of absorption lines or Lyα proxies. It is a promising method to make the most of observations with current ground-based facilities (e.g., Keck, Gemini, LBT, ALMA) and the upcoming JWST data that will grant access to many spectroscopic tracers up to redshift above 7 (e.g., Schaerer et al. 2022). However, as shown in Ramambason et al. (2020), reconciling the different tracers arising from a multiphase ISM and the escape fraction observed in known leaking galaxies requires a complex modeling step, which can be solved by introducing multicomponent models.

We built on this previous analysis presented in Ramambason et al. (2020) and on other previous works using topological models (presented in Section 3.2.1) to develop a new modeling approach based on the combination of several ISM component (see Chapters 4 and 7). In the following section, we present a first illustration of the use of IR emission lines to constrain the escape fraction of photons in a sample of local star-forming dwarf galaxies.

9.2 Preliminary analysis

9.2.1 Defining the cuts

As described in Chapter 5, all the `Cloudy` models are run until they reach an A_V of 10 (or until their electronic temperature T_e drops below 10 K). Each model is then cut in 17 submodels which correspond to shells of gas at varying depths, as described in Section 5.2.3. This step is essential in our modeling approach since it allows us to account for density-bounded models, corresponding to models stopping before they reach their .

Because of the multiphase nature of our models, which include a PDR and molecular zone, we also want to define a "cut" parameter that can consistently be defined throughout the different phases. Hence, proxies such as the Hβ fraction (used in, e.g., Stasińska et al. 2015; Asari et al. 2016; Ramambason et al. 2020) or the zero-age optical depth to LyC photons (τ_λ) (used in, e.g., Leboutteiller et al. 2017; Plat et al. 2019) could not be easily adapted to our use case. Several physical quantities could have been used to define the cut. In particular, we stored as post-processed observables the following quantities which could have been used instead of the "cut" parameter:

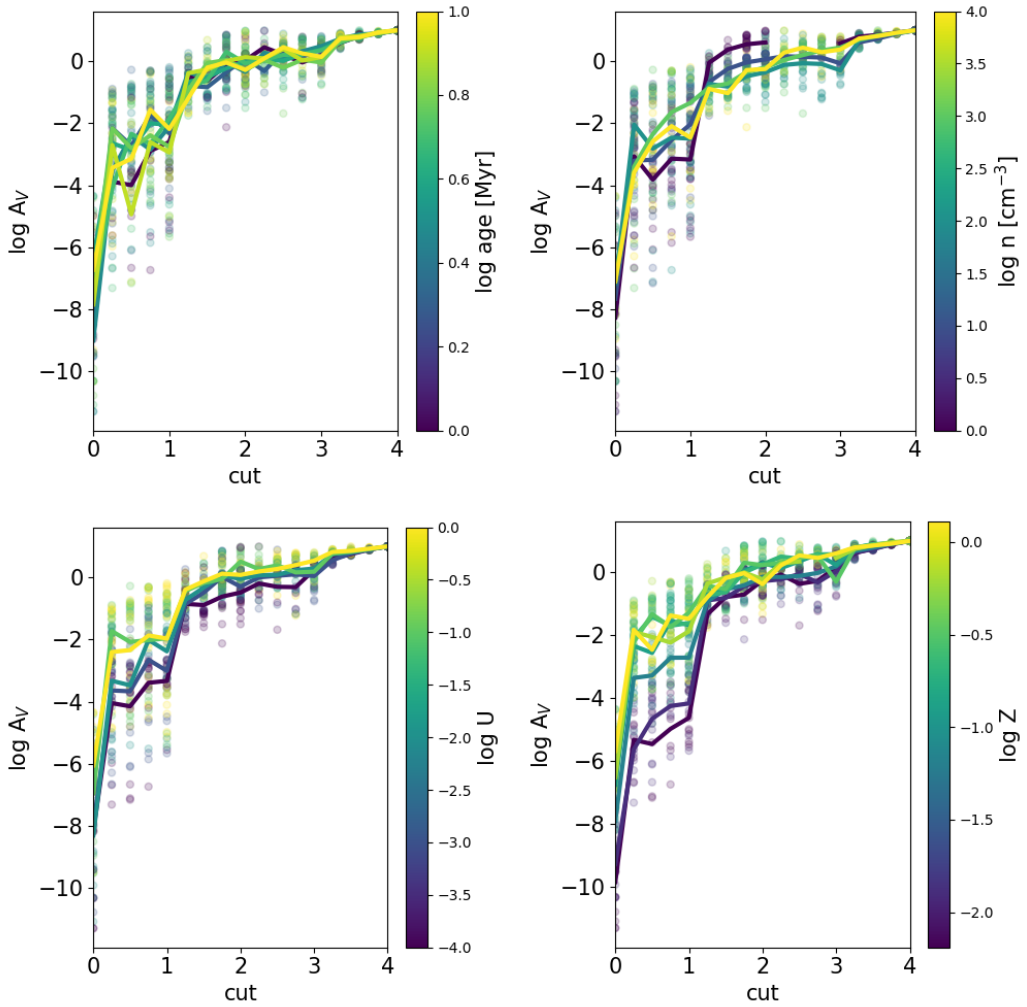

Figure 9.1: Variation of the visual extinction A_V corresponding to different cuts. The colored dots correspond to a random sample of 1000 models taken from the SFGX grid. The plain colored lines show the mean values of all models, grouped by bin of values corresponding to the color code.

- **Av_stop**: the visual extinction at the outer radius of each model.
- **depth_stop**: the physical depth at the outer radius in cm.
- **Nh_stop**: the hydrogen column density at the outer radius in cm^{-2}.
- **G$_0$**: the incident radiation field at the outer radius in Habbing unit.

Nevertheless, one specificity of our grid is that it spans a large range of physical conditions which strongly affect the shape and thickness of the resulting shell of gas, making it difficult to define use the previously mentioned parameters with fixed bins to describe all the models. Instead, we opted for a definition based on transitions between different ionization state of abundant elements, namely hydrogen and carbon, which can be used in all the models, regardless of their metallicity.

While several transitions could be considered to define the cuts, one needs to ensure that the cuts are always sorted in the same order in all the models. We check that this condition is respected for the five cuts defined in Section 5.2.3. This sorting issue prevents us, however, from using for example

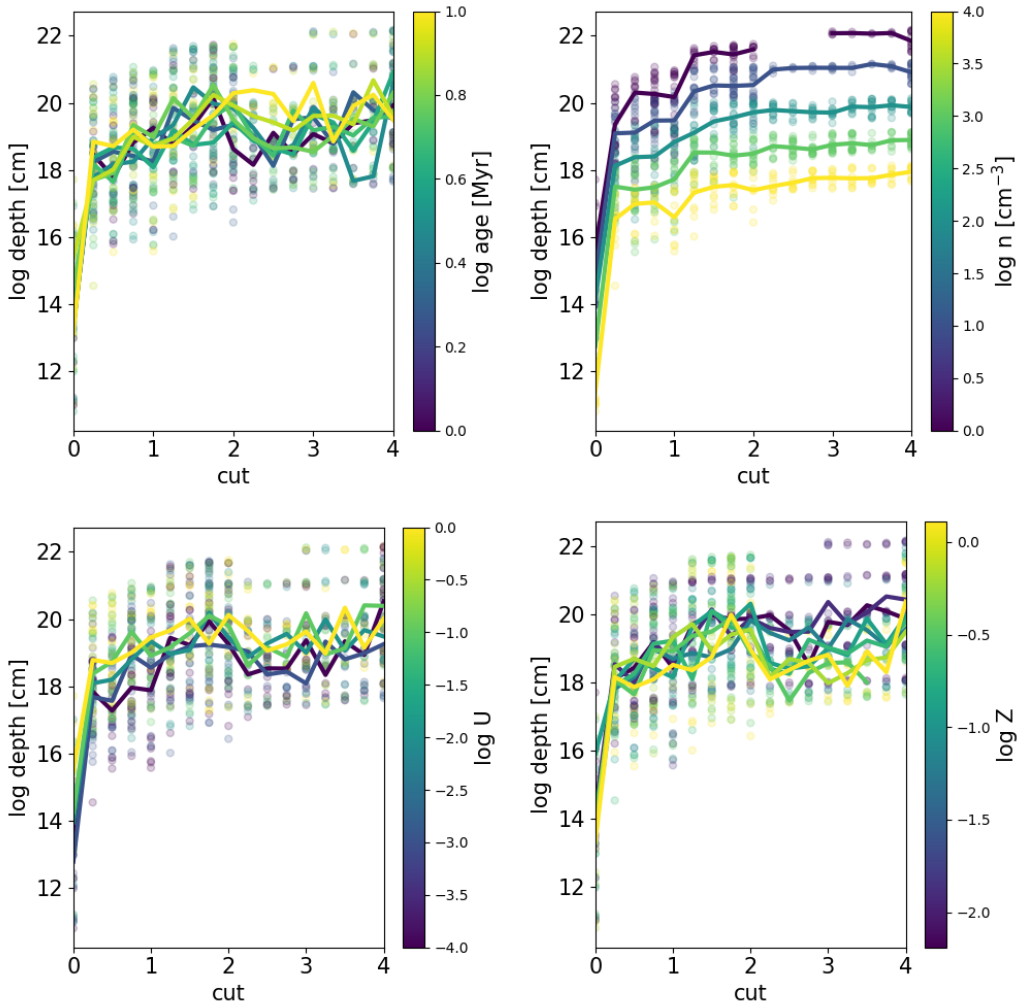

Figure 9.2: Variation of the physical depth corresponding to different cuts.

the C^+-C^0 transition which falls either before or after the H^0–H_2 transition, depending on the model. We recall here the five main cuts which were introduced in Section 5.2.3:

- cut=0: inner radius (illuminated edge).
- cut=1: H^+–H^0 transition (i.e., ionization front; electron fraction 0.5).
- cut=2: H^0–H_2 transition (photodissociation front; molecular gas fraction 0.5).
- cut=3: C^0–CO transition (photodissociation front; molecular gas fraction 0.5).
- cut=4: full depth corresponding to stopping criterion $A_V = 10$.

With this definition, radiation-bounded models have cuts ≥ 1 while density-bounded models (introduced in Section 1.2.1) correspond to cut < 1. In order to refine the cut sampling, we define fractional values in between the main cuts. While a first approach consisted in using abundance profiles (e.g., H^+, H^0) to define the cut at a given percentage of the abundance, this definition has proved to be difficult to use due to the sharp variation of some specific elements. For example, using fractional values based on H^+ profile resulted in sampling only a narrow region near the ionization front

where H^+ abundance drops steeply. Instead, fractional cut values (with a step of 0.25) are considered by sampling the extinction A_V profile, which allows a better coverage of the cloud depth.

In Figures 9.1 and 9.2 we show the variations of the visual extinction A_V and physical depth of the models, as a function of the cuts. We color code the variation with the main physical parameters controlling the geometry of the resulting models: the stellar age, the initial density, the ionization parameter ($U_{\rm in}$), and the metallicity. We note that the correspondence between the extinction and the cut parameter is mostly sensitive to variations in the metallicity and ionization parameter. In particular, the chemical transitions corresponding to the ionization front, dissociation front, and the C^0-CO transition occur at smaller visual extinctions in low-metallicity and low-U models.

In Figure 9.2, we see that the U-dependency of the cut-A_V relation corresponds to a physical depth effect: models with higher U, in which the gas is set closer to the stars (smaller inner radius), result in marginally thicker models than the models with lower U in which the gas is placed further away. The higher-A_V found in high-U models hence comes from the cumulated effect of dust extinction integrated over a larger slab of gas than for low-U models.

On the other hand, the metallicity-dependency of the cut-A_V relation does not correspond to a physical depth effect, since low-metallicity models tend to produce thicker models than the high metallicity ones. Hence, the enhanced visual extinction in high-metallicity model comes from the dust-to-gas mass ratio increase with metallicity (as shown in Figure 5.6). At fixed cut values, models with higher metallicity correspond to thinner geometries but with a higher dust content, which translates into larger A_V.

Finally, we note that the physical depth is primarily driven by the initial density of our models and spans up to 5 order of magnitude. The densest models correspond to rather thin shells ($\sim 10^{17}$ cm) of gas that quickly reach A_V=10 while the lowest density models reach much larger physical depths ($\sim 10^{22}$ cm).

9.2.2 Defining a proxy for the escape fraction

Although direct observations of LyC-escaping photons are challenging (as discussed in Section 9.1), several proxies can be used to provide an estimate on the escape fraction of ionizing photons, $f_{\rm esc}$. The different observational proxies that can be used do not exactly probe the same escape fractions. In particular, comparing the predicted escape fractions based on observational proxies with the results obtained with models or simulation is a difficult task which requires a careful investigation of the definition used for $f_{\rm esc}$.

In our analysis, we have focused on deriving an *intrinsic* escape fraction from the `Cloudy` models which accounts for all the photons escaping from density-bounded regions with an energy greater than 1 Ryd. The definition used in our analysis was refined over the time to obtain a rather flexible definition of the escape fraction of photons from modeled H II regions. We present here the different definitions that we considered and motivate the choice made for the first applications.

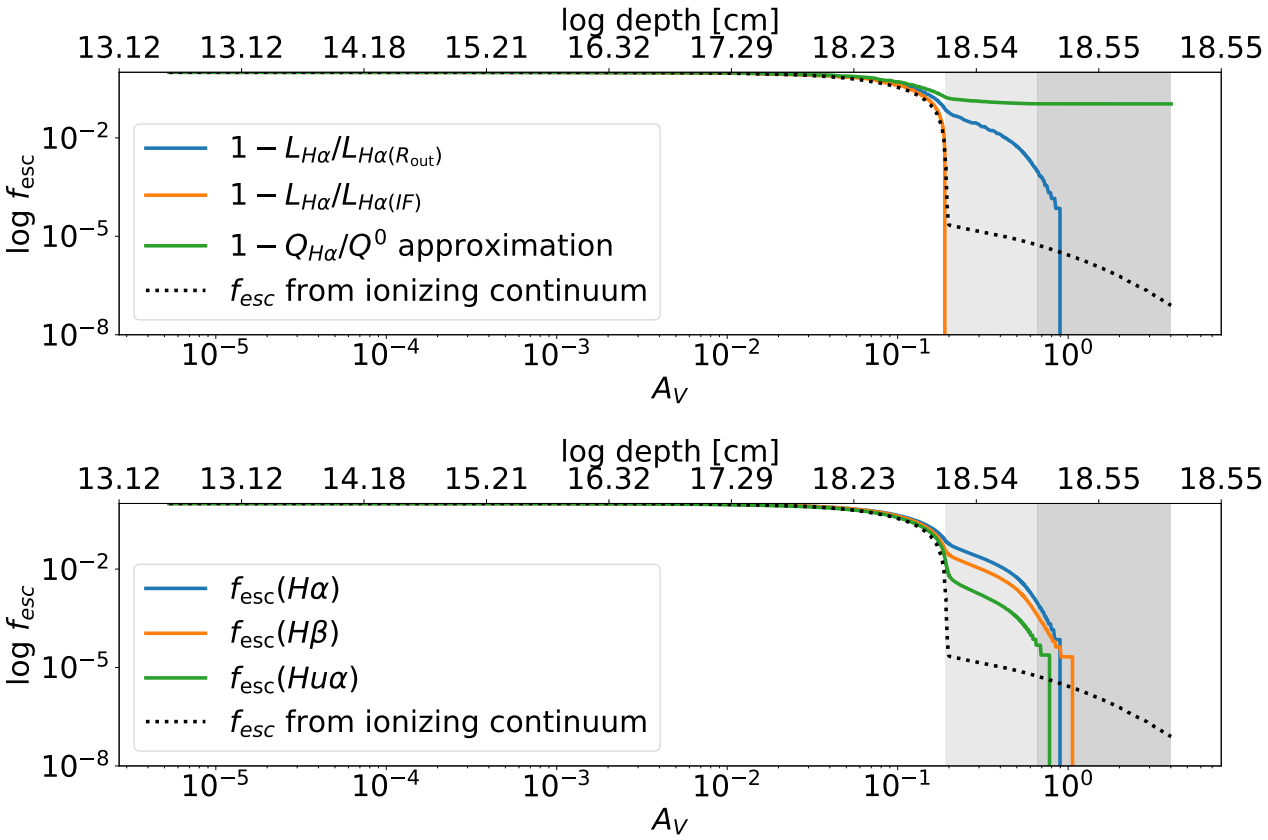

Figure 9.3: Comparison of different proxies to estimate $f_{\rm esc}$. **Top**: Evolution of $f_{\rm esc}$ vs. A_V for different proxies. The tests are run for a solar-metallicity model, without any X-ray source, and run up to $N_{\rm H}$=10^{22} cm^{-3}. The shaded areas mark the PDR location (light-gray) and the H_2 molecular zone (dark-gray). **Bottom**: Comparison of the escape fractions estimated from different hydrogen recombination lines: Hα, Hβ and Huα.

$f_{\rm esc}$ from H recombination lines

A first simple approach to estimate the escape fraction from a given model is to use one hydrogen recombination line which, at first order, is directly proportional to the number of ionizing photons $Q(H^0)$ (see Osterbrock & Ferland 2006).

Because the emission of hydrogen recombination lines is roughly constant in the H II region, a simple estimate of the number of ionizing photons escaping from density bounded-regions can be computed by comparing the luminosity produced by the density-bounded models (i.e., cut < 1 in the formalism presented in Section 5.2.3) and the luminosity of the same model when stopped at the ionization front (cut=1). The most convenient tracers are hence hydrogen lines from series easily observed such as the Balmer series (e.g., Hα, Hβ) in the optical or the Brackett, Pfund, and Humphreys series (e.g., Brα, Pfα, Huα) in the infrared domain. The escape fraction can then be approximated as follows:

$$f_{\rm esc}(r) \approx 1 - \frac{L_{\rm H_X}(r)}{L^{\rm IF}_{\rm H_X}} \tag{9.1}$$

where H_X is a hydrogen recombination line (such as Hα, Hβ), $L^{IF}_{H_X}$ is the total H_X luminosity produced by a radiation-bounded model truncated at its and $L_{H_X}(r)$ is the H_X luminosity produced between the inner radius and a given radius r before the ionization front. This definition is similar to the one based on the Hβ fraction, which we used in Ramambason et al. (2020).

Although this proxy can be computed in a very simple way, it introduces a dependency related to the definition adopted to locate the ionization front. This effect is shown in the first panel from Figure 9.3 in which the blue and orange lines both represent f_{esc} based on the intrinsic Hα emission. The orange line corresponds to estimates based on Equation 9.1 which uses the Hα value at the ionization front[1] to normalize the emission, while the blue curve uses the total Hα value (measured by integrated the cumulative Hα up to the outer radius) to normalize the emission. We note that the latter definition delays the drop of f_{esc} after the PDR zone. This is due to the fact that a small fraction of Hα is produced after the IF due to a residual fraction of free electrons in the PDR.

While the associated Hα emission is small, this effect translates into f_{esc}(Hα) estimates of the order of the few percent measured at the IF. In the low-metallicity models, this residual f_{esc}(Hα) measured at the IF can reach up to a few 10%. Although the residual fraction of ionized hydrogen is enough to power intrinsic Hα emission, the PDR is largely dominated by neutral hydrogen, which completely prevents any LyC-leakage beyond the IF. Subsequently, we conclude that the use of hydrogen recombination lines as proxies to estimate f_{esc} is valid only in pure H II region models. Multiphase models including a PDR component might produce additional Hα emission which is not associated with the H II region, i.e., corresponding to a partially-ionized neutral gas, and bias the f_{esc}(Hα) estimates that are derived. Although this effect is small for models with solar-like metallicity, this additional contribution can reach ~10% in the low-metallicity models (~1/10 $Z_\odot$) which becomes problematic and motivates the use of another proxy.

Finally, using hydrogen lines as proxies for f_{esc} is also subject to other caveats:

- The hydrogen recombination lines are not sensitive to high energy photons which can propagate in the PDR and molecular clouds and eventually contribute to the global f_{esc} (see Section 9.4.1).
- The estimate from Equation 9.1 implicitly assumes a constant hydrogen recombination coefficient throughout the whole H II region, while the latter varies with electronic density and temperature.
- Different recombination lines may be sensitive to opacity effects, especially for large metallicity runs; the results obtained by using Hα, Hβ, or Huα yield slight variations in the escape fraction estimate, as shown in the bottom panel from Figure 9.3.

$f_{esc}(Q(\mathbf{H}\alpha))$

Another approach was used in Niederhofer et al. (2016) where they use the following proxy to estimate escape fractions:

$$f_{esc}(Q(\mathrm{H}\alpha)) \approx \frac{Q(\mathrm{H}^0) - Q(\mathrm{H}\alpha)}{Q(\mathrm{H}^0)} \tag{9.2}$$

[1]The ionization front is defined at the location where $n(H^+)$=$n(H^0)$ in the models.

where $Q(H^0)$ is the number of ionizing photons produced by the incident source and $Q(H\alpha)$ is an estimate of the number of photons consumed throughout the H II region to produce Hα luminosity.

Considering that recombination coefficients show relatively small variations within typical H II regions conditions, a classical assumption is to a fixed factor calculated based on typical averaged conditions. For example, in Kennicutt et al. (1995) and Kennicutt (1998) they derive the following equation for T_e= 10^4 K and n_e=100 cm^{-3}):

$$Q(H\alpha) = 7.31 \times 10^{11} L_{H\alpha} \tag{9.3}$$

However, Niederhofer et al. (2016) stress that this proxy is highly sensitive to the stellar types and ages and yield very uncertain results. Similarly, in the test presented in Figure 9.3, the conversion of the total Hα intrinsic emission using Equation 9.3 does not allow us to recover the full $Q(H^0)$ from the model. Instead, the $Q(H\alpha)$ we derive with this approximation remains lower than the actual $Q(H^0)$ leading the estimated f_{esc}(Q(Hα)) (green curve from panel 1) to reach a plateau instead reaching zero after the IF. Similarly to Niederhofer et al. (2016), we conclude that this approximation should not be used to recover the total $Q(H^0)$ as it yields large uncertainties on the total number of ionizing photons produced by the incident source. Instead, it is crucial to use a method which accounts for the local physical conditions at each depth in the cloud and the associated variations of the recombination coefficients.

$f_{esc}(Q_E)$

`Cloudy` allows saving, for every zone, the number of photons reaching a given depth per frequency bin (in units of photons cm^{-2} s^{-1}) as a function of energy, for all energies above 1 Ryd. The number of ionizing photons reaching a radius r can then be expressed as:

$$Q_E(r) = 4\pi r^2 \int_{h\nu=E_1}^{E_2} \frac{F_\nu(r)}{h\nu} d\nu \tag{9.4}$$

where r is the radius (in cm) and $F_\nu/h\nu$ the flux per frequency bin in s^{-1} cm^{-2}.

The associated escape fraction of photons with energy between E_1 and E_2 can then be defined as:

$$f_{esc}(Q_E) \approx \frac{Q_E(r)}{Q_E(R_{in})} \tag{9.5}$$

where $Q_E(r)$ (in s^{-1}) is the number of photons that reach the radius r per time unit and $Q_E(R_{in})$ is the number of photons produced by the ionizing source per time unit, i.e., the number of photons reaching the edge of the cloud at the inner radius ($Q_E(R_{in})$=$Q(H^0)$ for E=[1 Ryd,∞]). This definition allows us to track the number of ionizing photons escaping from the H II region with a given energy in a flexible way and accounting for the actual local ionizing fluxes as saved at each depth of the model.

However, this definition relies on the predicted number of photons reaching the gas at a given depth per time unit. Although one would expect this quantity to monotonically decrease with increasing depth, as photons are being destroyed/absorbed by the gas and dust, this is actually not the case

when accounting for the nebular contribution (see discussion from Section 8.2.3). Indeed, for models including X-ray sources, high energy photons can be absorbed and re-emitted by the two-photons process, which causes $Q_{\rm E}$ to increase locally. While in most models the contribution from the nebular continuum is negligible, as its photons are completely absorbed within the nebula (i.e., on-the-spot approximation; Osterbrock & Ferland 2006), this contribution can become important, especially in low-metallicity environments exposed to hard radiation fields (Inoue 2010).

$f_{\rm esc}(L_{\rm E})$

In our models, we find pathological cases (mostly the densest models which include X-ray sources) where $Q_{\rm E}$ locally exceeds $Q({\rm H}^0)$ leading to escape fraction above unity. Although those cases are rare, we consider yet another definition based on the luminosity ratio instead of number of photons in order to use an observable which is conserved throughout the cloud. We hence define:

$$L_{\rm E}(r) = 4\pi r^2 \int_{h\nu=E_1}^{E_2} F_\nu(r) d\nu \tag{9.6}$$

where r is the radius (in cm) and F_ν the energy flux erg s^{-1} cm^{-2}.

The associated escape fraction of photons with energy between E_1 and E_2 is then defined as:

$$f_{\rm esc}(L_E) \approx \frac{L_{\rm E}(r)}{L_{\rm E}(R_{\rm in})} \tag{9.7}$$

where $L_E(r)$ (in erg) is the luminosity powered by photons that reach the radius r within the energy range E and $L_E(R_{\rm in})$ is the luminosity within the same energy range E produced by the ionizing source and reaching the edge of the cloud at radius 0. This definition is analog to the one used in Bianchi et al. (2018) to define $f_{\rm abs}$, the fraction of bolometric luminosity absorbed by dust. Similarly, our definition quantifies the fraction of bolometric luminosity which in *unabsorbed* by metals and dust and escape at a given depth in our models. This will be the definition adopted in the analysis presented in Chapter 9.

9.2.3 Summary and definition of $f_{\rm esc,HII}$

We have seen in Section 9.2.2 that the definition of $f_{\rm esc}$ is a crucial point which may lead to substantial changes in the values and trends that we derive. Our code currently computes several $f_{\rm esc}$ values based on the above definitions. Meanwhile, in the following applications (Chapter 9 and 10), we chose to use the last definition presented in Section 9.2.2, which accounts for the energy range of the escaping photons and define the escape fraction based on a luminosity ratio instead of considering numbers of photons. This definition is equivalent to the ratio of the total observed emission below 912Å and the intrinsic emission below 912Å. It consistently accounts for the absorption of photons in the gas and dust, as modeled by `Cloudy`. In that sense, this definition is as closed as possible to the "true" intrinsic escape fraction of photons in our models.

Nevertheless, we stress that this definition reflects the underlying modeling assumption. In particular, the topological models (see Section 3.2) that we use do not account for the diffuse phase of the ISM. To avoid possible confusions, we will refer to our estimates as the escape fraction from H II regions ($f_{\rm esc,HII}$). This denomination highlights one of the main hypothesis in our modeling scheme, which is that a galaxy can be modeled a sum of H II regions. In principle, this $f_{\rm esc,HII}$ definition also accounts for the effects of considering a PDR and a molecular phase, directly connected to the H II regions. However, they do not account for potential transmission effects through a more diffuse component in which the star-forming region is embedded. Linking our predicted $f_{\rm esc,HII}$ to a global $f_{\rm esc}$ at galactic scale would need further improvements, in particular to model and constrain that the diffuse component of the ISM which we discuss in Section 11.1.1.

9.2.4 Variation of $f_{\rm esc,HII}$ in the grid

Ultimately, we want to ensure that our cut definition allows a smooth sampling of the escape fraction of ionizing photons from H II regions. Before studying the escape fraction inferred for combined models and comparing it to observations, we explore the variation of $f_{\rm esc,HII}$ in our grid.

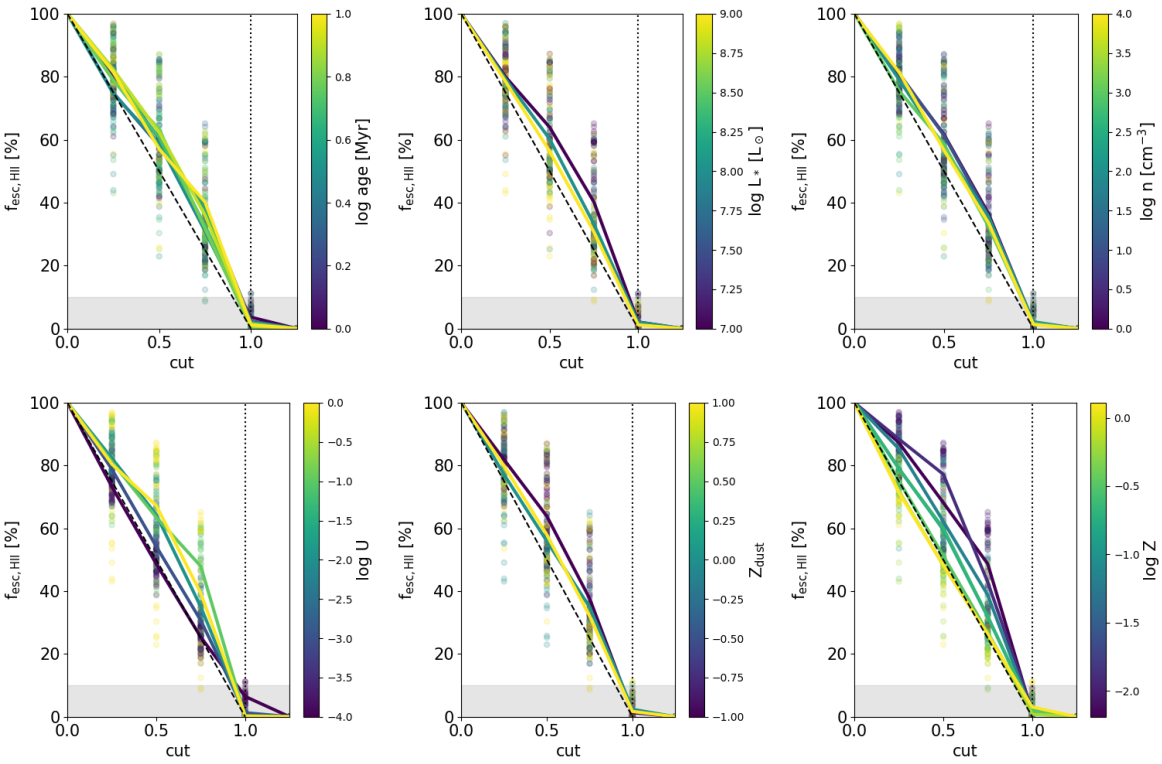

Figure 9.4: Variation of the escape fraction $f_{\rm esc,HII}$. The colored dots correspond to a random sample of 1000 models taken from the SFGX grid. The plain colored line show the mean values of models corresponding to the color code. The gray shaded area marks the threshold of $f_{\rm esc,HII}$=10% and the vertical dotted line is the position of the ionization front. Most models lie above $f_{\rm esc,HII}$=(1-cut) line, shown by the black dotted line between 0 and 1.

In Figure 9.4, we show the variations of $f_{\rm esc,HII}$ with the cut parameter for a subsample of models without any X-ray source. As one might expect, $f_{\rm esc,HII}$ varies from 100% at the illuminated edge (cut=0) down to 0% at the ionization front (cut=1). Because we are using the full ionizing continuum

to define $f_{\rm esc,HII}$, the values after the ionization front are not exactly zero, but they are down below 1% for all models without X-ray sources.

We note that the $f_{\rm esc,HII}$ values corresponding to a given cut are scattered around a roughly linear relation. The $f_{\rm esc,HII}$-cut relation is not very sensitive to the stellar age, total stellar luminosity, and dust-to-gas mass ratio $Z_{\rm dust}$. We note that our $Z_{\rm dust}$ parameter represents the spread around the relation from Galliano et al. (2021) (see Section 5.2) and does not correspond directly to the dust-to-gas mass ratio value. The absolute value can be calculated using the relations from Galliano et al. (2021) and scales with metallicity. Hence, the $f_{\rm esc,HII}$ values also exhibit a dependency on the dust-to-gas mass ratio, which is directly linked to the metallicity dependency.

However, the scatter seems to be mostly driven by the density, the metallicity, and the ionization parameter. Specifically, models with the lowest density, highest ionization parameters, and lowest metallicity exhibit a shallower drop of $f_{\rm esc,HII}$ with the cut parameter. This can be understood as the ionizing continuum is less quickly absorbed in low density and low-metallicity models, hence UV photons penetrates deeper in the cloud and are absorbed only at larger cuts.

Finally, we note that our binning remains somewhat coarse, with a rapid drop from a mean $f_{\rm esc,HII}$ around 20-40% at cut=0.75 going down to 0 at cut=1. This might lead to an overestimation of $f_{\rm esc,HII}$ especially in single-sector models. Possible workarounds that may be used in future works include performing a linear interpolation on the cut parameter as done in LR22 or increasing the number of bins used to sample the cut. We note, however, that in multicomponent models (i.e., at least two sectors), this effect is mitigated as the global escape fractions depends both on the escape fraction of each sector and on the mixing weights (see Equation 5.7) which are *continuous* random variables. Hence, for multicomponent models, the range of global $f_{\rm esc,HII}$ values is always continuously covered.

In Figure 9.5 we show the evolution of $f_{\rm esc,HII}$ as a function of the cut parameter for models which have an X-ray source. For those models, the escape fraction does not reach zero after the ionization front (cut=1) because the additional photons produced by the X-ray source penetrate deeper in the gas compared to models without an X-ray source and can ionize the gas even in the PDRs. We see that a moderate escape fraction of ionizing photons ($\leq$ 10%) is expected even in radiation-bounded models when an X-ray source is present.

Specifically, as shown in the bottom row from Figure 9.5, we find that, although this effect of escaping photons in the PDR is controlled by the two parameters (luminosity and temperature) of the X-ray source, it is mostly sensitive to temperature of the accretion disk ($T_{\rm X}$). This is due to the fact that higher $T_{\rm X}$ implies a relatively larger number of photons with higher energies that can penetrate deeper in the cloud. In particular, models with $T_{\rm X}$=10^7 K reach escape fraction of ~10% on average in the PDR.

Finally, we note that the steepness of the $f_{\rm esc,HII}$-cut relation is controlled by the same parameters as in models without X-ray source (metallicity, density, and ionization parameter). Interestingly, when an X-ray source is present, the total stellar luminosity L_* becomes an important parameter. The models with the lowest bolometric luminosity ($10^7\,L_\odot$) have increased $f_{\rm esc,HII}$ with respect to the highest bolometric luminosity ($10^9\,L_\odot$). This effect is counter-intuitive, since the absolute X-ray luminosities are larger with the highest bolometric luminosities. However, in models with lower bolometric luminosities, the gas is located closer to the incident sources, hence increasing the effect of the X-ray source.

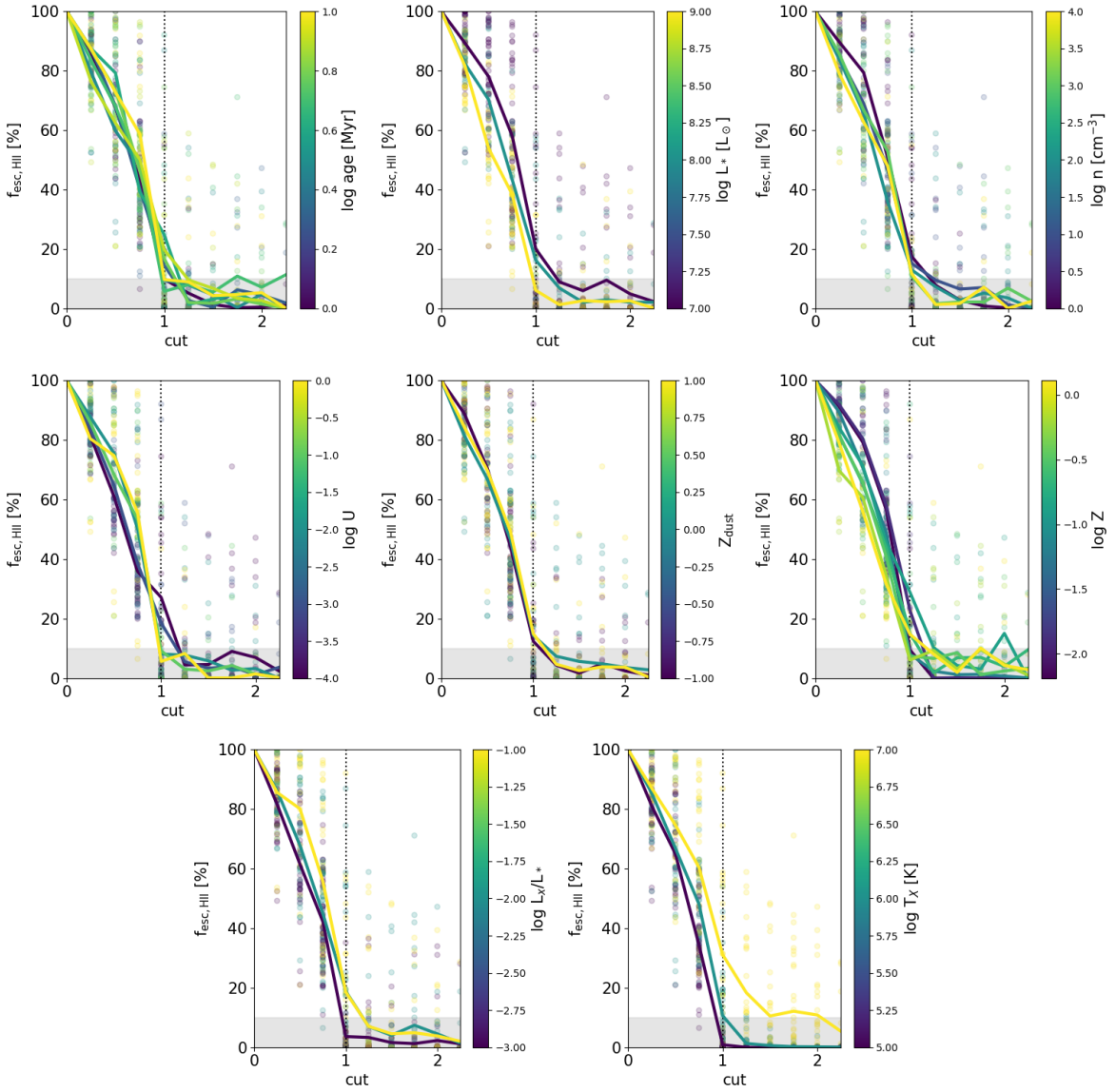

Figure 9.5: Same Figure as 9.4 but for models with X-ray sources.

9.3 Inferring $f_{\mathrm{esc,HII}}$ in the Dwarf Galaxy Survey

As a first application, we apply `MULTIGRIS` to a sample of local ($\lesssim$ 100 Mpc) well-known star forming dwarf galaxies from the Dwarf Galaxy Survey (DGS; Madden et al. 2013). This sample is ideally suited to test our new code, thanks to the wealth of IR tracers detected with the space telescopes *Herschel* and *Spitzer*. The DGS galaxies cover a wide range of physical properties, including high star formation rates and low-metallicities (down to 1/35 $Z_\odot$) which make them interesting targets to study how stars form and interact with the surrounding gas in low-metallicity environments.

This study is a direct continuation of the work from C19 which modeled the DGS galaxies using single- and 2-sector models and constrained their PDR covering factor. They found PDR covering factors below unity for all objects and a correlation between the PDR covering factor and the metallicity. Their results hint at an increased porosity to UV radiation fields of the low-metallicity , which might translate into increased LyC escape fractions. Such results have important implications on

our understanding of the physics of high-redshift primordial galaxies, which may have significantly contributed to the reionization budget.

The following study aims at providing, for the first time, a quantitative estimate of the ionizing porosity to UV photons of the ISM of local low-metallicity dwarf galaxies. To do so, we constrain their H II region escape fractions ($f_{\mathrm{esc,HII}}$) based on infrared emission lines arising from both the ionized and neutral phases of their ISM. We also provide an overview of the possibilities opened by the multicomponent modeling framework developed in Chapter 4 and 7 of this thesis. This first application to local, well-known galaxies paves the way for future studies of larger samples of galaxies, including high-redshift objects with fewer spectral constraints (see Chapter 11).

A&A 667, A35 (2022)
https://doi.org/10.1051/0004-6361/202243866

Astronomy & Astrophysics

Inferring the HII region escape fraction of ionizing photons from infrared emission lines in metal-poor star-forming dwarf galaxies

L. Ramambason[1], V. Lebouteiller[1], A. Bik[2], C. T. Richardson[3], F. Galliano[1], D. Schaerer[4], C. Morisset[5], F. L. Polles[6], S. C. Madden[1], M. Chevance[7,8], and I. De Looze[9,10]

[1] Université Paris-Cité, Université Paris-Saclay, CEA, CNRS, AIM, 91191 Gif-sur-Yvette, France
e-mail: lise.ramambason@cea.fr
[2] Department of Astronomy, Oskar Klein Centre, Stockholm University, AlbaNova University Centre, 106 91 Stockholm, Sweden
[3] Physics Department, Elon University, 100 Campus Drive CB 2625, Elon, NC 27244, USA
[4] Observatoire de Genève, Université de Genève, 51 Ch. des Maillettes, 1290 Versoix, Switzerland
[5] Instituto de Astronomía, Universidad Nacional Autónoma de México, AP 106, 22800 Ensenada, BC, Mexico
[6] SOFIA Science Center, USRA, NASA Ames Research Center, M.S. N232-12, Moffett Field, CA 94035, USA
[7] Astronomisches Rechen-Institut, Zentrum für Astronomie der Universität Heidelberg, Mönchhofstrasse 12-14, 69120 Heidelberg, Germany
[8] Institut fur Theoretische Astrophysik, Zentrum für Astronomie, Universität Heidelberg, 69120 Heidelberg, Germany
[9] Sterrenkundig Observatorium, Ghent University, Krijgslaan 281 – S9, 9000 Ghent, Belgium
[10] Department of Physics & Astronomy, University College London, Gower Street, London WC1E 6BT, UK

Received 26 April 2022 / Accepted 11 July 2022

ABSTRACT

Context. Local metal-poor galaxies stand as ideal laboratories for probing the properties of the interstellar medium (ISM) in chemically unevolved conditions. Detailed studies of this primitive ISM can help gain insights into the physics of the first primordial galaxies that may be responsible for the reionization. Quantifying the ISM porosity to ionizing photons in nearby galaxies may improve our understanding of the mechanisms leading to Lyman continuum photon leakage from galaxies. The wealth of infrared (IR) tracers available in local galaxies and arising from different ISM phases allows us to constrain complex models in order to estimate physical quantities.

Aims. Primitive galaxies with low metal and dust content have been shown to host a patchier and more porous ISM than their high-metallicity counterparts, with numerous density-bounded regions from where ionizing photons might leak out. To what extent this peculiar structure contributes to the leakage of ionizing photons remains to be quantitatively studied. Such effects can only be investigated by accounting for the complexity and inhomogeneity of the ISM. We aim to provide a new statistical framework to quantify various galactic observables by constraining a representative multiphase and multisector topology using a combination of 1D models.

Methods. To address these questions, we built a refined grid of models that include density-bounded regions and a possible contribution of an X-ray source. Using MULTIGRIS, a new Bayesian code based on Monte Carlo sampling, we combined the models as sectors under various assumptions to extract the probability density functions of the parameters and infer the corresponding escape fractions from H II regions ($f_{\rm esc,HII}$). We applied this new code to a sample of 39 well-known local starbursting dwarf galaxies from the Dwarf Galaxy Survey.

Results. We confirm previous results that hinted at an increased porosity to ionizing photons of the ISM in low-metallicity galaxies and provide, for the first time, quantitative predictions for $f_{\rm esc,HII}$. The predicted $f_{\rm esc,HII}$ for low-metallicity objects span a large range of values, up to ~60%, while the values derived for more metal-rich galaxies are globally lower. We also examine the influence of other parameters on the escape fractions, and find that the specific star-formation rate correlates best with $f_{\rm esc,HII}$. Finally, we provide observational line ratios that could be used as tracers of the photons escaping from density-bounded regions. Among others, we discuss the possible caveats of diagnostics based on [C II]158 µm in low-metallicity environments as we find a strong metallicity dependence of the fraction of [C II]158 µm emitted in the different ISM phases.

Conclusions. The new framework presented in this paper allows us to use suites of unresolved IR emission lines to constrain various galactic parameters, including the escape fraction of ionizing photons from H II regions. Although this multisector modeling remains too simple to fully capture the ISM complexity, it can be used to preselect galaxy samples with potential leakage of ionizing photons based on current and forthcoming spectral data in unresolved surveys of local and high-redshift galaxies.

Key words. galaxies: starburst – galaxies: dwarf – ISM: structure – radiative transfer – infrared: ISM – methods: numerical

1. Introduction

Young stars that have just been formed irradiate the surrounding gas and ionize the interstellar medium (ISM). Those pockets of ionized gas (i.e., H II regions) often dominate the emission at galactic scales in star-forming galaxies. At equilibrium, H II regions are surrounded by atomic neutral hydrogen, the photodissociation regions (PDRs), which may recombine further from the stars to form H_2 in dense molecular clouds. In some cases, part of the ionizing radiation (Lyman continuum below 912 Å; LyC) can leak out of H II regions and irradiate a diffuse ionized gas (DIG) reservoir (Zurita et al. 2002; Weilbacher et al. 2018;

Herenz et al. 2017; Bik et al. 2018; Menacho et al. 2019, 2021). Depending on the exact morphology of the gas distribution, the ionizing photons can freely travel on large scales to escape in the surrounding circum-galactic medium (CGM) and potentially in the inter-galactic medium (IGM).

The contribution of LyC-leaking galaxies to the total ionizing budget in the epoch of reionization (EoR, $z \sim 6$–9) is a key element in our understanding of the reionization process. Recent simulations indicate that such populations of numerous, low-mass, LyC-leaking galaxies with average escape fractions (f_{esc}(LyC)) of 10–20% would be sufficient to reionize the whole universe without invoking any other contribution from ionizing sources such as active galactic nuclei (AGN; Robertson et al. 2013, 2015). Under favorable assumptions on ionizing photons production and accounting for a subdominant contribution from AGN, Finkelstein et al. (2019) find that even an average escape fraction below 5% throughout the bulk of the EoR would be enough to match observational constraints. Naidu et al. (2020) propose an alternative model where reionization is not driven by the lower mass galaxies but by a few (<5%) highly star-forming galaxies with stellar masses above $10^8\,M_\odot$ and extreme escape fractions (the "oligarchs").

Regardless of which galaxy population drives the reionization process, the inclusion of binary stars, which provide ionizing photons at later stellar evolution stages than single stars, might also play a crucial role in providing energetic photons over large timescales. Simulations of Ma et al. (2016) and Rosdahl et al. (2018) find that this leads to significantly higher time-averaged escape fractions of ionizing photons. In low-metallicity environments hosting very massive stars, stellar-mass black holes might also contribute significantly to the ionizing photon production (e.g., Mirabel et al. 2011). While escaping LyC photons have been directly observed in the UV domain at redshifts below 0.5 (e.g., Bergvall et al. 2006; Leitet et al. 2013; Borthakur et al. 2014; Leitherer et al. 2016; Izotov et al. 2016a,b, 2018a,b; Wang et al. 2019, 2021; Izotov et al. 2021; Flury et al. 2022a,b), and at $z \sim 2$–3 (e.g., Vanzella et al. 2015, 2016, 2018, 2020; Shapley et al. 2016; de Barros et al. 2016; Steidel et al. 2018; Bian et al. 2017; Fletcher et al. 2019; Rivera-Thorsen et al. 2019; Pahl et al. 2021, 2022) with observed f_{esc}(LyC) ranging from 2 to 72%, direct detections of leaking LyC radiation is not possible – or extremely unlikely – above $z \sim 4$ due to the absorption by neutral hydrogen in the IGM, preventing any direct observation of potentially LyC-leaking galaxies directly within the EoR. This observational barrier makes it difficult to perform quantitative studies of primordial LyC-leaking galaxies, which are much needed to understand their role in the reionization process. Moreover, since LyC detections only probe a single line of sight, the measured values are sensitive to viewing angle dependences.

To overcome these constraints, several indirect methods to trace the escape fraction have been explored. They rely on the fact that photons escaping from H II regions are quite sensitive to the structure and properties of the surrounding gas. This view is supported by hydrodynamical simulations of the ISM, which account for an inhomogeneous gas distribution produced by turbulence and/or stellar feedback (e.g., Fujita et al. 2003; Trebitsch et al. 2017; Kimm et al. 2017, 2019; Kim et al. 2018; Kakiichi & Gronke 2021; Yoo et al. 2020). First, far-UV absorption lines can serve as a promising proxy to infer the covering fraction of H I gas, which places an upper limit on the amount of escaping photons (e.g., Reddy et al. 2016; Gazagnes et al. 2018, 2020; Chisholm et al. 2018; Saldana-Lopez et al. 2022). Another approach consisting in looking at far-UV colors selection diagrams has been proposed in, for example Vanzella et al. (2015) and Naidu et al. (2017). Finally, emission lines arising from different phases of the ISM can also be very useful probes of the global porosity to ionizing photons. In the UV domain, the Lyman alpha (Lyα) line profile has proven to be an interesting proxy with the presence of double-peaked or triple-peaked profiles being associated with LyC and Lyα leakage (e.g., Verhamme et al. 2015, 2017; Henry et al. 2015; Izotov et al. 2020; Maji et al. 2022). In particular, the velocity separation of the blue peak and red peak has been shown to strongly anti-correlate with the measured escape fraction of Lyα photons. Recent studies have also highlighted the interest of using proxies such as weak helium lines (Izotov et al. 2017) or the Mg II $\lambda\lambda$2796,2803 Å doublet (e.g., Henry et al. 2018; Chisholm et al. 2020; Xu et al. 2022; Katz et al. 2022a) for probing the low-density lines of sight necessary to allow LyC-photons to escape.

In the optical range, several proxies can serve as indicators of LyC leakage. In particular, line ratios involving ions with different ionization potentials, produced at different depths, have been proposed as indicators to discriminate between radiation-bounded and density-bounded H II regions. Radiation-bounded regions correspond to ionized spheres delimited by their Strömgren radii set by the equilibrium between production of photons by stars and ionization of the surrounding gas. Density-bounded regions are instead delimited by the lack of matter, which sets their outer radius before the Strömgren radius. Hence, they allow part of the LyC-photons produced by stars to escape from H II regions. The oxygen line ratio [O III]λ 5007 Å/[O II] $\lambda\lambda$3726, 3728 Å (O32) proposed by Jaskot & Oey (2013) and Nakajima & Ouchi (2014) was successfully used to select LyC-leaking candidates but no strong correlation was found with the measured values of escape fraction (see Izotov et al. 2018b; Naidu et al. 2018; Bassett et al. 2019; Nakajima et al. 2020, and discussions therein). Based on a similar idea, the lack of emission from ions with low ionization potentials like [S II] $\lambda\lambda$6716,6731 Å has also been proposed to target leaking candidates (Wang et al. 2019, 2021; Katz et al. 2020). This lack of emission of some low ionization species was first interpreted as the signature of a density-bounded galaxy where the outer part of H II regions were completely stripped out. However, using simple photoionization models, Stasińska et al. (2015) have shown that on average galaxies with high O32 cannot have massive escapes of ionizing photons, since low ionization lines like [O I]6300 Å are often also detected in these galaxies, implying the presence of radiation-bounded regions. Subsequently Plat et al. (2019) and Ramambason et al. (2020) noted that several strong LyC-emitters show surprisingly strong [O I]6300 Å emission, and proposed several explanations. While Stasińska et al. (2015) and Plat et al. (2019) suggested that such emission could be powered by the presence of AGN or radiative-shocks, we proposed in Ramambason et al. (2020) a 2-component model combining both density- and ionization-bounded regions.

A complementary picture has been provided by studies of galaxies in the local universe with resolved H II regions. Recently, Della Bruna et al. (2021) estimated the average escape fraction of ionizing photons from H II regions in NGC 7793 to be 67% from MUSE observations with a ~10 pc resolution. A large fraction of those escaping photons is, however, likely reabsorbed within galactic scales and contributes to create a DIG reservoir also seen with MUSE (Della Bruna et al. 2020). This picture is in line with recent PHANGS-MUSE

observations of resolved H II regions in nearby spiral galaxies (Belfiore et al. 2022; Chevance et al. 2022). Other indirect methods applied to local objects based on the mapping of the ionization parameter (e.g., Zastrow et al. 2011, 2013), on the estimation of the intrinsic ionizing photon production rate from resolved stars (Choi et al. 2020), and on the ionized gas kinematics (Eggen et al. 2021) are also suggestive of large escape fractions of ionizing photons from H II regions. Aditionally, results from Polles et al. (2019) on the local starbursting galaxy IC10 indicate that the derived porosity depends on the spatial scale, with most clouds being matter-bounded at small scales (~60 to 200 pc), while larger regions become more and more radiation-bounded at galactic scales. This result highlights the complexity of the ISM in which the energy produced by feedback is deposited at various spatial scales and over different dynamical timescales, hence producing a highly inhomogeneous internal structure.

While the indirect methods relying on spatially resolved information can be applied in the local universe, they are not applicable in more distant unresolved galaxies. Disentangling the contribution from the ionized and neutral gas phase is a crucial step to better understand the interplay between ionizing photons and the surrounding ISM. To do so, the infrared (IR) domain offers interesting tracers, not only arising from H II regions but also from the PDR and molecular phases. Unfortunately, this part of the spectra is often inaccessible in samples of known LyC-leaking galaxies and only a few objects with LyC detections were also observed in the IR. As an alternative, recent studies (e.g., Cormier et al. 2012, 2015, 2019) have tried to constrain the covering factor of neutral gas (PDR covering factor) in local galaxies to quantify the porosity to ionizing radiation by estimating the fraction of gas residing in neutral atomic and molecular phases. This complex approach was first introduced in Péquignot (2008) that provided an unprecedentedly detailed analysis of the proto-typical, low-metallicity galaxy I Zw 18 by developing refined multisector topological models representing the contribution of each phase to the total emission. Similar models were adapted and successfully applied to local objects (e.g., Haro 11: Cormier et al. 2012, I Zw 18: Lebouteiller et al. 2017, IC 10: Polles et al. 2019), to a sample of local dwarf galaxies (Cormier et al. 2019) drawn from the Dwarf Galaxy Survey (DGS; Madden et al. 2013) and to resolved regions in the Small and Large Magellanic Clouds (SMC/LMC; Lambert-Huyghe et al. 2022). The results from those studies indicate that nonunity PDR covering factors of neutral gas are necessary to reproduce the emission lines of most local objects. Such findings appear at odds with the few UV observations that detect little to no LyC-leakage in the local universe (e.g., Bergvall et al. 2006; Leitet et al. 2013; Borthakur et al. 2014; Leitherer et al. 2016).

In this context, it becomes crucial to understand what properties of the ISM are responsible for its porosity to ionizing radiation and determine if and how well integrated emission lines of unresolved galaxies can be used to constrain their escape fractions of ionizing photons. The question is especially interesting in the context of high-redshift studies, as more and more galaxies are detected with facilities like ALMA and NOEMA at $z \sim 4$–9 (e.g., Inoue et al. 2016; Carniani et al. 2017; Walter et al. 2018; De Breuck et al. 2019; Hashimoto et al. 2019; Harikane et al. 2020; Falkendal et al. 2021; Bakx et al. 2020; Meyer et al. 2022). Such observatories and the advance of future ones such as the *James Webb* Space Telescope (JWST) are opening a new window to observe galaxies close to or within the EoR.

In this paper, we present a first application of MULTIGRIS (Lebouteiller & Ramambason 2022, hereafter LR22), a new Bayesian code designed to constrain multisector models using spectra of unresolved galaxies. Our work builds on previous studies (e.g., Cormier et al. 2012, 2019; Lebouteiller et al. 2017; Polles et al. 2019; Lambert-Huyghe et al. 2022) in which multisector models were constrained using frequentist methods. In particular, this paper is a direct continuation of Cormier et al. (2019) that used a χ^2 minimization method to select the best-fitting configurations between 1- and 2-sector models and derived PDR covering factors. We revisit those results using the same sample of galaxies but adopting a new method. Using a Bayesian framework allows us to overcome some major issues of the χ^2 method: difficulty to derive errorbars, sensitivity to outliers, impossibility to include complex priors etc. Most importantly, it allows us to infer probability density functions (PDF) of various parameters, including for the first time the escape fraction of ionizing photons from H II regions ($f_{\rm esc,HII}$).

The paper is organized as follows. In Sect. 2 we present our sample and the tracers used in the analysis. The grid of models and the Bayesian code are presented in Sects. 3 and 4 and our results in Sects. 5 and 6. We discuss the limits and possible improvements of this new framework in Sect. 7. Our main conclusions are summarized in Sect. 8.

2. Sample

2.1. Overview

Our sample is drawn from the *Herschel* Dwarf Galaxy Survey (DGS; Madden et al. 2013), which gathers photometric and spectroscopic observations of 50 nearby (0.5−191 Mpc) galaxies in the far-infrared (FIR) and submillimeter domains performed with the *Herschel* Space Telescope. All of these galaxies were also observed in the mid-infrared (MIR) domain with the *Spitzer* observatory and spectrocopic measurements are available for all but 5 objects. We focus on a sub-sample used in Cormier et al. (2019) that selected 38 compact galaxies with at least three spectral lines detected in the MIR and FIR domain (~5 to 120 μm) among the 50 observed galaxies. The DGS sample exhibits a wide range of physical properties, which makes it an ideal laboratory to study the variation of ISM properties over a range of physical and chemical conditions. In particular, this sample is ideally suited to study how the escape fraction evolves in local, low-metallicity galaxies whose ISM may resemble primordial galaxies from the early universe. More specifically, the intense sources of radiation (see Sect. 2.2) as well as the low masses, compact sizes and metal-poor gas reservoirs (see Sect. 2.3) of these galaxies may, to some extent, resemble the physical and chemical conditions of the primordial dwarf galaxies. We note, however, that our selection criterion favors IR-bright galaxies hosting actively star-forming regions that have formed in a previously enriched ISM. The possible analogy with primordial galaxies should hence be taken with caution.

We use the line fluxes provided in Cormier et al. (2019) which combine *Herschel*/PACS (60–210 μm) data with *Spitzer*/IRS (5–35 μm) data. The latter are available for all galaxies in our sample except three (HS 0017+1055, UGC 4483 and UM 133). The fluxes correspond to integrated measurements for galaxies that were fully covered by the instrumental apertures, except for one galaxy. The only exception is NGC 4214, which was observed in two pointings (central and southern star-forming regions) that are studied separately in this study. The extraction procedures and corrections applied to extended

sources are detailed in Cormier et al. (2015). We corrected one line flux from Cormier et al. (2019) ([Ne V]14 μm for NGC 5253) that corresponded to a false detection due to an error in the fitting process. Finally, we instead use a detection upper limit at 2-σ of 0.112×10^{16} W m^{-2}. We also include the total IR luminosities ($L_{\rm TIR}$) derived by modeling the dust spectral energy distributions (SEDs) in Rémy-Ruyer et al. (2015). All the available upper limits were used in the analysis, contrary to Cormier et al. (2019) that manually selected a suite of classical emission lines arising from H II regions and PDR. Our study aims to extend the multiphase picture of the ISM that was provided in Cormier et al. (2019) by including more lines arising from different phases of the ISM and tracing different physical processes. The list of observables used as constraints is summarized in Table 1.

Although it requires several lines to derive the various parameter values with reasonable uncertainties, the code can run properly as long as it is provided with at least one line and one upper limit (which are necessary to set the prior distributions). In this case, the given solution corresponds to a relatively wide PDF (defined in Sect. 4.3) and large errorbars. While little can be said about the parameter values of such individual objects, they do not, however, bias the analysis of the global trends in our sample. We thus decided to include all objects with at least one detection and one upper limit, which adds one more galaxy to the sample of Cormier et al. (2019) (HS 2352+2733, detected in [O III]88 μm with two upper limits on [C II]158 μm and $L_{\rm TIR}$), hence leading to a total of 39 galaxies.

In addition to including upper limits for the first time, we can also enlarge the selection of lines since our method is more robust to outliers than the previously used χ^2 method. The number of constraints available for each galaxy varies from 1 single detection (in addition to two upper limits) up to 22 detected emission lines. The exact number of detections and instrumental upper limits available for each object is provided in Fig. A.4. The choice of the suite of lines used in the analysis has an impact on the best solution that is selected by our code and, in turn, on the parameter values that are derived. Choosing the optimal suite of lines to be used and the minimal number needed to constrain a given parameter is a complex problem and we refer to LR22 for a more detailed discussion.

In this study we focus only on IR lines although most galaxies in our sample are also detected in the optical domain. We check a posteriori that our models predict values consistent with the available Hα measurements (see Sect. 5.2). Combining optical and IR lines is possible with MULTIGRIS but would require an additional treatment of the dust attenuation and of systematic uncertainties due to instruments. Increasing the number of lines used as constraints can also raise specific issues regarding redundancies and the risk to over-constrain some parameters. We postpone this study to a future work and refer to LR22 for an example combining both IR and optical lines.

2.2. Radiation sources and feedback

The DGS galaxies are starbursting galaxies with prominent MIR and FIR emission lines that hint at the presence of a population of young UV emitting stars, which strongly irradiate the ISM (Madden et al. 2013). Some galaxies in our sample (15/39) have also been reported to host a population of Wolf-Rayet stars and signatures associated with massive stars have been reported in Schaerer et al. (1999).

Additionally, evidence suggests that several galaxies in our sample may host X-ray sources with the claimed detections reported in the following papers: HS 1442+4250, VII Zw 403 (Papaderos et al. 1994; Kaaret et al. 2011; Brorby et al. 2014), NGC 1569, NGC 5253, NGC 4214 (Ott et al. 2005; Binder et al. 2015; McQuinn et al. 2018) and NGC 625 (McQuinn et al. 2018). Some of them even host Ultra-Luminous X-ray sources (ULX) with measured X-ray luminosities above 10^{39} erg s^{-1}: Haro 2 (Otí-Floranes et al. 2012), Haro 11 (Prestwich et al. 2015; Gross et al. 2021), He 2–10 (Ott et al. 2005; Reines et al. 2011), I Zw 18 (Thuan et al. 2004; Ott et al. 2005; Kaaret et al. 2011; Kaaret & Feng 2013; Brorby et al. 2014) and SBS 0335-052 (Thuan et al. 2004; Prestwich et al. 2013). We note that in some objects, other physical mechanisms are considered to explain the X-ray emission and the claimed detections of compact objects have been actively debated (e.g., He 2–10; Cresci et al. 2017, HS 1442+4250; Senchyna et al. 2020, and NGC 5253; (Zastrow et al. 2011, and references therein). Regardless of their exact nature (e.g., high-mass X-ray binaries, intermediate mass black hole or AGN), the contribution of such sources may have an important impact in low-metallicity, transparent environments in which high energy photons can travel freely over large scales. In I Zw 18, the galaxy with the lowest metallicity in our sample, Lebouteiller et al. (2017) have shown that the X-ray radiation emitted by a single point source ULX dominates the energy balance in the neutral gas over galactic scales.

We also expect X-ray heating of the ISM to produce specific spectroscopic signatures such as the presence of emission lines associated with very high ionization potentials (e.g., above 40eV) ions. Among the X-ray-sensitive tracers that we consider in this study (see Table 1), [Ne III] is detected in 33/39 sources and [O IV] detected in 17/39 sources. While we only have upper limits on [Ne V]$\lambda\lambda$24,14 μm, other lines produced by ions with high ionization potentials have been detected in the optical domain (e.g., [Fe IV], [Fe V], [Ne V] and even [Fe VI],

Table 1. IR tracers used as constraints and corresponding ionization potentials for ionic lines.

	Tracers
Molecular hydrogen	H_2 S(0), H_2 S(1), H_2 S(2), H_2 S(3)
Neutral and ionized gas tracers	[O I]$\lambda\lambda$63,145 μm [Fe II]$\lambda\lambda$17,25 μm (7.9 eV), [Si II]λ34 μm (8.2 eV), [C II]λ158 μm (11.3 eV), Hu$\alpha\lambda$12 μm (13.6 eV), [N II]$\lambda\lambda$122,205 μm (14.5 eV), [Ar II]λ7 μm (15.7 eV), [Fe III]λ23 μm (16.2 eV), [Ne II]λ12 μm (21.6 eV), [S III]$\lambda\lambda$18,33 μm (23.3 eV), [Ar III]$\lambda\lambda$9,21 μm (27.6 eV), [N III]λ57 μm (29.6 eV), [S IV]λ10 μm (34.7 eV), [O III]λ88 μm (35.1 eV), [Ne III]λ15 μm (40.9 eV) [O IV]λ26 μm (54.9 eV), [Ne V]$\lambda\lambda$14,24 μm (97.1 eV),
Total IR luminosity	$L_{\rm TIR}$ (1 μm–1000 μm)

Notes. We report the ionization potentials corresponding to the energy thresholds required to create the ion producing a given emission line, either by de-excitation or by recombination.

[Fe VII]; Izotov et al. 2001, 2004a,b). We note, however, that the origin of such lines is debated with two main hypothesis being either the presence of X-ray sources (e.g., Lebouteiller et al. 2017; Schaerer et al. 2019; Simmonds et al. 2021) or the presence of fast radiative shocks (e.g., Allen et al. 2008; Izotov et al. 2012). In this study we do not take into account shocks since most of the lines that we consider arise from the H II region and PDR where the dominant source of feedback is likely radiation pressure (Lee et al. 2016, 2019). In any case, although shocks may contribute to the emission of some of the lines we consider (e.g., lines associated with very high ionization potentials and H_2 lines), we do not have enough constraints to disentangle the contribution of shocks to the emission in the current study. For a similar reason, we do not study the impact of varying the cosmic rays (CR) rate in our models although CR might significantly contribute to the heating of neutral gas, especially at low-metallicity (Lebouteiller et al. 2017).

Even when nonstellar sources do not significantly contribute to the energy balance, their presence can be linked to mechanical feedback mechanisms that may facilitate the escape of LyC photons produced by stars. Any additional kind of feedback that modifies the gas reservoirs surrounding the stars (e.g., through fragmentation or creation of low-density channels) might affect the resulting global porosity to ionizing photons. The DGS sample exhibits a wide variety of feedback mechanisms, which impact the gas structure and kinematics. In particular, several indications of the ISM being disrupted have been discussed in previous studies that have reported signatures of outflows (in Haro 2: Otí-Floranes et al. 2012, Haro 11: Menacho et al. 2019), ionization cones (NGC 5253: Zastrow et al. 2011), superbubbles possibly associated with galactic winds (NGC 1569: Westmoquette et al. 2008; Sánchez-Cruces et al. 2015), and other signatures of strong feedback phenomena (NGC 1705: Zastrow et al. 2013, NGC 625: Cannon et al. 2004; McQuinn et al. 2018, UM 461: Carvalho & Plana 2018, NGC 4214: Martin 1998; McQuinn et al. 2018 and Pox 186: Eggen et al. 2021). Such dynamical effects strongly affect the chemical and physical conditions of the surrounding gas. They might even lower the metallicity of galaxies since metal-enriched outflows can remove newly formed metals from H II regions (e.g., Amorín et al. 2010; Hogarth et al. 2020). Another direct effect might also be the creation of low-density channels formed by gas ejection that may favor escaping photons.

2.3. Gas and dust properties

Our sample spans a large range of sub-solar[1] metallicities ranging from 12+log(O/H) = 7.14 ($\sim$1/35 $Z_\odot$) up to 8.43 ($\sim$1/2 $Z_\odot$), derived from empirical strong line methods (Madden et al. 2013; Rémy-Ruyer 2013). Their dust-to-gas mass ratios also span a large range of values (0.07–0.33) and have been carefully investigated in Rémy-Ruyer et al. (2014, 2015) and Galliano et al. (2021) using continuum measurements from *Spitzer* and *Herschel*.

Madden et al. (2020) have shown that in such environments, the UV photons can penetrate deeper in the clouds and photodissociate CO, hence creating a layer of CO-dark gas. While H_2 and CO are only detected in a few galaxies in our sample, evidence suggests that their molecular gas reservoir could be largely underestimated when using CO lines as a tracer. Using in particular the [C II]158um line, Madden et al. (2020) developed a method to estimate masses of CO-dark H_2 gas residing in the envelop where CO is photodissociated by strong radiation field and find that the DGS galaxies most likely host unseen molecular gas reservoir (>70% of the total H_2 in all the galaxies in our sample) that can explain their high star-formation rates (SFR; $-2.2 < \log \mathrm{SFR} < 1.4$) compared to the CO-based estimates of molecular gas content.

Additionally, some of the DGS galaxies are associated with large H I reservoirs with masses ranging from 10^7 to $\sim 10^{11}\ M_\odot$ (Rémy-Ruyer et al. 2014) and specific gass masses ($M_{\rm HI}/M_*$) ranging from 0.03 to 17.3. Although this neutral component is important, little is known on the actual distribution of this gas in our sample. Some of the objects observed in absorption have been associated with large column densities reaching values greater than 10^{21} cm^{-2} (see Table A.3 and associated references). Such high values completely rule out photons escaping along these lines of sight. However, previous studies have also shown the inhomogeneous distribution of the neutral component and line-of-sight effects (e.g., Gazagnes et al. 2020).

3. Models

3.1. Modeling strategy

One of the main challenges driving our modeling strategy is the necessity to deal with spatially unresolved observations in which the structure of the ISM is not directly accessible. In such cases, emission arising from the different phases of the ISM (e.g., H II regions, PDRs, molecular gas, DIG) is blended into one single beam.

To overcome this problem, we recover the underlying gas distribution in the different ISM phases from unresolved spectra. Using a multisector topology as a representative view of the galaxy, such as proposed in Péquignot (2008), it becomes possible to disentangle the relative contribution of each phase. Each "sector" corresponds to a fraction of a sphere of gas where radiative transfers are performed in a 1D continuous way throughout the H II region, PDR, and molecular zone. While this model simplifies the actual geometry of a galaxy, there is much to gain from combining sectors (even independent ones) as compared to a single model, especially when dealing with tracers coming from different phases or depths. The need to automate and generalize this multisector topological approach led to the development of a new code, MULTIGRIS, which allows a flexible combination of models within a grid. We briefly describe the code in Sect. 4 and refer to LR22 for a detailed description of the general strategy used in combining sectors.

The models used in the combination must account for the main physical and chemical mechanisms producing the line emission. Photoionization and photodissociation codes with complex chemical networks and refined prescriptions for radiative transfer are well suited to that purpose. They have been extensively used to study H II regions (e.g., with Cloudy; Ferland et al. 2017, MAPPINGS V; Sutherland et al. 2018) or PDR (e.g., with MeudonPDR; Le Petit et al. 2006). We chose to use Cloudy models that allow a consistent treatment of the emission line physics throughout the ionized phase, PDR, and molecular phase. Combining such models into a representative topology somewhat compensates for their simplistic geometry, which, by itself, cannot capture the full complexity of the multiphase ISM. In particular, several studies pointed out the necessity to combine several regions having different physical depths to successfully reproduce the emission lines of local objects (e.g., Binette et al. 1996; Péquignot 2008;

[1] We use the solar value from Asplund et al. (2009) of 12+log(O/H)= 8.69.

Lebouteiller et al. 2017; Cormier et al. 2012, 2019; Polles et al. 2019; Ramambason et al. 2020; Lambert-Huyghe et al. 2022). The present study is a direct continuation of the previous analysis from Cormier et al. (2019) of the DGS sample (presented in Sect. 2.1) using topological models. Based on 1-sector and 2-sector models, with varying PDR covering factors, they selected the best fitting models using a χ^2 minimization. In each model, at least one sector was computed until $A_V = 5$ and the PDR scaling factor was defined a posteriori as a linear scaling, between 0 and 1, applied to the lines arising from the PDR. For 2-sector models, an additional sector stopping at the ionization front was added. Based on this definition, they find nonunity covering factors for most of the galaxies in our sample, and a strong correlation between metallicity and covering factor.

Similarly to Cormier et al. (2019), we model each galaxy as a weighted combination of sectors where the mixing weights represent the contribution of each sector (see Sect. 4.1). The models are computed with a fixed cluster luminosity and scaled to match the observed line fluxes used as constraints with a free scaling factor defined within the Bayesian model (see Sect. 4.2). The prescriptions used for this new grid are inspired by Cormier et al. (2019) with a few modifications. One major change is that the line luminosities are saved for each tracer in a cumulative mode, for each depth computed in Cloudy, allowing us to include density-bounded sectors (i.e., models stopping before the ionization front), which was not the case in Cormier et al. (2019). This setting also allows a refined sampling of the parameter controlling the depth of each model (see Sect. 3.2.3). We discuss the changes in the model grid and their implications in the following section.

3.2. Cloudy models

The grid used in this article was built using the photoionization and photodissociation code Cloudy v17.02 (Ferland et al. 2017). Each model consists of a spherical shell of gas placed at a fixed inner radius of the incident radiation source. We use a closed, spherical geometry taking into account the transmitted and reflected radiation, assuming a unity covering factor of the gas. The radiative transfer is computed along each line of sight (1D) in a continuous way throughout the H II region, PDR, and molecular zone. We summarize the main input parameters and the range of values spanned for each parameter in Table 2. The grid contains 28 800 models with an X-ray source and 3200 models without X-ray source, which results in a total of 32 000 Cloudy models. Each model is then truncated a posteriori with 17 cuts to create 544 000 sub-models. The current grid includes predictions for 516 emission lines from UV to IR.

Table 2. Input parameters of Cloudy models.

	Fixed parameters
Stellar population	BPASS v2.1[(1)] (Eldridge et al. 2017) Broken power-law IMF α_1 (0.1 $M_\odot$ – 0.5 $M_\odot$) = −1.30 α_2 (0.5 $M_\odot$ – 100 $M_\odot$) = −2.35 Single burst, including binary stars
X-ray component	Multicolor blackbody (Mitsuda et al. 1984), $T_{out} = 10^3$K, variable T_{in}
Other fields	Cosmic microwave background Cosmic ray background 0.5log
Grains	SMC graphite/silicate model (Weingartner & Draine 2001)
Z_{dust}, q_{PAH}	Values scaled with metallicity based on the analytical fits from Galliano et al. (2021)
X/O profiles [(2)]	N[(1)], C[(1)], Ne[(2)], S[(2)], Ar[(2)], Fe[(2)], Si[(3)], Cl[(4)]
Other elements	Table ISM from Cloudy
Density law	$n_H = n_{H,HII} \times$ (1+N(H)/[10^{21} cm^{-2}])
Stopping criteria	A_V=10 or T_e = 10K
	8 varied parameters
log($L_*/L_\odot$)	[7, 9]
L_X/L_*	[0, 0.001, 0.01, 0.1]
log($T_{X,in}$)	[5, 6, 7] K
log([O/H])	[−2.190, −1.889, −1.190, −0.889, −0.667, −0.491, −0.190, 0.111]
Age of burst	[1,2,3,4,5,6,8,10] Myr
log(n_H)	[0,1,2,3,4] cm^{-3}
log(U_{in}) [(3)]	[−4, −3, −2, −1, 0]
Cut [(4)]	[0,4], step=0.25

Notes. [(1)]Metallicity bins in BPASS are $Z = 10^{-5}$, 10^{-4}, 0.001, 0.002, 0.003, 0.004, 0.006, 0.008, 0.010, 0.014, 0.020, 0.030 and 0.040 (i.e. 0.05% of solar to twice solar). We note that $Z_\odot = 0.02$ while the default value we use here is $Z_\odot = 0.014$. [(3)]The input ionization parameter, U_{in}, is set by varying the inner radius for a given luminosity. It is described in Sect. 3.2.3. [(4)]The cut parameter controls the stopping depth of a given model. It is described in Sect. 3.2.3.
References. [(2)]References for elemental abundances. (1): Nicholls et al. 2017, (2): Izotov et al. 2006, (3): Izotov & Thuan 1999, (4): fixed at −3.4.

3.2.1. Radiation field

In our grid we consider a stellar component and an X-ray component as represented in Fig. 1. The stellar population consists of a single burst population from BPASSv2.1 (Eldridge et al. 2017) that includes the contribution from binary stars. We note that the stellar population in Cormier et al. (2019) was instead computed using a continuous star-formation history (SFH) for a 10 Myr old cluster simulated with Starburst99 (Leitherer et al. 1999). Switching to a continuous SFH instead of single-bursts would result in a change in our ionizing spectrum as young O/B stars would provide ionizing photons over longer timescales. Since we consider only one single burst, our grid spans ages below 10 Myr, after which the Hα emission drops. We stress, however, that for the lowest metallicity models, the contribution from binary stars delays this drop in Hα and that ages up to ∼30 Myr (Xiao et al. 2018) should be explored. Considering that our solutions favor ages below 6 Myr and that the addition of a new bin in stellar age would significantly increase the size of our grid of models, we postpone this potential improvement to a future work. In practice, an older population of stars is present in many of the galaxies in our sample but single bursts of later ages alone would not match the spectral signatures that we study here. We note that considering a combination of bursts instead of single-burst model could have a substantial impact on the ionizing continuum, since mixed-age models, which allows contributions from extremely young stars, typically produce more ionizing photons and harder ionizing spectra than single-burst

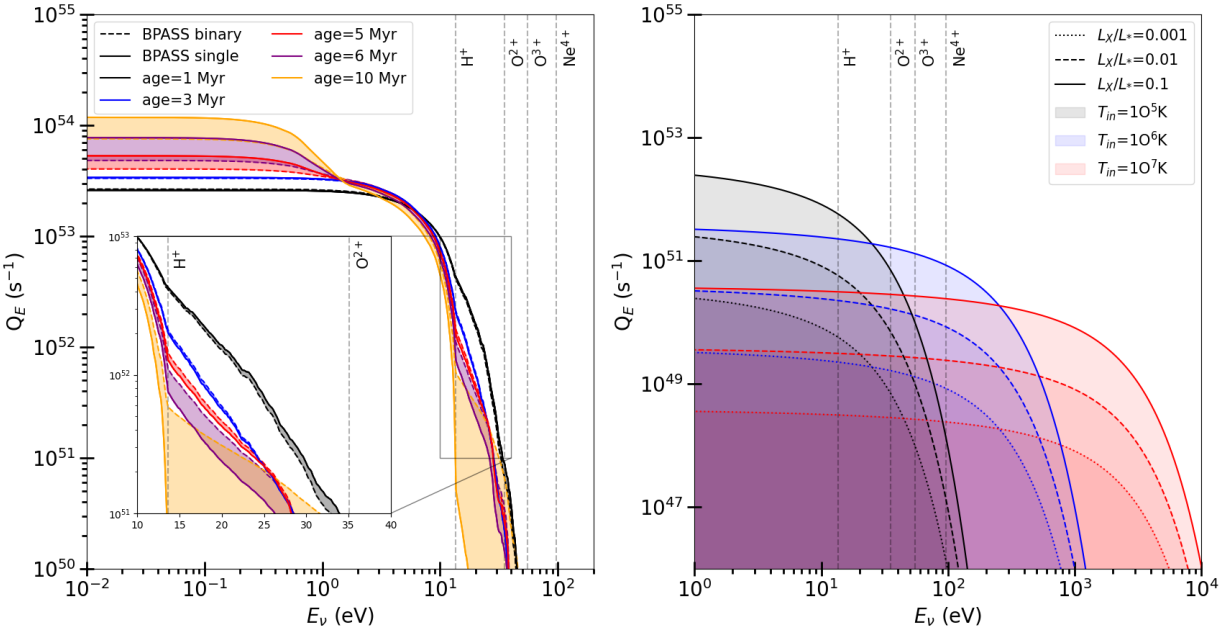

Fig. 1. Stellar (*left*) and X-ray (*right*) incident SEDs of Cloudy models for a solar metallicity and bolometric luminosity of $10^9\,L_\odot$. Q_E shows the number of photons produced above a given energy E_ν. The vertical dashed lines represent the ionization potentials of H^+ (13.6 eV), O^{2+} (35.1 eV), O^{3+} (54.9 eV) and Ne^{4+} (97.1 eV). *Left-hand side*: the dashed lines represent BPASS models with binary stars while the solid lines of the same color correspond to BPASS models without binary stars for a single stellar population of the same age. The change in the spectra due to the inclusion of binary stars is represented by the shaded area in between both lines. The insert shows a zoom around the Lyman edge at 912 Å where the additional contribution to ionizing photons is visible for ages above 3 Myr and increases with the age of the burst. *Right-hand side*: the colors represent different inner temperature of the multicolor blackbody and the different linestyles to different percentage of stellar luminosity: 0.001%, 0.01% and 0.1%.

and constant star-formation rate models corresponding to similar ages (Chisholm et al. 2019).

The stellar evolutionary tracks incorporate the effect of mass transfers between members of binary systems and the stellar initial mass function (IMF) includes a distribution of binaries tuned to reproduce the binary fractions observed in the local Universe. We use a broken power-law IMF with two indices of −1.3 and −2.35 with a change of slope at $0.5\,M_\odot$, which is the default IMF in Eldridge et al. (2017). Although very massive stars with masses above $100\,M_\odot$ may exist in local, low-metallicity galaxies (Crowther et al. 2010; Wofford et al. 2021), such objects remain largely unconstrained in models. Since we do not need to invoke very massive stars to reproduce the IR lines considered in this study, we choose to use the BPASS default mass cut-off at $100\,M_\odot$.

In Fig. 1, we illustrate the effects of including binary stars by comparing the single-star and binary-stars BPASS (Eldridge et al. 2017) SEDs. A clear difference is visible at ages above 3–4 Myr where models that include binary stars produce more ionizing photons. This feature has also been pointed out in Xiao et al. (2018): while single-star and binary-star populations produce fairly similar hydrogen- and helium-ionizing spectra at young ages, the inclusion of binaries produces a shallower drop in ionizing flux than for their single-star counterparts at later ages. The inclusion of binaries with ages between 1–10 Myr has a most profound effect on lines with ionization potentials between 13.6 up to ∼54 eV (see Table 2). Indeed, X-ray binaries provide high energy photons with an energy sufficient to power the emission of several species with very high ionization potentials.

However, it is well-known that these high ionization lines cannot be reproduced by classical photoionization models (e.g., Stasińska et al. 2015), even when including effects of binary stars (cf. Stanway & Eldridge 2019). Other sources producing a harder ionizing continuum need to be invoked to reproduce the high ionization potential lines observed in local and high-redshift galaxies (e.g., He II: Kehrig et al. 2018; Schaerer et al. 2019; Stanway & Eldridge 2019; Senchyna et al. 2020, 2021; or [C IV]: Stark et al. 2015; Senchyna et al. 2021, 2022). For example, high-mass X-ray binaries have been proposed in Schaerer et al. (2019) to explain the presence of ubiquitous He II emission. While Senchyna et al. (2020) find that the addition of an accretion disk associated with high-mass X-ray binaries is insufficient to fully explain the He II emission, several recent studies have argued that the addition of an X-ray source is still needed to simultaneously reproduce the emission lines arising from different ISM phases, including the emission from ions with high ionization potential (e.g., Simmonds et al. 2021; Olivier et al. 2021; Umeda et al. 2022).

Similarly, we find that in order to produce both [O IV] and [Ne V] emission, we need to add an X-ray source that produces harder ionizing photons than BPASS models alone. As discussed in Sect. 2.1, the exact nature of such compact objects is unknown (e.g., high-mass X-ray binaries, intermediate mass black hole or AGN). Moreover, the X-ray spectra themselves are widely unconstrained and rely on strong modeling assumptions (Simmonds et al. 2021). Hence, we choose to use a general prescription to model a compact object surrounded by an accretion disk. To do so, we used a multicolor blackbody spectrum as defined in Mitsuda et al. (1984). This spectrum is defined by an outer temperature fixed at 10^3 K and an inner temperature varying from 10^5 to 10^7 K. The luminosity is set with respect to the stellar luminosity and varies from 0% to 10% of the stellar cluster luminosity. The right-hand side panel of Fig. 1 shows

the X-ray component when varying the inner temperature of the disk and the relative luminosity. Although simplistic, our prescription for the X-ray source is more general than that used in previous studies that considered single blackbody spectra (e.g., Lebouteiller et al. 2017; Cormier et al. 2019).

Finally, we include the cosmic microwave background (CMB) and CR background. Following Cormier et al. (2019) we use a CR rate ∼3 times higher than the standard CR rate (2×10^{-16} s^{-1}; Indriolo et al. 2007) to account for the recent star-formation history in the DGS and match the observed [O I] IR line ratio. We note that CRs and X-rays both impact the ionization and heating in the PDR but we do not have means to disentangle both effects. This somewhat arbitrary choice for the CR rate has a strong effect at low metallicities (below $1/10\,Z_\odot$) where the emission of low ionization potential species and recombination lines is boosted even at large A_V, deep inside the molecular zone. However, in the range of A_V, density and temperature favored in the results presented here, the CR effects should not be significant. Because we have no means to discriminate between CR and X-ray effect, we do not further discuss their potential contribution in this paper.

3.2.2. Metal and dust abundances

Our grid is designed to match the abundance patterns of low-metallicity dwarf galaxies that belong to the group of blue compact dwarf (BCD) galaxies. The prescriptions used in our models are summarized in Table 2. The abundances for nitrogen and carbon are based on Nicholls et al. (2017) who derive analytical curves accounting for primary and secondary production. Their fit is derived from a large sample of stellar measurements in the Milky Way spanning a wide range of metallicities ($6 \leq 12+\log(\mathrm{O/H}) \leq 9$). Their analytical curves are compatible with the gas-phase measurements for carbon and nitrogen in the BCDs from Izotov & Thuan (1999), which have little depletion due to their poor dust content. Neon, sulfur, argon, and iron abundance profiles follow the regressions of Izotov et al. (2006), based on a sample of low-metallicity BCDs. To avoid extrapolating at high metallicities, we used flat profiles for metallicities above 8.2 for those four elements. Because of the low statistics of Izotov et al. (2006) for chlorine, we fixed the [Cl/O][2] value to the median of their sample (−3.4). The silicon abundance profile is based on Izotov & Thuan (1999). For all the other metals we used the values from the Cloudy ISM table, assuming they scale linearly with the oxygen abundance. The profiles used in our grid are given in Fig. A.1.

We apply the same scaling relative to the solar value to both the gas and stellar metallicity. In practice, since BPASS uses a slightly different solar value ($Z_\odot = 0.020$) than the Cloudy one ($Z_\odot = 0.014$), the metallicities are slightly offset. The helium abundance is fixed to the value provided in the ISM abundance set in Cloudy. The dust-to-gas mass ratio and abundance of polycyclic aromatic hydrocarbons (PAH) are computed following the metallicity-dependent prescription of Galliano et al. (2021) based on Bayesian dust SED fits of 798 galaxies (including our sample) using the code HerBIE (Galliano 2018) with the THEMIS grain properties (Jones et al. 2017). The dust-to-gas mass ratio hence follows a 4th degree polynomial relation at metallicity above $12+\log(\mathrm{O/H}) = 7.3$ and scales linearly with metallicity below this threshold (see Eqs. (8)–(9) from Galliano et al. 2021).

Galliano et al. (2021) also provide an analytical prediction for the mass fraction of aromatic features emitting grains ($q_{\rm AF}$), assuming those features are carried by small a–C(:H) grains. Here, we assume that such features are carried by PAHs instead and estimate the mass fraction of PAHs ($q_{\rm PAH} = M_{\rm PAH}/M_{\rm dust}$) as $q_{\rm PAH} \sim q_{\rm AF}/2.2$. The abundance of PAHs is known to strongly vary under different physical conditions: they are especially sensitive to metallicity effects and to the strength of the interstellar radiation field (ISRF). Galliano et al. (2021) find that $q_{\rm PAH}$ is primarly driven by metallicity effects while correlations with ISRF indicators are weaker. We adopt a single value of $q_{\rm PAH}$ per metallicity bin, corresponding to the analytical fit from Galliano et al. (2021). We note that this prescription sets the total PAH abundance in our Cloudy models while the abundance profile follows the default one in Cloudy, that is, scaling as H^0/H. This results in a profile where PAHs are mostly present in the PDR but completely destroyed in the H II regions and molecular zone. This assumption is consistent with observational studies that show that PAHs are absent in the ionized region (e.g., Relaño et al. 2018; Chastenet et al. 2019); however, PAHs can also be present in the molecular phase (e.g., Chastenet et al. 2019), which have not been considered in the Cloudy modeling. Considering that all the PAHs reside in the PDRs might in turn result in a slight overestimation of the PDR heating by photoelectric effect.

3.2.3. Geometry of H II regions

Each of our models consists in a spherical shell of gas around a central ionizing source, whose inner radius is set to match the input ionization parameter defined as:

$$U_{\rm in} = \frac{Q(\mathrm{H}^0)}{4\pi n c R_{\rm in}^2}, \tag{1}$$

where n is the hydrogen density at the inner radius, $U_{\rm in}$ is the input ionization parameter, c is the speed of light, and $Q(\mathrm{H}^0)$ is the total number ionizing photons produced by the central source defined as follows:

$$Q(\mathrm{H}^0) = \int_{E_\nu=13.6\,\mathrm{eV}}^{\infty} \frac{L_*(\nu) + L_{\rm X}(\nu)}{\mathrm{h}\nu}\mathrm{d}\nu, \tag{2}$$

where $L_*(\nu)$ is the stellar luminosity and $L_X(\nu)$ the luminosity of the X-ray source emitted with a given frequency ν. We note that this definition of $U_{\rm in}$ only constrains the ionization parameter at the inner irradiated edge and that the volume-averaged ionization parameter could be significantly different depending on the model parameters. In particular, we note that this definition is not appropriate to use for thick shells of gas for which the inner ionization parameter ($U_{\rm in}$) at the illuminated front can be very different from the ionization parameter at the ionization front $U_{\rm out}$. As an illustration, the relation between $\log U_{\rm in}$, $\log U_{\rm out}$ and the volume-averaged $\log\langle U\rangle$ is provided in Fig. A.5. Since the inner radius is set automatically to match a given pair of input ionization parameter and gas density, the only possibility to change the geometry of the region is to change the incident luminosity. At fixed $U_{\rm in}$, a large cluster luminosity result in a more shell-like geometry while a lower cluster luminosity result in a more filled-sphere geometry in which the gas lies closer to stars. Stasińska et al. (2015) show that such geometrical effects can affect the low-ionization to high-ionization line ratio (e.g., [O I]/[O III]) as the emission from outer regions is boosted in a compact configuration. Although this geometrical effect is only secondary for most lines arising from the H II regions, it has a

[2] The brackets indicate values normalized by the solar abundance ratio.

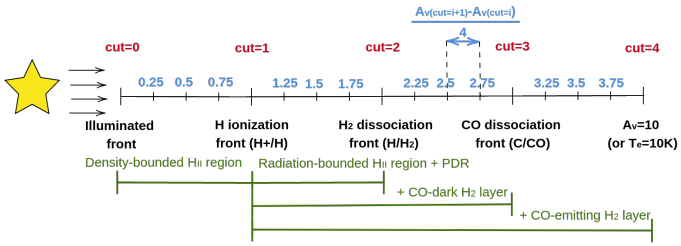

Fig. 2. Schematic view of the 17 cuts used to create sub-models.

much stronger effect on lines emitted near the ionization front and in the PDR and molecular zone.

The bolometric luminosity of $10^9\,L_\odot$ chosen by Cormier et al. (2019) corresponds to the case of a galaxy dominated by one single to a few giant H II regions. For the range of ionization parameters and stellar ages covered in our grid this corresponds to a thick shell geometry with log Hα between 39.8 and 40.7 (in erg s^{-1}). Such super-giant H II regions with Hα luminosities above 10^{39} erg s^{-1} are numerous amoung BCD and starburst galaxies, but not always present (Youngblood & Hunter 1999). They populate the upper end of the observed Hα luminosity function derived from resolved galaxies (Bradley et al. 2006) and are likely to be optically thin (Pellegrini et al. 2012). We also consider a lower luminosity $L_{\rm bol} = 10^7\,L_\odot$ ($37.8 \leq \log {\rm H}\alpha\,({\rm erg\,s^{-1}}) \leq 38.7$), which corresponds to a typical H II region, similar to those that dominate the Hα luminosity function just before the cut-off value ($\log {\rm H}\alpha = 38.6\,{\rm erg\,s^{-1}}$) from Bradley et al. (2006). This case mimics a bursty star-formation where a few hundred to a few thousand compact clusters are responsible for the total emission.

We adopt the same density law as in Cormier et al. (2019) in which $n_{\rm H}$ is nearly constant in the H II region and scales with column density above 10^{21} cm^{-2}. This law provides a simple first order prescription of a smoothly varying density, which can describe both the density profile expected in dynamically expanding H II regions (Hosokawa & Inutsuka 2005) and in the interior of turbulent molecular clouds (Wolfire et al. 2010).

One major change as compared to Cormier et al. (2019) is that luminosities are compiled in a cumulative fashion, meaning that the we can access the intrinsic luminosity of each line at a given depth in the Cloudy model instead of considering only the total resulting luminosity, which obviously depends on the stopping criterion. This is a crucial step to study the escape fraction of ionizing photons as this parameter is sensitive to the stopping depth of the model, especially near the ionization front. As illustrated in Fig. 2, each initial Cloudy model is used to create 17 sub-models stopping at different A_V controlled by the "cut" parameter. The original model is cut at the inner radius (cut = 0), ionization front (cut = 1), H_2 dissociation front (cut = 2), CO dissociation front (cut = 3) and outer radius (cut = 4). To sample the different phases (H II region, PDR, CO-dark H_2 region and CO-emitting H_2 region) defined by those cuts, three additional cuts are added between each integer i (cut = i + 0.25, i + 0.5, i + 0.75), equally spaced in A_V between cut = i and cut = i + 1. We stress that stopping the model at a given cut and truncating it a posteriori are not strictly equivalent but our tests have shown that this is a secondary effect.

Our cut parameter is analogous to other parameters that were used to describe density-bounded models such as the Hβ fraction used in Stasińska et al. (2015) and Ramambason et al. (2020) or the zero-age optical depth to LyC photons (τ_λ) used in Lebouteiller et al. (2017) and Plat et al. (2019). Although the cut parameter does not have a physical meaning, it allows us to ensure a good sampling of the region near the ionization front and provides a simple characterization of density-bounded regions (cut < 1) vs. ionization-bounded models (cut > 1). Additionally, it is defined consistently regardless of the metallicity (as opposed to A_V).

Since each Cloudy model is cut a posteriori into sub-models, the stopping criterion is not crucial in this study. However, to incorporate the emission from different phases, models should be deep enough to include the neutral and molecular zone for each set of parameters. This transition is metallicity-dependent and requires going deep enough in A_V for the lowest metallicity models. Most of our Cloudy models are computed until they reach a maximum $A_V = 10$. Only the densest models cannot reach $A_V = 10$ because their electronic temperature drops below 10 K before reaching this optical depth.

3.2.4. Escape fraction from H II regions

We calculate the escape fraction of ionizing photons from H II regions by using the ionizing continuum that is provided by Cloudy at each depth in the model. This continuum saves the number of photons per frequency bins, at a given depth, for all energies greater than 1 Rydberg. We compute the escape fraction in a given range of energy as follow:

$$f_{\rm esc,HII}(E_1 \leq h\nu \leq E_2)(R) = \frac{L(E_1 \leq h\nu \leq E_2)(R)}{L(E_1 \leq h\nu \leq E_2)(R_{\rm in})}, \quad (3)$$

where the numerator is the luminosity produced by the central sources reaching the radius R with energy between E_1 and E_2, and the denominator is the luminosity impinging the cloud at the inner radius $R_{\rm in}$, within the same energy range. The number of ionizing photons at a given radius is calculated as:

$$L(E_1 \leq h\nu \leq E_2)(R) = 4\pi R^2 \int_{\nu_1}^{\nu_2} F_\nu(r){\rm d}\nu, \quad (4)$$

where F_ν in the flux of photons with frequency ν at a given radius. In practice we consider $f_{\rm esc,HII} = f_{\rm esc}(1{\rm Ryd} \leq h\nu \leq \infty)(R_{\rm cut})$ that includes all ionizing photons even in the X-ray regime, with $R_{\rm cut}$ the radius at which the sub-model is cut. This definition is equivalent to the ratio between the total observed emission below 912 Å and the intrinsic emission below 912 Å. We ensured that the sampling of the cut values results in a fine enough sampling for $f_{\rm esc,HII}$ as well. This definition of $f_{\rm esc,HII}$ consistently accounts for the absorption of photons in the gas and by dust. However, it assumes that the integrated emission of a galaxy is dominated by H II regions and that photons reaching the edge of a density-bounded region freely escape in the IGM. In practice, this tends to overestimate the global galactic escape fraction as some photons are likely to get reabsorbed by clumps or diffuse gas after having escaped from H II regions. This $f_{\rm esc,HII}$ is to be considered as the escape fraction from H II regions and cautiously compared to other measurements, except for the two regions in NGC 4124 (see Sect. 2.1). While this definition corresponds to the true escape fraction from our model, accounting for the total ionizing luminosity below the Lyman edge, it differs from the definition commonly used in observational studies to measure $f_{\rm esc}$(LyC), which most often relies on the flux ratio at 912 Å and 1500 Å. This difference should be taken into account when comparing with LyC measurements. This will be discussed in Sect. 7.

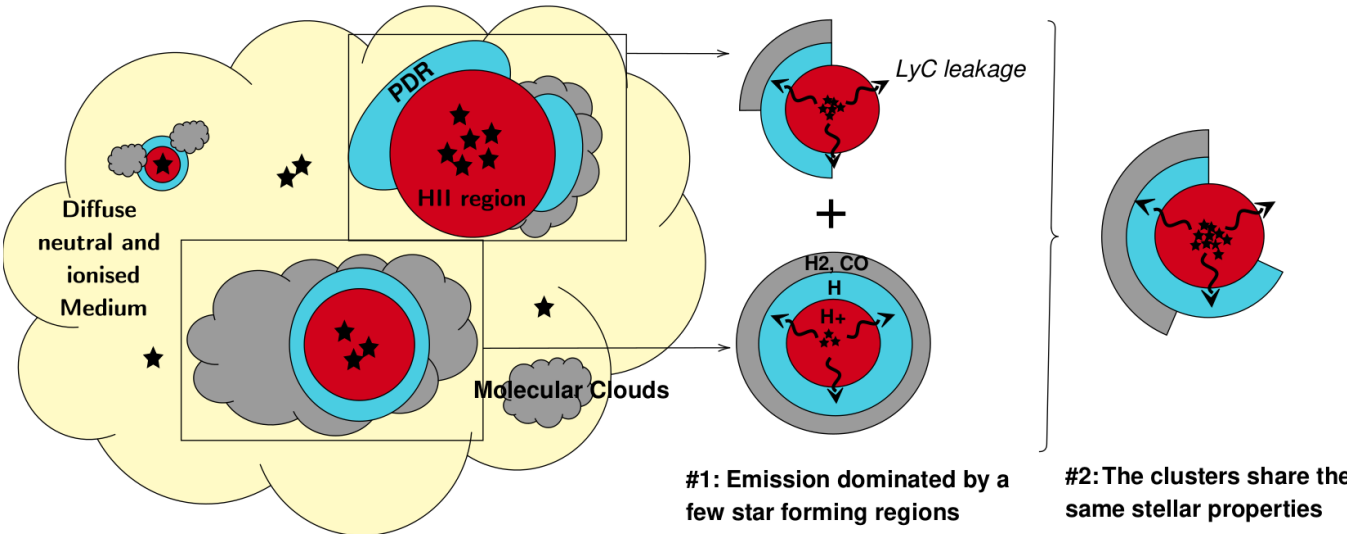

Fig. 3. Schematic view of the ISM of a starburst galaxy and associated representative topology.

4. MULTIGRIS runs

In this section we recall some of the most important features of the code used in the current study. A more detailed description of the code can be found in LR22.

4.1. Topological representation

The notion of topology introduced in Sect. 3.1 is especially important in the context of the escape fraction. Indeed, regardless of the nature of the physical mechanisms at the origin of escaping photons, it seems that the LyC escape fraction is tightly linked to the distribution of gas in the ISM. In particular, Gazagnes et al. (2020) have shown that the escape fractions of LyC and Lyα photons are quite sensitive to the distribution of neutral gas and dust, both at galactic scale and on specific lines of sight. This effect is even more pronounced for LyC photons that are easily absorbed by the neutral gas while Lyα photons scatter and are destroyed only by dust.

One of the main advantages of our grid of models compared to previous topological studies (see Sect. 1) is that it includes a free parameter that allows us to constrain the depth of each sector based on the observed line ratios (Sect. 3.2.3). This avoids having to fix an arbitrary stopping criteria that is not trivial to determine in studies that include PDRs and molecular regions (e.g., as in Cormier et al. 2019). The topological models aim to provide representations of the distribution of matter (number and covering factors of sectors) and phases (stopping depth) in a galaxy. The different phases we consider are represented in Fig. 3: an atomic ionized phase dominated by H II regions near star clusters, a neutral atomic phase dominated by a warm component around H II regions where hydrogen is photodissociated (PDR), and a neutral molecular phase concentrated in molecular clouds. Two phases are not accounted for in our Cloudy models: the DIG and the diffuse neutral gas (see discussion in Sect. 7).

In essence, topological models are not designed to determine the spatial distribution of the gas but only the relative contribution of each phase and sector. As shown in Fig. 3, we model a galaxy as a sum of representative star-forming regions that can be combined into one single representative model, assuming that the star clusters that dominate the emission share the same properties. In practice, we assume a single SED for each galaxy we model. This representation is especially well-suited for unresolved observations for which we precisely do not have access to spatial information. Additionally, most current photoionization codes cannot account for precise gas distribution that would require expensive 3D radiation transfer models. In our single cluster topological models, the emission lines are computed as a weighted linear combination of the emission from each sector as follows:

$$L = \sum_{\mathrm{i}=1}^{N_{\mathrm{sectors}}} w_{\mathrm{i}} L_{\mathrm{i}}, \tag{5}$$

where w_{i} is the mixing weight and L_{i} the predicted luminosity of a given line in the ith sector.

This topological approach allows predictions for any other physical quantities available in Cloudy as long as it can be expressed as an analytical combination of the observables from individual sectors. We use the same formula for "extensive" quantities, meaning, that scale with luminosity (e.g., gas masses in the different phases, number of ionizing photons Q). In the current study, we consider configurations having a single cluster and different number of sectors (from 1 to 3). Under this hypothesis of a single cluster, the $f_{\mathrm{esc,HII}}$ is also an extensive quantity and the averaged $f_{\mathrm{esc,HII}}$ can be derive as follows:

$$f_{\mathrm{esc,HII}} = \sum_{\mathrm{i}=1}^{N_{\mathrm{sectors}}} w_{\mathrm{i}} f^{\mathrm{i}}_{\mathrm{esc,HII}}. \tag{6}$$

For a configuration with N sectors, the free parameters correspond to N times the 8 free parameters in the Cloudy models from Table 2 determined for each sector plus the mixing weights w_i. In addition, the global luminosity scaling factor of the cluster is free. This amounts to a total of $9N+1$ parameters. The effective number of free parameters can be reduced by imposing priors that link some parameters together. To mimic the exposure of all sectors to a single cluster we impose that the stellar luminosity, X-ray luminosity, inner temperature of the of the X-ray-emitting accretion disk, metallicity, and age of the stellar burst are identical in all sectors. An additional constraint imposes that all w_{i} sum up to 1 (no hole in the model). The effective number of free parameters thus reduces to $4N+5$ parameters.

The last constraint on the sum of w_{i} translates into one major modeling assumption: we assume that the cluster is fully surrounded by the different sectors and that we can constrain its

total luminosity. If the cluster luminosity is unconstrained, the total luminosity scaling factor and the covering factors of each sector can be degenerate. In practice, this would result in quite different configurations that could all match a suite of emission lines but with different scaling factors. This caveat is particularly important for quantities that strongly vary with different configurations (in particular, the escape fraction) and will be discussed in Sect. 7. Hence, $L_{\rm TIR}$, which is used as a constraint (see Table 1), is essential in our analysis to constrain the scaling factor and match the observed total luminosity.

4.2. Inference of the parameters

MULTIGRIS can be used with different Markov chain Monte-Carlo (MCMC) samplers, which are compared in LR22. We used the Sequential Monte Carlo Sampler (SMC) that is adapted to multidimensional grids with several likelihood peaks. Interpolating on all parameters is computationally expensive and raises some issues for models located at the edges of our grid. Instead, we use a nearest neighbor interpolation on all the parameters of the grid (described in Table 2) except for the metallicity, for which we perform a linear interpolation.

We define the likelihood of our data $p(\boldsymbol{O}|\boldsymbol{\theta}, \mathcal{M})$, where $\boldsymbol{O}$ is the data and $\boldsymbol{\theta}$ the set of parameters of a model $\mathcal{M}$, by considering our suite of emission lines as independent identically distributed random variables (RV). Each RV is described as Gaussian distribution centered on the measured value and with a σ corresponding to the uncertainty of the observation. Hence, the likelihood can be expressed as:

$$\mathcal{L} = p(O|\boldsymbol{\theta}, \mathcal{M}) = \prod_{i=0}^{N_{\rm obs}} \mathcal{N}(\mu = O_{\rm i}, \sigma^2 = U_{\rm i}^2), \tag{7}$$

where N is the number of emission lines with observed fluxes $O_{\rm i}$ and uncertainties $U_{\rm i}$. For undetected lines with instrumental upper limits, the Gaussian distribution is replaced by a half-Gaussian. We consider that upper limits correspond to a 2σ signal. To avoid possible biases due to lines detected with unrealistically small uncertainties, we force a minimal uncertainty of 10% for all lines.

In practice, single-sector models should be preferred if they are able to simultaneously reproduce all the emission lines of a given galaxy. However, galaxies having numerous lines are more likely to require a greater number of sectors. To which extent the addition of a supplementary sector improves the agreement with observations needs to be quantitatively studied. To compare models, one needs to estimate how well a model performs at predicting data. To do so, we estimate the marginal likelihood corresponding to each configuration by evaluating the model at the posterior distribution of the parameters. The marginal likelihood is defined as follows:

$$\mathcal{L}_{\mathcal{M}} = \int_{\theta} p(O|\boldsymbol{\theta}, \mathcal{M}) {\rm d}\boldsymbol{\theta}. \tag{8}$$

Because of the integration on all the free parameters $\boldsymbol{\theta}$, this metric penalizes models that require a large number of sectors without significantly improving the agreement with the data. Although the code can be used with any number of components, those models become less and less likely to be selected by the marginal likelihood criterion. Additionally, the computing time required for the MCMC sampling step scales with the number of free parameters. We hence limit this study to combinations involving 1, 2, or 3 sectors. An example of a 4-sector model for I Zw 18, which includes additional constraints from the optical lines, is presented in LR22.

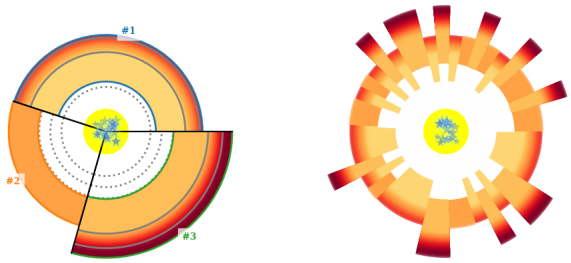

Fig. 4. Schematic view of the single cluster 3-sector configuration for He 2–10. The number of stars in the central cluster scales with the intensity of the radiation field, the proximity of the gas representing varying ionization parameters and the orange shading increases with density.

In Table A.1, we report the best (i.e., having the highest marginal likelihood) configuration for each galaxy, the corresponding accuracy at 3σ and the marginal likelihood values in logarithm, for a varying number of sectors. In practice, we combine the results obtained for configurations having 1, 2, or 3 sectors by considering a weighted mean of the 3 configurations with the weights are defined as follows:

$$\alpha_{\rm i\ sectors} = \frac{\mathcal{L}_{\mathcal{M},{\rm i}}}{\sum_{\rm j=1}^{3\ \rm sectors} \mathcal{L}_{\mathcal{M},{\rm j}}}, \tag{9}$$

where $\mathcal{L}_{\mathcal{M},{\rm i}}$ is the marginal likelihood associated with the configuration having i sectors (with $1 \leq i \leq 3$). This weighted combination allows us to account for configurations having marginal likelihoods that are close to the best configuration. Since the weighted combination is performed directly on the MCMC draws, the uncertainties we derive for the parameters incorporate uncertainties on the combination of configurations, which can be considered as an uncertainty on the model itself. Other metrics could be used to define the weight, which are detailed in LR22. Using a different metric yields variations for a few individual objects in our sample but does not affect the global trends that we discuss in the following sections.

As an illustration, Fig. 4 shows the best configuration found by MULTIGRIS for He 2–10. This is one of the galaxies for which we have the most constraints available (22 detections and 6 upper limits, see Fig. A.4). For this galaxy, the 3-sector configuration is favored. The left-hand side plot represents the relative contribution of each of the 3-sectors and the right-hand side plot shows another version where the sectors have been randomly redistributed around the central source. We stress that the two topologies are completely equivalent in our study.

4.3. Probability density functions

One of the advantages of using a Bayesian method is that the ensemble of solutions can be represented by a continuous distribution, which is well-suited to identify possible degeneracies or multimodal solution. While frequentist methods like the χ^2 only provide a group of the most likely values of a parameter, MULTIGRIS outputs a sample of draws from the posterior PDF $p(\boldsymbol{O}|\boldsymbol{\theta}, \mathcal{M})$ of any given parameter.

The PDF can be described using various estimators. The mean, mode, and median values of the parameters can be significantly different in the case of multimodal or asymetric distribution. Ideally, one would like to represent the full posterior PDFs that are used in this study to derive trends and correla-

tions. We also investigate the combined PDF concatenating all sources. To that purpose, we use kernel density estimate (KDE) plots, which provide a smoothed convolution (using a Gaussian kernel) of the 2D-PDF of our whole sample. Additionally, to represent 2D-PDFs of individual objects we adopt the skewed uncertainty ellipse (SUE) representation (see e.g., Appendix F from Galliano et al. 2021). A SUE represents the 1σ contour of a 2 dimensional split-normal distribution adjusted to have the same three first moments as the underlying PDF. The center of the SUE marks the location of the robust mean of the 2D-PDF. While the representation is convenient to locate the parameter space of highest probability (e.g., 1σ) for a given object, SUEs are not always representative of the underlying PDF, especially when several modes are present. In the following plots we show either the KDE of the full sample or the individual SUEs corresponding to each galaxy. While the KDE of the full sample should always be used to study the statistical trends and correlations, the individual SUEs can also be useful to visualize the region of maximal likelihood associated with each object.

5. Consistency checks

The escape fraction of ionizing photons is a complex parameter with multiple dependences. Before diving into the interpretation of this complex observable (see Sect. 6), we perform consistency checks to see how the predictions from MULTIGRIS compare to classical diagnostics found in the literature. In particular, since the escape fraction of ionization photons is sensitive to both the gas and stellar content of a galaxy, we first examine some key parameters that may control the predicted values of $f_{\rm esc,HII}$: the metallicity and the star-formation rate.

5.1. Metallicity estimates

To ensure that our metallicity estimates are not affected by the metallicity sampling of the model grid (see Sect. 3), we perform a linear interpolation on this parameter. For the other parameters, we take at each draw the nearest neighbor in the grid. Our approach differs from classical methods that usually rely on a single tracer or a combination of a few lines to derive the metallicity. Instead, our code uses the combined information from all emission lines to constrain model parameters, including the metallicity. Although this approach is more flexible (no need for a specific suite of emission lines), we need to rely on a particular abundance pattern because we do not have enough constraints on the abundance of all species. In this study, we use abundances tailored for compact, low-metallicity, dwarf galaxies, as described in Sect. 3.2.4.

The metallicity could be subject to multiple degeneracies. In particular, varying the cut parameter, that is, the A_V at which each sector stops, impacts the emission of tracers used to estimate the metallicity. Previous works have reported that accounting for escaping ionizing photons tends to bias the metallicity estimates toward lower values (Xiao et al. 2018; Jiang et al. 2019). The metallicity is also degenerate with the stellar age parameter. As mentioned in Xiao et al. (2018), an old stellar population with no leakage can produce line ratios similar to a younger stellar population where part of the radiation is leaking through density-bounded regions.

To avoid such degeneracies between metallicity, stellar age and cut parameters, we allow the metallicity to vary under a weakly informative prior. This ensures the inference preferentially starts around the mean of the prior and prevents drawing too far from it. We use a Gaussian centered at the measured metallicity with a large width parameter ($\sigma = 0.1 + \sigma_{\rm mes}$ in logarithmic scale, where $\sigma_{\rm mes}$ is the uncertainty associated with the measured metallicity). We emphasize that although we use only constraints in the IR domain, the prior on metallicity incorporates information derived from optical lines since it is centered on the metallicity estimates derived from optical strong lines method. Galaxies that do not fall within the $\sigma = 0.1$ dex envelop are galaxies for which the information given by all emission lines have pushed the posterior PDF away from the prior distribution.

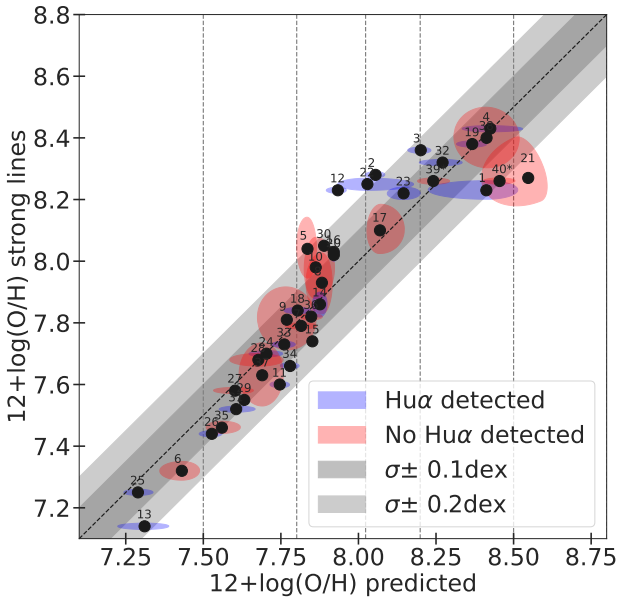

Fig. 5. Predicted metallicity of MULTIGRIS vs. measured metallicity from the strong lines R23-P method from Pilyugin & Thuan (2005) based on optical oxygen lines ratio. The black dots represent the robust means of the posterior PDF. The 1σ contours of the SUEs are calculated by simulating mock data for the y-axis with a dispersion corresponding to the measured uncertainty. The vertical dashed lines correspond to the metallicity bins of the grid. The galaxies are labeled with the numbers reported in Table A.1. The numbers that are flagged with a star correspond to two pointings in NGC 4214, which are excluded from the KDE. The color code indicates whether the Humphreys α line (7–6) is detected or not.

In Fig. 5, we compare this estimated metallicity to that measured from strong lines methods. The galaxies are labeled with the numbers reported in Table A.1. The measured metallicities from the literature come from Madden et al. (2013), and references therein, based mainly on the optical oxygen lines ratio (R23-P) method from Pilyugin & Thuan (2005). Their method is based on a two-parameter calibration involving the R23 ratio and an excitation parameter P, which corrects the estimates by accounting for the physical conditions in the H II regions. The method used in Madden et al. (2013) allows them to derive metallicity measurements for almost all the DGS galaxies except for three galaxies (HS 0017+1055, HS 0822+3542 and HS2351+2733) for which the R23-P method cannot be applied and the measured metallicity is derived using the T_e-direct method (Ugryumov et al. 2003; Izotov et al. 2006, 1994). We note that the latter method is more reliable than the R23-P method (e.g., Maiolino & Mannucci 2019) that was chosen only because it could be applied in a consistent way to almost all objects in the DGS sample.

The uncertainties derived (1σ uncertainties of SUEs defined in Sect. 4.3) from our method are somewhat larger than those found in the literature, as they incorporate uncertainties on all

emission lines and on the modeling (e.g., in particular our prescription for the abundance patterns vs. metallicity). When a hydrogen recombination line (e.g., the Humphreys α line (7–6); Huα) is available, the estimated metallicity tends to be more tightly constrained as the code can directly infer elemental abundances relative to hydrogen. This is illustrated in Fig. 5 where galaxies that have Huα detection tend to have smaller 1σ contours than the ones without. Interestingly, our code is also able to estimate metallicities without hydrogen recombination line measurements, in some cases with a precision comparable to galaxies with Huα measurements. As a consequence, the metallicity is inferred by constraining metal-to-oxygen abundances, which depend on metallicity in the grid (see Sect. 3). In this case the results we get are strongly dependent on the assumed metal-to-oxygen profiles, and in particular the N/O and C/O vs. O/H relations.

The metallicities are consistent with the corresponding strong line measurements within 0.1 dex for 31 out 40 galaxies and withing 0.2 dex for all galaxies except one (12: II Zw 40), for which we predict a significantly smaller metallicity than that derive with the R23-P method. We note, however, that our measurement is consistent with the value measured for this galaxy (8.09 ± 0.02) using the T_e-direct method from Izotov et al. (2006). The scatter we obtain is somewhat smaller than that derived for the DGS sample in Madden et al. (2013) when comparing different calibration methods using the R23 ratio (Pilyugin & Thuan 2005) and the T_e-direct method. (Izotov et al. 2006). We find that our predictions tend to systematically predict slightly higher metallicities than the R23-P method for metallicities below 7.8. This might be due to the fact that we find galaxies with numerous density-bounded regions (i.e., with small cut parameters) among the lowest metallicities in our sample (see Sect. 6.1). This effect might be linked to the fact that the R23-P diagnostic relies on an empirical calibration based on samples of galaxies, regardless of the presence or not of density-bounded H II regions. If the calibration is dominated by radiation-bounded H II regions, this diagnostic might slightly underestimate the metallicity of galaxies in which numerous density-bounded regions are present. Nevertheless, based on the weak prior assumption and on the emission lines given as constraints, we find that our code infers metallicities in agreement with previous measurements for all galaxies in our sample. This is a necessary condition to derive parameters such as $f_{\rm esc,HII}$ which is expected to be metallicity-sensitive.

5.2. Star-formation rate estimates

We now use the predictions provided by MULTIGRIS to infer SFR estimates in our sample and compare them to classical diagnostics. Although the SFR is not one of the primary (used for inference) or secondary (output from Cloudy) parameters, we use classical SFR proxies (Hα, $L_{\rm TIR}$ and $Q(H^0)$), which we convert into SFR. We emphasize that Hα observations are not used as constraints in our models but our code can provide PDFs of both observed and unobserved emission lines. In this work, we use the constraints from IR lines that, to first order, are not affected by extinction by dust, to estimate the number of ionizing photons produced by the central cluster, $Q(H^0)$. The predicted Hα luminosity depends of the incident flux $Q(H^0)$ and on the geometry of the gas in our representative model. The main advantage of our method is that it takes into account information coming from all the emission lines available. However, like for other diagnostics, our estimates ultimately depend on the assumption made regarding the stellar population.

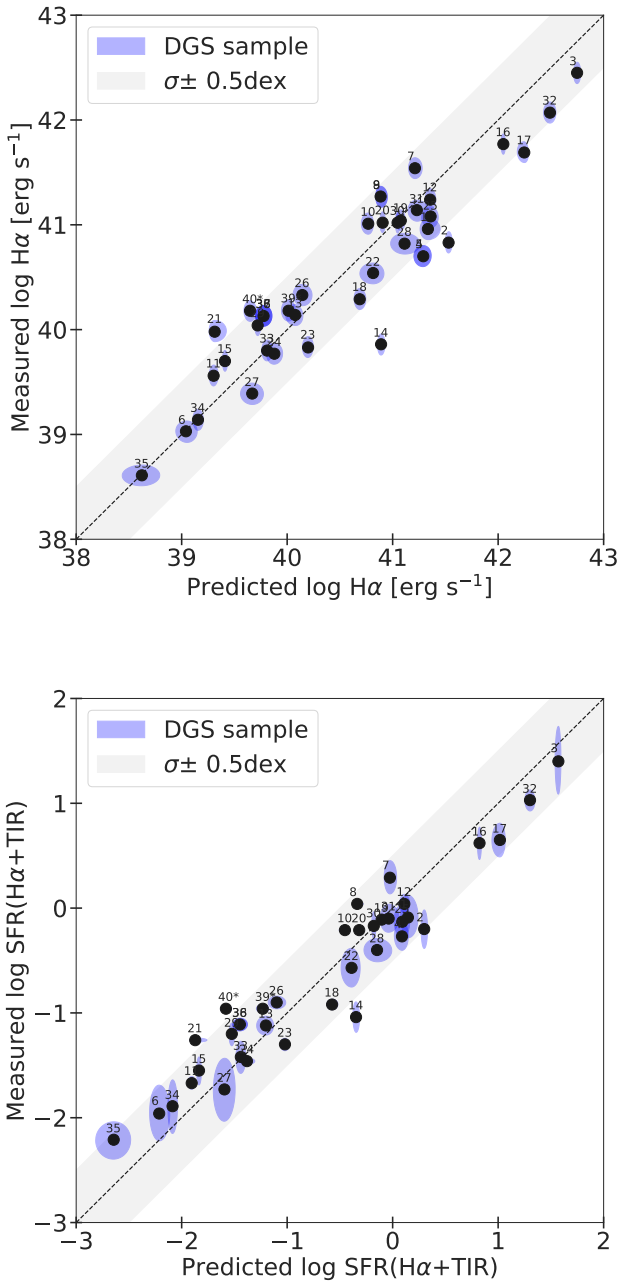

Fig. 6. Agreement between Hα and SFR predictions with empirical measurements. *Top*: predicted intrinsic Hα emission vs. measured Hα corrected for Galactic extinction from (Rémy-Ruyer et al. 2015, and references therein). *Bottom*: predicted SFR vs. SFR(Hα+$L_{\rm TIR}$) from Rémy-Ruyer et al. (2015). The 1σ contours of the SUEs are calculated by simulating mock data for the y-axis with a dispersion corresponding to the measured uncertainty. For 4 galaxies (HS 0017+1055, HS 0052+2536, HS 1319+3224 and HS 2352+2733) no Hα measurements are available. For 2 of those galaxies (5: HS 0052+2536 and 9: HS 1319+3224), their measured SFR is derived in using diagnostics based on FUV.

We compare our predictions to a classical diagnostic based on composite diagnostics that combine tracers of dust-obscured (e.g., $L_{\rm TIR}$) and dust-unobscured (e.g., Hα) star-formation (see review from Calzetti et al. 2012). First, we compare the predicted Hα luminosities to the values from the literature gathered in Rémy-Ruyer et al. (2015) in Fig. 6. The latter correspond to Hα luminosities corrected for underlying stellar absorption,

[N II]6548,6584Å lines contamination, and foreground Galactic extinction. No correction has been applied to account for the intrinsic attenuation within galaxies. For consistency, we compare the corrected measurements to the predicted intrinsic Hα emission i.e. what escapes from the galaxy without considering dust extinction.

The predicted Hα luminosity is compatible with observations within 0.5 dex for all galaxies in our sample except for two galaxies (14: Mrk 153 and 2: Haro 3) for which our prediction is significantly larger. The fact that our models find Hα values close to observed measurements is remarkable, especially since no optical lines were used as constraints. This means that the intrinsic values we predict are relatively robust even though we do not properly account for Hα emitted by the DIG (see Sect. 7). It also indicates that taking into account the actual geometry of the gas by including density-bounded regions in our models only yields minor variations in the predicted intrinsic Hα emission and remains compatible with observations using somewhat simple prescriptions to correct the observed Hα emission for foreground Galactic extinction. We stress that the intrinsic fluxes predicted by our models do not account for internal attenuation within the galaxy. The fact that our predictions are in agreement with observations that were not corrected for internal attenuation suggests that this additional attenuation term has only a minor effect in low-metallicity galaxies. This could, however, be an important effect for more metal-rich sources and may also explain part of the scatter see in Fig. 6.

We then use our line predictions to estimate the SFRs. To allow the comparison with the estimates provided in Rémy-Ruyer et al. (2015), we use the same empirical calibration from Kennicutt et al. (2009) based on mixed tracers (Hα and $L_{\rm TIR}$). This diagnostic corresponds to an SFR conversion assuming a Kroupa IMF[3] and with coefficients calibrated on the *Spitzer*-SINGS sample. For consistency, we use $L_{\rm TIR}$ luminosities predicted by our code. We note, however, that the observed $L_{\rm TIR}$ is used as a constraint and always well reproduced by our models within 0.1 dex. Similarly to the Hα prediction, we find that our predictions for the SFR are consistent with measurements from Rémy-Ruyer et al. (2015) within 0.5 dex for all galaxies but one (14: Mkr 153), which is linked to our overestimation of Hα emission[4].

One advantage of our approach is that it allows us to account for the photon losses due to escape fractions in density-bounded sectors, which are ignored in empirical calibrations. In Fig. 7 we convert the incident flux of ionizing photon, $Q(\mathrm{H}^0)$, into SFR using the analytical formula provided in Calzetti et al. (2012) assuming a Kroupa IMF. We then compare it to our prediction based on Hα + $L_{\rm TIR}$. Our SFR estimates based on $Q(\mathrm{H}^0)$ are in most cases shifted to higher values. If we consider SFR($Q(\mathrm{H}^0)$) as the intrinsic SFR, the correction to apply to SFR(Hα+TIR) follows the analytical curve 1/(1-$f_{\rm esc,HII}$) with small deviations, which can be attributed to several effects. First, although Hα emission is directly proportional to the number of LyC photons absorbed by the gas, $(1 - f_{\rm esc,HII})\mathrm{Q}(\mathrm{H}^0)$, this conversion depends on the hydrogen recombination coefficient that varies with physical conditions (density and electronic temperature). Second, the

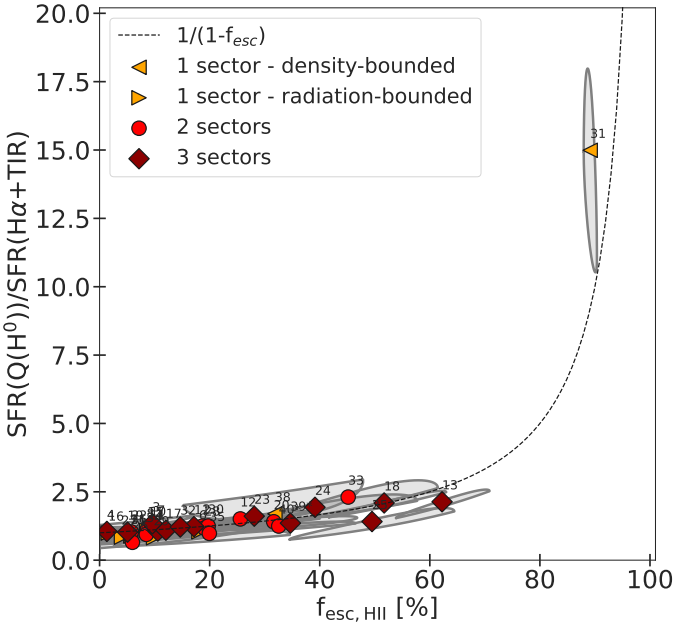

Fig. 7. Predicted SFR($Q(\mathrm{H}^0)$)/SFR(Hα+TIR) vs $f_{\rm esc,HII}$. The symbols represent the robust means of the posterior PDF and the gray contours show the 1σ uncertainties of the SUEs. The dashed line corresponds to the analytical curve 1/(1–$f_{\rm esc}$).

$L_{\rm TIR}$ emission that is used to derive the SFR has more complex dependences; it is sensitive to the dust content of our models and can be partially powered by X-ray photons heating the PDR. Hence, the SFR estimates we derive from Hα+TIR should, in theory, be sensitive to variations of the average physical and chemical conditions in each galaxies of our sample. Nevertheless, we find a rather narrow dispersion around the $1/(1 - f_{\rm esc,HII})$ relation that may indicate that, to first order, such effects are small. Assuming that SFR(Q(H^0)) is the intrinsic SFR accounting for all the photons produced by the central source, the correction to apply to SFR(Hα+TIR) varies from a factor 1 to ~2.5 for well constrained galaxies having at least two sectors, and reaches a factor ~15 for one poorly-constrained galaxy (31: Tol 1214-277) for which the best solution is a single-sector, completely density-bounded model (see Fig. 7).

The difference between the predicted SFR(Hα+TIR) and SFR($Q(\mathrm{H}^0)$) can therefore be explained by the presence of density-bounded regions leading to escaping photons. Indeed, photons that escape from density-bounded regions are accounted for in $Q(\mathrm{H}^0)$ but escape before ionizing the gas that produces Hα. Nevertheless, this effect is enhanced by our model assumptions (see Sect. 7) in which photons reaching the edge of density-bounded sectors can freely escape without further interaction. In principle, part of the leaking radiation should be reabsorbed and produce a DIG contributing to the global Hα emission (Mathis 1986; Sembach et al. 2000; Wood et al. 2010; Belfiore et al. 2022). This effect would likely reduce the correction factors that we predict here in order to obtain the intrinsic SFR($Q(\mathrm{H}^0)$). Assuming that all the ionizing photons escaping from H II regions are reabsorbed by the surrounding DIG and assuming a maximum value for the recombination coefficient of hydrogen (corresponding to densities typically found in H II region conditions), we can provide an upper limit on the additional Hα emission, which might be powered by leaky H II regions. In Fig. 8, we show the predictions for the maximum additional Hα emission powered by photon leakage,

[3] We apply a corrective factor of 0.945 (assuming that all the luminosity comes from stars with masses above 1 $M_\odot$ and that all the mass comes from stars with masses below 8 $M_\odot$) to our predicted SFR to compare them to measurements that assumed a Kroupa IMF.

[4] There is also a good agreement with the measured SFR(UV+24 μm). However, this comparison should be considered with caution due to the bursty nature of star formation in local metal-poor dwarfs (e.g., De Looze et al. 2014).

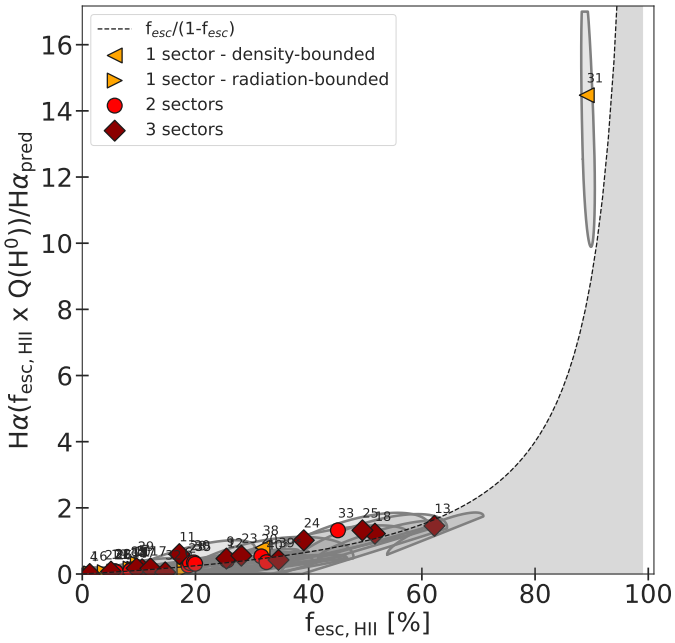

Fig. 8. Ratio of the maximum Hα emission associated with photon leakage from H II regions, assuming that all photons are reabsorbed and of the Hα value predicted by our models vs. $f_{\rm esc,HII}$. We assumed Q(H^0)-to-Hα conversion coefficient of 7.31×10^{11} for a case B recombination with T_e = 10 000 K and n = 100 cm^{-3} from Kennicutt et al. (1995) to estimate Hα($f_{\rm esc,HII}\times$ Q(H^0)). The H$\alpha_{\rm pred}$ values are computed consistently in the Cloudy models. The dashed line correspond to the analytical curve $f_{\rm esc}/(1-f_{\rm esc})$.

assuming a Q(H^0)-to-Hα conversion coefficient of 7.31×10^{11} for a case B recombination with T_e = 10 000 K and n = 100 cm^{-3} (Kennicutt et al. 1995) to derive Hα($f_{\rm esc,HII}\times$ Q(H^0)). This quantity follows the analytical curve $f_{\rm esc,HII}/(1-f_{\rm esc,HII})$. The small variations around this relation come from the fact that we assume a fixed recombination coefficient to estimate Hα($f_{\rm esc,HII}\times$Q(H^0)) while H$\alpha_{\rm pred}$ is calculated by integrating Hα emission at each depth, assuming a varying recombination coefficient that corresponds to the local physical conditions. We note that this upper limit was calculated with an unrealistically high recombination coefficient for DIG conditions and only yields an upper limit on the additional Hα emission. The range of plausible values corresponding to the reabsorption of leaky LyC-photons by DIG is represented by the gray shaded area. Unsurprisingly, the large $f_{\rm esc,HII}$ we derive translate into large upper limits of the Hα emission powered by DIG; this maximum emission exceeds 50% of the total Hα value for the larger $f_{\rm esc,HII}$ we infer (i.e., $f_{\rm esc,HII} > 50\%$). Although a significant fraction of the Hα emission is expected to be powered by the DIG (up to ~50–60%, see e.g., Oey et al. 2007; Lacerda et al. 2018), this is not sufficient to explain the $f_{\rm esc,HII} > 50\%$ that are found for a few objects in our sample. Most importantly, this additional Hα component can hardly be reconciled with the fact that our predicted Hα values (which do not account for the DIG) are in good agreement with the observed values (see Fig. 7). Possible explanations of this discrepancy are further discussed in Sect. 7.2.2.

6. Escape fraction of ionizing photons

We have seen in Sect. 5 that our new code is able to predict several galactic observables that are well constrained by the suite of lines given as inputs and compatible with previous observational studies. Those observables do not depend much on the inferred topology and can be estimated in a robust way, with little influence from, for example, the number of sectors considered (see LR22). The agreement of our predictions with empirical studies shows that simple prescriptions used in previous studies (e.g., single-component and no treatment of density-bounded regions) are sufficient, to first order, to estimate metallicities and SFR. Nevertheless, several other quantities can be predicted by our code, including more complex ones for which we have only a few, sometimes biased, observational proxies, and that strongly depend on the assumed topology. This possibility is especially interesting to study escape fractions for which direct observations are challenging and indirect proxies often yield different results (see Sect. 1). We now present the results that we obtain for $f_{\rm esc,HII}$ and discuss the observed trends and their possible physical interpretations.

6.1. Escape fraction vs. metallicity

In Fig. 9 (upper panel) we show the relation between $f_{\rm esc,HII}$ (defined in Sect. 3.2.4) and metallicity. The inferred mean $f_{\rm esc,HII}$ in our sample ranges from 0 up to around ~60%. Since we used the ionizing continuum above 13.6 eV, it also includes high energy photons that may propagate unabsorbed throughout the neutral gas. Due to such photons, the inferred escape fraction is never exactly zero, although very low for some objects. In practice, even for completely radiation-bounded models, we predict escape fractions of the order of 10^{-10}–10^{-6}. We find that 11 galaxies are best modeled with one single-sector. In most of those cases (9 out of 11), the best solution (meaning, having the highest marginal likelihood; see Sect. 4.2) is radiation-bounded. When such single-sector, radiation-bounded models are favored, the resulting escape fraction is biased toward low $f_{\rm esc,HII}$ values.

Density-bounded, single-sector models have been proved to not reproduce well the optical emission lines from H II regions (e.g., Stasińska et al. 2015; Ramambason et al. 2020). They are even less likely to match the observational constraints in the IR domain where some tracers arise from the PDR. For two galaxies, however, our code favors single-sector solutions that are completely density-bounded (38: HS 2352+2733 and 31: Tol 1214-277). HS 2352+2733 has only one detection ([O III]) and an upper limit on [C II], while for Tol 1214-277 most of the detections are lines arising from the ionized region and only [C II] and [O I] constrains the PDR. However, the errorbars on its [C II] and [O I] detections are rather large, hence allowing the code to favor a completely density-bounded solution since there is not enough information to constrain the neutral gas properties in those galaxies. For such galaxies, the predicted LyC-leakage can reach high values since photons escape directly from one single density-bounded sector. Thus, for all the galaxies for which the best model has only a single-sector (whether completely radiation-bounded galaxies or completely density-bounded), the $f_{\rm esc,HII}$ estimates that we derive are relatively uncertain, with a bias toward low values for radiation-bounded models and toward high values for density-bounded models. They can, however, be considered as lower and upper limits on the escape fraction. In the rest of the analysis, we focus on a subsample of 27 out of 40 galaxies for which the best models have at least two sectors and that correspond to galaxies covered in one single pointing to derive the main trends of $f_{\rm esc,HII}$ with galactic parameters. In practice, we plot the KDEs and correlation coefficients for this subsample of galaxies while the individual SUEs are shown for

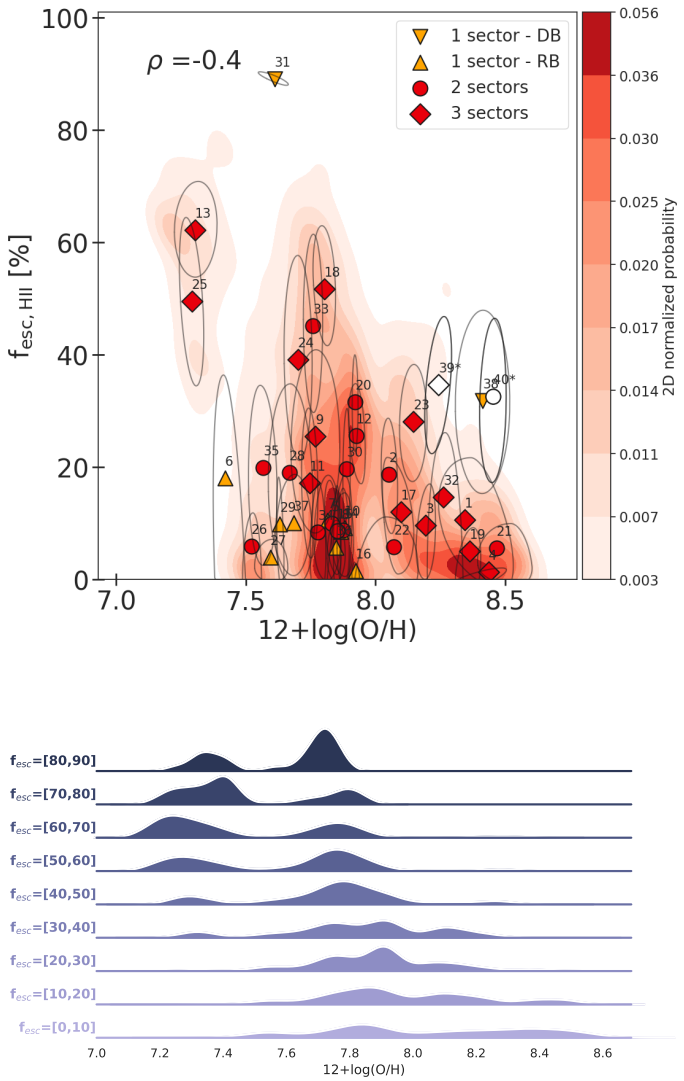

Fig. 9. Escape fraction dependence on metallicity. *Top*: escape fraction as a function of metallicity. The symbols and ellipses represent the robust means and associated uncertainties of the MCMC draws for individual galaxies, weighted by the marginal likelihood of each configuration (see Sect. 4.3). The symbols and colors indicate the configuration having the highest marginal likelihood for each galaxy. The underlying KDE represents the distribution of all the MCMC draws for galaxies whose best model has $N_{\rm sectors} \geq 2$ only and ρ the associated Spearman correlation coefficient. Orange triangles represent galaxies for which the best model is a single component either density-bounded (DB) or radiation-bounded (RB). Their values correspond respectively to upper and lower limits of $f_{\rm esc,HII}$. The white symbols that are flagged with a star correspond to two pointings in NGC 4214, which are excluded from the KDE. *Bottom*: KDE per bins of $f_{\rm esc,HII}$ for galaxies with at least two sectors.

the whole sample, including single-sector models and the two pointings in NGC 4214. Although the KDE accounts for the full posterior probability distribution described by all the MCMC draws for each galaxy, we note that our sample remains stastically limited and that the trends derived should be taken with caution.

The predicted $f_{\rm esc,HII}$ spans a wide range of values, which appear to be linked to the metallicity of the galaxy. By examining the KDE of the galaxies that are best constrained, we find a global trend of an increasing $f_{\rm esc,HII}$ with decreasing metallicity. More specifically, in Fig. 9 (upper panel), the upper right part of the KDE is completely unpopulated meaning that high-metallicity galaxies have very low probability of reaching high values of $f_{\rm esc,HII}$. The dispersion increases with lower metallicity. While low-metallicity galaxies do not systematically exhibit leakage, if they do, they are more likely to reach high values of $f_{\rm esc,HII}$. In Fig. 9 (bottom panel), we show 1D-KDE plots of the metallicity by bins of $f_{\rm esc,HII}$, for galaxies having at least 2 sectors. While this representation is quite sensitive to the intrinsic metallicity distribution of our sample, it provides a clear visual representation of the $f_{\rm esc,HII}$-metallicity relation. Although we find rather large distributions of metallicity for most of the $f_{\rm esc,HII}$ bins, we see that the main peaks in metallicity are shifted toward lower metallicities when the escape fraction increases. We note that the current sample is poorly populated in the low-metallicity regime (only 3 galaxies below 12 + log(O/H) = 7.5) and does not allow us to derive a robust trend with metallicity. Nevertheless, we find enhanced values of $f_{\rm esc,HII}$ in the low-metallicity regime, in which all the highest predicted $f_{\rm esc,HII}$ are found. Future observations of galaxies in the low-metallicity regime may reveal objects with small $f_{\rm esc,HII}$. The large scatter observed among galaxies that fall within a given metallicity bin may be linked to dependences in other parameters, which will be further examined in Sect. 6.2. The metallicity-$f_{\rm esc,HII}$ trend we present here remains clearly visible when we force the number of sectors to be the same for all galaxies (i.e., no selection of the best configuration) as shown in Fig. A.3. By looking at the individual PDFs represented by the SUEs, we confirm that the highest mean $f_{\rm esc,HII}$ are associated with the lowest metallicities in our sample. Specifically we infer $f_{\rm esc,HII} > 40\%$ in 5 galaxies with metallicities below 7.8. and $f_{\rm esc,HII} > 50\%$ in 2 galaxies with metallicities below 7.3. However, galaxies with $f_{\rm esc,HII}$ below 10% are found at all metallicities above 7.5 and are not preferentially associated with high metallicities. We also note that some objects exhibit somewhat high escape fractions on the high-metallicity part of our plot (e.g., 39: NGC 4214-c and 40: NGC 4214-s). However, those two objects correspond to separate pointings within one galaxy instead of an integrated measurement. They will be further examined in Sect. 7.

The $f_{\rm esc,HII}$-metallicity relation presented in this section is reminiscent of the PDR covering factor vs. metallicity relation presented in Cormier et al. (2019), that show that the PDR covering factor (calculated as a common scaling factor of PDR lines) correlates with metallicity. It is also globally consistent with observational studies in which LyC-leaking galaxies are predominantly found at low-metallicity, especially for the most extreme escape fractions. (e.g., Leitet et al. 2013; Izotov et al. 2018a; Flury et al. 2022b). In the recent Low-redshift Lyman Continuum Survey, which gathers 89 LyC measurements and associated metallicity estimates, Flury et al. (2022b) find that the strong LyC-emitters with measured $f_{\rm esc} > 10\%$ are all found at metallicities below 8.1. However, they report a significant scatter in the $f_{\rm esc}$-metallicity relation and do not observe any significant correlation for their whole sample. We further discuss possible explanations of this scatter in Sect. 7.

Several, possibly combined, effects might lead to the $f_{\rm esc,HII}$-metallicity trend that we observe: first, the UV opacity of the diffuse gas in the ISM is strongly dependent on metallicity. This UV opacity is controlled, in particular, by the dust content of galaxies since dust grains have been shown to absorb a significant (<50%) part of the LyC radiation in H II regions (Inoue et al. 2001). Hence, in the low-metallicity regime where the dust-to-gas mass ratio drops steeply (Rémy-Ruyer et al. 2014; Galliano et al. 2021), the ISM becomes more transparent and photons can travel larger distances. Second, the escape of

LyC photons from H II regions also depends on the properties of the host molecular clouds in which stars form. Molecular clouds are optically thick in the UV at all metallicities. Therefore, the escape fraction strongly varies with their covering factor. Recent simulations (e.g., Kimm et al. 2019; Yoo et al. 2020) have shown that the neutral gas covering factor is sensitive to both metallicity and stellar age effects with ionizing photons escaping more efficiently from their birth cloud when the metallicity is lower. Yoo et al. (2020) emphasize that the metallicity of the cloud in which stars form seems to be an important driver of the cloud disruption timescales. They find that in the low-metallicity runs, stars break out of their birth cloud within a few megayears only while at higher metallicities they tend to stay enshrouded for a longer period.

6.2. Dependences on other parameters

We further examine the $f_{esc,HII}$-metallicity relation by looking at possible secondary dependences that would explain the dispersion of this relation. In particular, we focus on examining the effect of various properties of the radiation sources: stellar mass, SFR, specific SFR (sSFR = SFR/M_*), age of the stellar burst, and X-ray-to-stellar luminosity ratio. We revisit the relation from Fig. 9 but split our sample by bins of values for secondary parameters (see Fig. A.6). We note that the various parameters that we examine are all interdependent. Correlations between those fundamental parameters have been extensively studied in samples of local and high-redshift galaxies (e.g., Nakajima & Ouchi 2014; Duarte Puertas et al. 2022) and, more specifically, among compact star-forming galaxies (Izotov et al. 2021; Flury et al. 2022b). In particular, the stellar mass, SFR, and sSFR are functions of the metallicity. Hence, we find that the highest $f_{esc,HII}$ that we infer are associated with low-metallicity, low stellar masses, and high sSFR. We find that those three parameters are the main drivers of large $f_{esc,HII}$ predicted in our sample. However, our analysis is limited by the observational bias of our sample in which only highly star-forming galaxies are populating the low-metallicity and low mass regime. With all those correlated parameters, it is not possible to infer which one has the strongest effect on $f_{esc,HII}$. We also find that stellar populations with ages above 3 Myr and an X-ray-to-stellar luminosity ratio above 0.5% (−2.3 in log) seem to be associated with enhanced escape fractions (see panels 4 and 5 from Fig. A.6). These results motivate the study of the $f_{esc,HII}$ dependences on galactic parameters other than metallicity, which we now examine in more details.

6.2.1. Stellar mass

It has been suggested that galaxies with shallow gravitational wells, associated with low stellar masses, would be good candidates for potential LyC-leakage. In particular, this $f_{esc} - M_*$ trend has been observed in samples with masses down to 10^8 $M_\odot$ (Leitet et al. 2013; Izotov et al. 2018b). However, a recent follow up study carried out in Izotov et al. (2021) targeting galaxies with masses below 10^8 $M_\odot$ finds no clear correlation between the global measured LyC escape fraction and the stellar masses.

In Fig. 10 (panel a), we find that large $f_{esc,HII}$ tend to be associated with low stellar masses, although with a large scatter. All the galaxies with masses above 10^9 $M_\odot$ have mean escape fractions below 40%. On the other hand, galaxies with masses below 10^9 $M_\odot$ do not always have high escape fractions. We note that this effect might be dominated by the mass-metallicity distribution of our sample in which lower masses are associated with the lowest metallicity. The correlation between $f_{esc,HII}$ and metallicity is, nevertheless, much clearer. Globally we observe a weak anti-correlation between $f_{esc,HII}$ and M_*, although with a large scatter, especially in the low-mass regime.

6.2.2. Star-formation rates

Recent simulations of photon leakage have shown that the process is a feedback driven mechanism (Trebitsch et al. 2017; Kimm et al. 2017, 2019; Kim et al. 2018, 2019, 2021; Kakiichi & Gronke 2021; Yoo et al. 2020). Although the main driver, whether stellar or supernovae (SN) feedback, remains elusive, the escape fraction is predicted to increase drastically when the star-formation efficiency is enhanced.

Surprisingly, we find no correlation between $f_{esc,HII}$ and SFR (see panel b from Fig. 10). The highest values of $f_{esc,HII}$ are associated with moderately star-forming galaxies in our sample. This absence of trend can be understood by looking at the KDE in Fig. A.6 (first panel); the highest SFRs in our sample correspond to the highest metallicities, which do not allow high amounts of photon leakage. On the other hand, we find that $f_{esc,HII}$ increases with increasing sSFR. We show on Fig. 10 the evolution of $f_{esc,HII}$ with respect to measured sSFR (panel c) and estimated sSFR from $Q(H^0)$ (panel f) as defined in Sect. 5.1. Although the trend is already visible when using the sSFR measurements from the literature, we find, not surprisingly, that applying a correction to account for the photon leakage through density-bounded sectors increases the dynamical range of the relation. Similarly to the M_*-$f_{esc,HII}$ relation and SFR-$f_{esc,HII}$ relation, the sSFR-$f_{esc,HII}$ relation is possibly driven by metallicity effects as the lower-metallicity sources are those with higher sSFR.

6.2.3. Stellar ages

Stellar evolution and the feedback mechanisms responsible for clouds dispersal may also impact the escape fraction. In our code, we only model very young bursts of ages below 10 Myr that represent the bulk of the H II regions that dominate the emission. We expect both rotation and binary stars to affect our prediction for the H II region escape fractions by producing more LyC photons for a longer period of time (Choi et al. 2020, and reference therein). In particular, Choi et al. (2020) finds that at low-metallicity (<1/3 $Z_\odot$) models that include binary stars produce 60% more ionizing flux compare to model without.

In Fig. 10 (panel d), we see that most of the stellar populations in our sample are dominated by young starbursts with ages of 2 to 6 Myr. These timescales suggest that leakage might occur at very early stage of stellar evolution. We find no clear trend of $f_{esc,HII}$ with stellar age, although large $f_{esc,HII}$ above ~50% seem to be associated with ages greater than 3–4 Myr. We emphasize, however, that large ages are not necessarily associated with leaking galaxies. As previously mentioned in Sect. 5.1, we note that the age and metallicity are interdependent, with larger ages tending to be associated with the lowest metallicities.

While our findings are consistent with scenarios in which photons might escape at very young ages (e.g., Kakiichi & Gronke 2021), we emphasize that our modeling assumptions based on a single burst stellar population prevent us from examining the effect of an older underlying population. Other sources of feedback (e.g., turbulence, SN feedback) might also play a role at later stellar ages. While at early ages, the feedback is dominated by pressure and ionization mechanisms from young H II regions, older stars can also contribute through stellar winds, especially in the Wolf-Rayet phase. We note that this

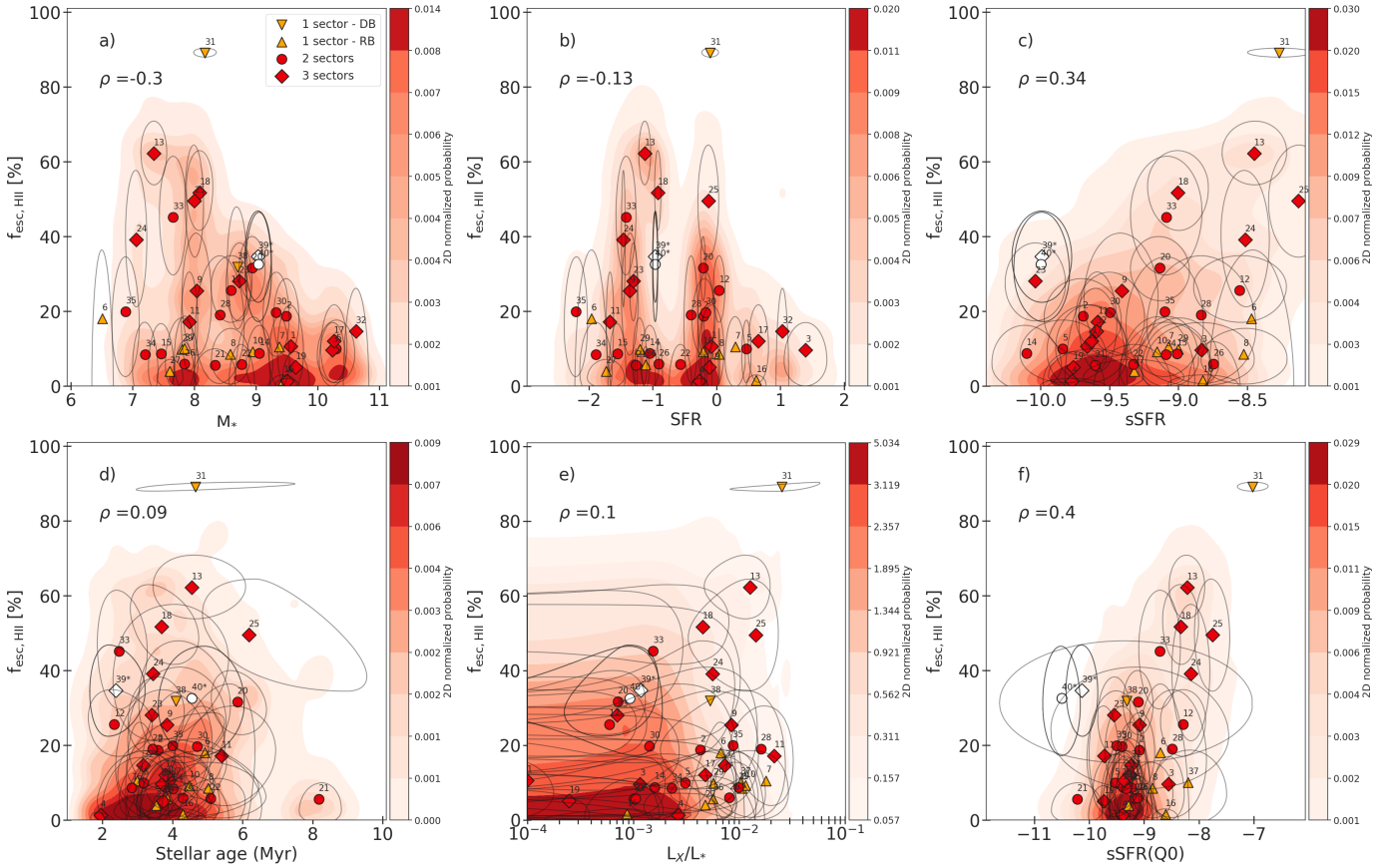

Fig. 10. Escape fraction dependences on measured and inferred galactic parameters. *Top panel*: escape fraction vs. some measured galactic parameters: M_*, SFR and measured sSFR. *Bottom panel*: escape fraction vs. some inferred galactic parameters: stellar age, Lx, sSFR($Q(H^0)$). The underlying KDE represents the distribution of all the MCMC draws for galaxies with $n_{\rm sectors} \geq 2$ only and ρ the associated Spearman correlation coefficient. Orange triangles represent galaxies for which the best model is a single component either density-bounded (DB) or radiation-bounded (RB). The white symbols that are flagged with a star correspond to two pointings in NGC 4214, which are excluded from the KDE.

stellar evolution also depends on metallicity effects with low-metallicity stars producing more ionizing photons over larger timescales (Eldridge et al. 2017; Xiao et al. 2018). Finally, older generation of stars might be responsible for creating low-density channels in the ISM, either due to mechanical feedback of the starburst (e.g., Ma et al. 2020) or through SN explosions. The presence of such low-density channels carved by previous generations of stars might favor leakage associated with young bursts that formed in regions where gas has already been partially cleared out (e.g., Ma et al. 2020). Nevertheless, it is worth noting that the stellar ages we derive are consistent with the disruption timescales found in simulations from Yoo et al. (2020). They are also very close to the feedback timescales of massive O/B stars measured in resolved giant molecular clouds (GMCs) from nearby spiral galaxies (e.g., Chevance et al. 2020a, 2022; Kim et al. 2022). Chevance et al. (2022) also point out that the efficiency at which the injected energy couples with the GMCs is low (a few tens of percent) and conclude that most of the energy and momentum produced by young stars escape outside of the host GMC, which is consistent with the high $f_{\rm esc,HII}$ we derive.

6.2.4. X-ray sources

As discussed in Sect. 2.1, several bright X-ray sources have been observed in our sample. Due to our simple modeling assumption and the suite of lines used in this study, the X-ray luminosities we derive are not tightly constrained. One of the motivations to include additional feedback mechanisms is the presence in our sample of extreme emission lines emitted by species with very high ionization potentials (>54 eV): [O IV] or [Ne V]. Such lines are, nevertheless, not numerous and not available for all our sample (see Fig. A.4). As a consequence, for most of the galaxies, the 1σ SUEs are large and the X-ray luminosity is not tightly constrained.

We note, however, that the X-ray-to-stellar luminosity ratio is also constrained by emission-lines in the neutral gas. This is due to the penetration of X-ray photons deep into the neutral and molecular gas because of their small photoionization cross-sections. Such photons are able to deposit their energy at large A_V, beyond the ionization front of H II regions. Because of this effect, our code is able to place an upper limit on the X-ray-to-stellar luminosity ratio based only on the emission lines arising from the neutral gas (e.g., [Fe II], [Si II], [O I]), even when no high ionization lines are detected. In particular, photoelectric heating on dust grains contributes to the heating of neutral gas in our models and is controlled by the assumed dust-to-gas mass ratio (fixed in our models) and total stellar luminosity (constrained by H II region emission lines). Hence, the additional X-ray contribution to neutral gas heating is limited by the emission lines in the neutral gas.

Although we derive uncertainties on the absolute log L_X above 1 dex for most galaxies, we predict a relatively well-constrained X-ray luminosity for a few galaxies. Specifically, we obtain a total X-ray luminosity of log $L_X = 40.33^{+0.53}_{-0.49}$ erg s^{-1}

for I Zw 18[5]. The value obtained for I Zw 18 is slightly higher but compatible with the measurement of 1×10^{40} erg s^{-1} (intrinsic X-ray luminosity; 0.3–10 keV) reported in Kaaret & Feng (2013). In our case, the predictions are for the total X-ray luminosity integrated over the whole energy range covered by the multicolor blackbody assumed as an input. While the X-ray-to-stellar luminosity ratio may allow us to pinpoint some of the most extreme escape fractions in our sample, we note, however, that no global trend seems to be visible, with some galaxies having little to no contribution coming from X-ray being associated with large escape fractions.

6.3. Tracers of density-bounded regions and escaping photons

Our code allows us to disentangle the emission of tracers that emit throughout different phases of the ISM. This is in particular the case of the [C II]158 µm line, which is emitted both in the ionized and neutral gas phase. This impacts the [O III]88 µm/[C II]158 µm (O3C2) diagnostic as a tracer of f_{esc}, which is quite popular in high-redshift studies (e.g., Harikane et al. 2020; Katz et al. 2022b), as it can be detected with ALMA up to redshifts above 6. This O3C2 ratio has been extensively studied with results indicating that high O3C2 values could be explained by specific physical conditions such as a high burstiness parameter (Vallini et al. 2021) or a low covering factor of PDR (Harikane et al. 2020). Such a trend has also been reported in Chevance et al. (2016, 2020b) that find that O3C2 is sensitive to the PDR filling factor defined for each pixel in their resolved analysis of 30Dor in the LMC. Nevertheless, at galactic scale, Cormier et al. (2019) find little correlation of O3C2 with their estimated PDR covering factors in the DGS. We now discuss the complex origin of the [C II] line and how it might affect the predictive power of the O3C2 line ratio. We then discuss the potential use of other line ratios to trace the presence of density-bounded regions and of associated LyC-photon leakage from H II regions, both in the IR and optical domain.

6.3.1. The complex origin of [C II] 158 µm emission

The [C II]158 µm line is an important diagnostic, which constrains in principle the neutral gas reservoir and the potential presence of density-bounded regions. However, the interpretation of this tracer is challenging as [C II] may be produced both throughout the neutral and ionized regions (e.g., Oberst et al. 2011; Ramos Padilla et al. 2021, 2022). In this section, we intend to disentangle the [C II] contribution from different phases by examining predictions for the fraction of [C II] luminosity arising from the ionized phase of the ISM (f[C II]$_{H+}$).

To do so, we weigh the cumulative [C II] line luminosity by the relative abundance profile of the hydrogen in each phase. For the ionized phase, we hence scale the cumulative line luminosity by the H^+ relative abundance profile and we integrate over the depth for each sector. The total [C II] luminosity arising from the ionized phase is obtained by summing over all the sectors of the model. We then derive f[C II]$_{H+}$ by computing the ratio of the prediction for [C II] in the ionized phase over the prediction for [C II] total luminosity.

In Fig. 11, we note that for one of the 1-sector, density-bounded galaxies (31: HS 2352+2733), log f[C II]$_{H+}$ = 0 since the gas in those models is completely ionized. For the other galaxies, log f[C II]$_{H+}$ varies between −3 to −0.5 and its value correlates with the metallicity of the galaxy. This confirms previous results from Croxall et al. (2017) and Cormier et al. (2019) that show that the fraction of [C II] arising from the ionized phase decreases at low metallicities. We also find values in the same dynamical range as previously found in Cormier et al. (2019) but with a somewhat smaller dispersion. According to Croxall et al. (2017), this result might be linked to the presence of harder radiation fields in low-metallicity environments. As a result, the relative abundance of C^{2+} in the ionized gas may increase with respect to the C^+ abundance, leading the [C II] emission to be dominated by the neutral phases. We discuss the implications on the use of [C II] as a potential tracer of the escape fraction in Sects. 6.3.2 and 7.

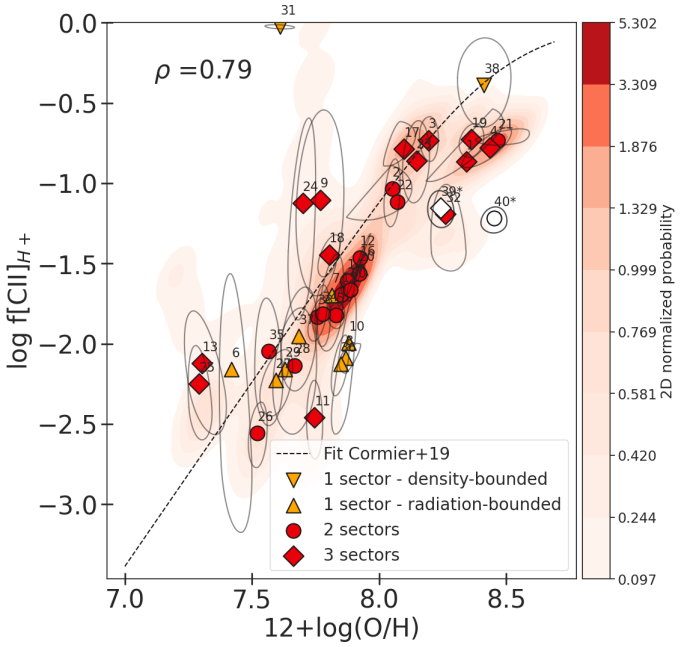

Fig. 11. Fraction of [C II]158 µm in the ionized phase vs. metallicity. The single-sector, density-bounded models represented by orange stars pointing down are PDR-deprived; all their [C II] emission arises from the ionized gas (log f[C II]$_{H+}$ ∼ 0). The underlying KDE represents the distribution of all the MCMC draws for galaxies with $N_{sectors} \geq 2$ only and ρ the associated Spearman correlation coefficient. The dashed line represents the relation derived by Cormier et al. (2019) for the DGS, assuming a fixed electron density $n_e = 30\,cm^{-3}$.

6.3.2. IR line ratios

We now provide some observed or unobserved tracers of the escape fraction based on what is predicted by our multisector models. The MCMC algorithm used in our code (see Sect. 4.2) does not only sample the PDFs of primary parameters but can also infer the PDFs of secondary parameters and unobserved tracers. In particular, we can extract predictions for unobserved line fluxes and the associated uncertainties that we represent using the same formalism as for primary parameters (i.e., SUEs and KDE, see Sect. 4.3).

Line emission produced by ions with different ionization potentials allow us to trace different physical depths. Hence, ratios involving lines emitted closer to the stars, in highly excited gas and lines emitted further away in regions near or past the ionization front provide an efficient way to probe the structure of the gas around a cluster of stars. This method can help identify galaxies in which numerous density-bounded regions

[5] The interval corresponds to the highest density interval at 94%.

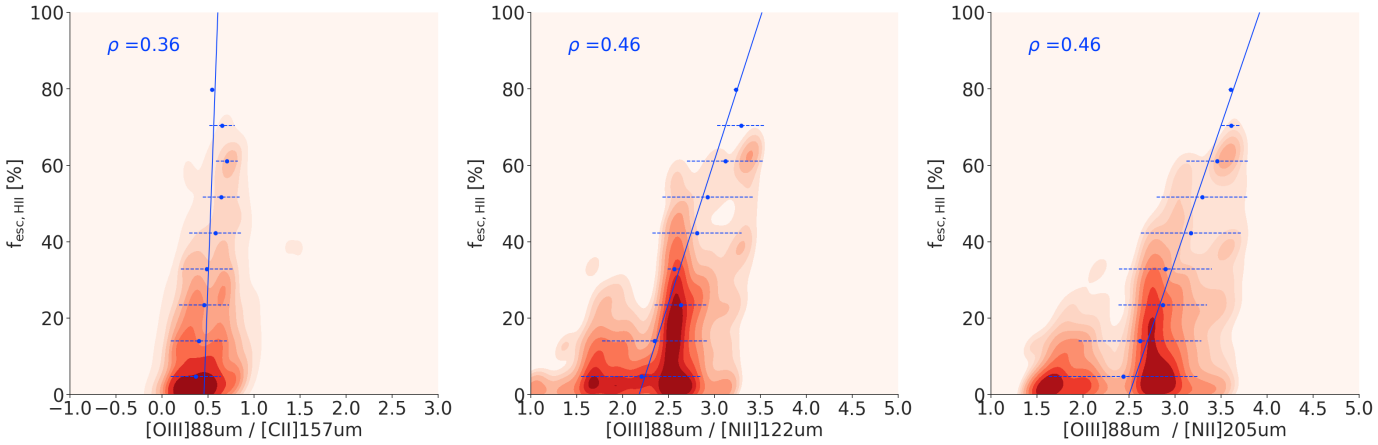

Fig. 12. KDE showing $f_{esc,HII}$ as a function of IR ratios. The KDE represent the distribution of all the MCMC draws for galaxies with $N \geq 2$ only and ρ the associated Spearman correlation coefficient. For this, we group the MCMC draws for our selected sample by bins of $f_{esc,HII}$ and fit each bin with a gaussian $\mathcal{N}(x_0, \sigma^2)$. The blue dots and dashed lines respectively represent the gaussian mean x_0 and the gaussian width σ. The plain blue line shows a weighted linear regression applied to the gaussian centers, accounting for the widths of each bin. The corresponding fit values are provided in Table A.2.

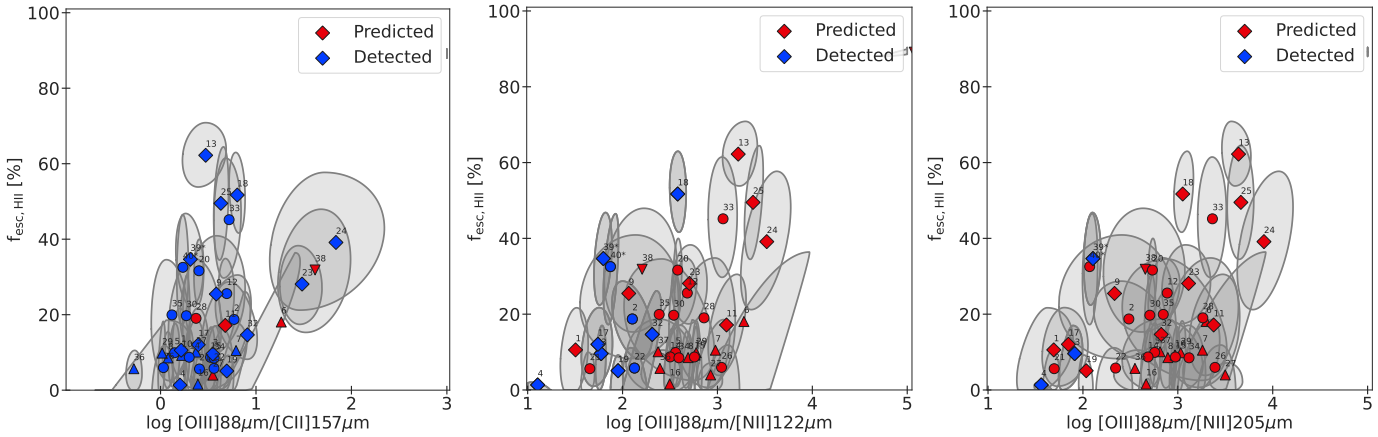

Fig. 13. Individual SUEs showing $f_{esc,HII}$ as a function of IR ratios. We note that Tol 1214-277 (31) most often lies outside of the range shown in x-axis. This is due to the lack of constraints on this galaxy for which the best model (one sector, completely density-bounded) predicts unrealistically low emission from ions with low ionisation potential.

are present. We expect such regions, which are bounded by the lack of gas, to have a reduced emission of low ionization lines emitted further away from the stars. We show in Figs. 12–15 the line fluxes predicted by our models for all lines, regardless of whether or not they are detected. We note, however, than when a line is detected and used as a constraint, the prediction should be close to the observed value.

In Figs. 12–13 we show how classical ratios of IR lines relate with $f_{esc,HII}$. In particular, we examine the O3C2 ratio, which involves the [C II]158 µm whose complex origin was discussed in the previous subsection. In the framework of our multisector modeling, we do find a correlation between O3C2 and $f_{esc,HII}$. Although this correlation is not clearly seen by looking at individual SUEs, it is visible on the KDE representing the whole sample (i.e., the concatenated MCMC draws for all the galaxies with at least 2 sectors). We emphasize, however, that for most galaxies the dynamical range of the O3C2 in our sample is rather narrow (<2 dex), which do not allow us to easily identify leakers based on their O3C2 ratio.

Although the [N II] lines are fainter and more challenging to detect, we find that [O III]88 µm/[N II]122 µm and [O III]88 µm/[N II]205 µm (O3N2) are better tracers of $f_{esc,HII}$ both because of the tightness of the relation and of their wider dynamical range (over 3 orders of magnitudes), which makes them potentially interesting diagnostics for observation-based classification. These correlations might be explained by the fact that both [O III] and [N II] originate from H II regions but are emitted by ions with different ionization potentials (35.1 eV for O^{2+} and 14.5 eV for N^{+}). This means that [O III] is predominantly emitted in a highly irradiated region close to the stars while [N II] arises from the outer part of the H II region, and hence their ratio depends on the depth of each sector. On the other hand, [C II] has a complex origin and arises both from H II regions and PDRs. As shown in Fig. 11, only a fraction of [C II] arises from the ionized gas. As one goes deeper in the neutral gas phase beyond the ionization front, the [C II] emission increases while $f_{esc,HII}$ remains unchanged. Hence, in low-metallicity environments in which [C II] is predominantly emitted in the neutral and molecular phases, the O3N2 ratio could be a better tracer of density-bounded H II regions as it is less contaminated by PDR emission than O3C2. We emphasize that, despite their interest for observational studies, the ratios presented here are strongly sensitive to the abundance profiles (C/O and N/O) assumed in this study. Other tracers, including ratios of the same

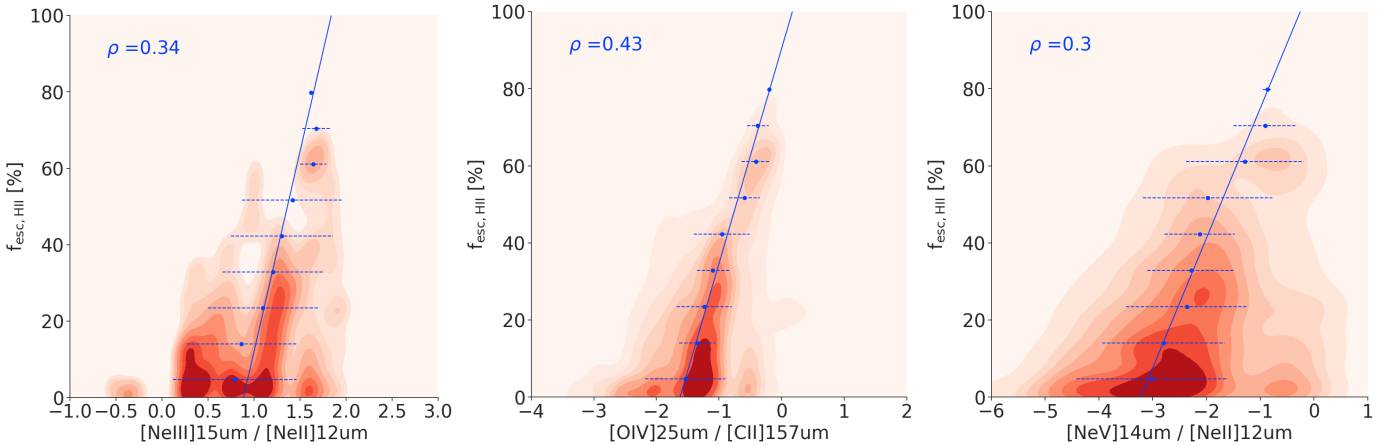

Fig. 14. KDE showing $f_{esc,HII}$ as a function of line ratio involving ions with high ionization potentials.

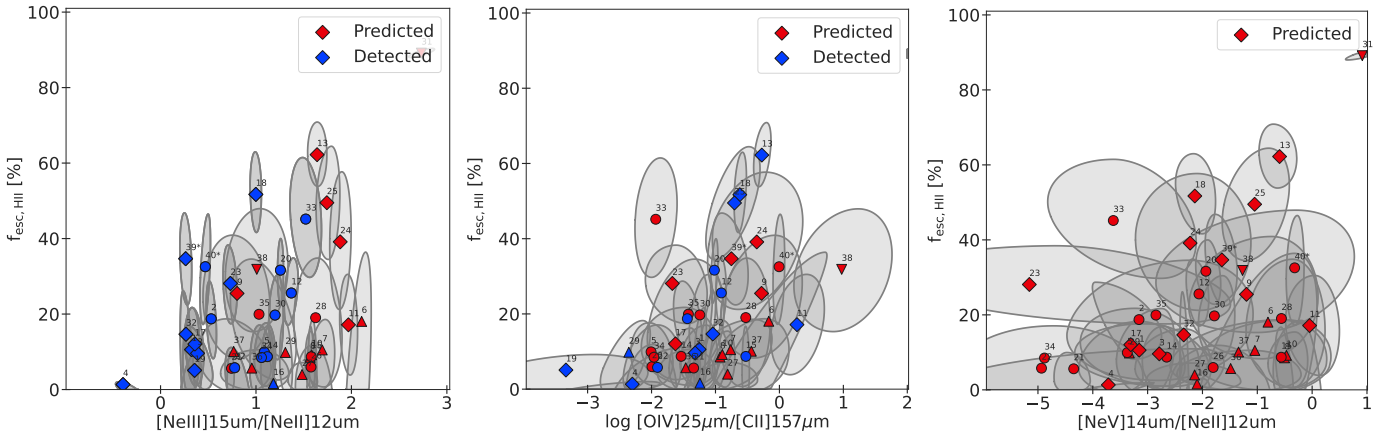

Fig. 15. Individual SUEs showing $f_{esc,HII}$ as a function of line ratio involving ions with high ionization potentials.

element that are not sensitive to abundance profiles are provided in Table A.2 where we report the best potential lines ratio having a Spearman coefficient above 0.3.

In Figs. 14–15, we examine some ratios involving the lines emitted by ions with high ionization potentials: [Ne III]15 µm, [O IV]25 µm and [Ne V]14 µm. The luminosity of these lines strongly depends on the properties of the potential X-ray source and can span a large range of values. Preliminary tests performed with single-star population have shown that our code always needed an additional contribution of an X-ray component to match the observed values of such highly ionized species (as well as neutral gas lines). In the current study, which includes a contribution from binary stars with BPASS, an additional contribution from X-ray is not systematically needed and the X-ray-to-stellar luminosity ratios span several orders of magnitudes (see panel e from Fig. 10). We also note that typical ages found by MULTIGRIS are shifted to later ages when including effects from binary stars.

We find a correlation between $f_{esc,HII}$ and [Ne III]15 µm/[Ne II]12 µm (Ne32), [O IV]25 µm/[C II]158 µm (O4C2) and [Ne V]14 µm/[Ne II]12 µm (Ne52). Nevertheless, we find that O4C2 and Ne52 cover much larger dynamical ranges. In Fig. 15, the large uncertainties of the ellipses reflect the uncertainty on the X-ray component, which is poorly constrained when no [O IV] and [Ne V] detections are provided. Although such ratios involving lines emitted by ions with high ionization potentials are subject to large uncertainties due to the unknown X-ray source properties, they could become of high interest if more constraints on the X-ray spectrum are available. In particular, a more complete statistical sample would provide interesting insights on whether or not leakage is linked to the presence of an X-ray source. Compared to our current predictions, it would also help pinpoint some of the missing physics in our models and might provide insights on the nature of X-ray sources.

6.3.3. Optical line ratios

Although our study focuses on IR tracers, we also examined a few well-known optical and mixed tracers that have been discussed in the literature as possible tracers of LyC escape fractions (see Figs. 16–17). We find that the L_{TIR}/L_{bol} ratio anti-correlates tightly with $f_{esc,HII}$. While L_{bol} accounts for the total luminosity produced by the central stellar cluster (and the additional X-ray contribution), L_{TIR} traces the photons that have been reprocessed by the surrounding gas and dust and re-emitted in the IR. Thus, this ratio traces well the escaping energetic photons that are produced by the stellar cluster but escape before being reprocessed by the gas. We note, however, that this method either requires modeling assumptions for the stellar population or fitting the continuum to derive L_{bol}.

Alternatively, we also put to test the [O III] $\lambda\lambda$4959,5007 Å/[O II] $\lambda\lambda$3726, 3728 Å (O32) and [S III] $\lambda\lambda$9068,9532 Å/[S II] $\lambda\lambda$6716,6731 Å (S32) optical line ratios. The O32 ratio has been claimed to efficiently pick up

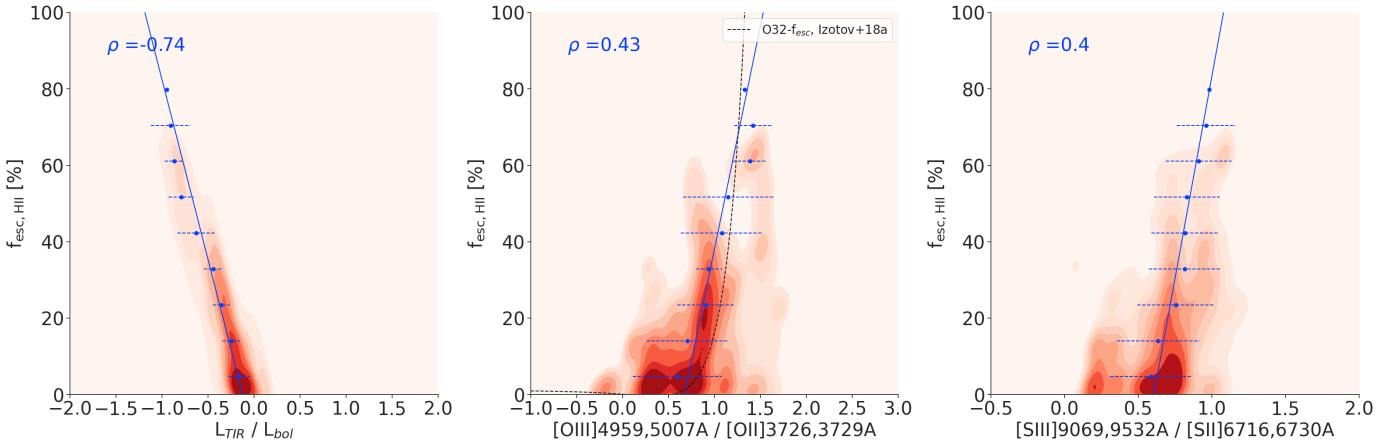

Fig. 16. KDE showing $f_{\rm esc,HII}$ as a function of IR and optical line ratios. The black dashed line on the second panel shows the relation O32-$f_{\rm esc}$(LyC) from Izotov et al. (2018a), assuming that [O III]λ 5007 Å/[O III]λ 4959 Å $\approx$ 3.

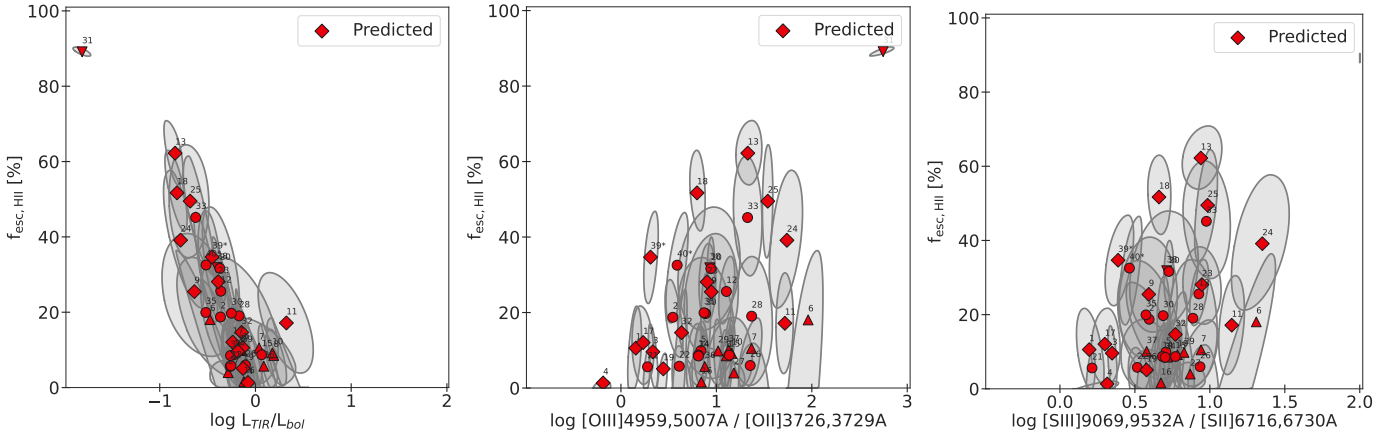

Fig. 17. Individual SUEs showing $f_{\rm esc,HII}$ as a function of IR and optical line ratios.

some of the most extreme leakers (Nakajima & Ouchi 2014; Izotov et al. 2016a,b, 2018a,b; Flury et al. 2022b). In particular, Flury et al. (2022b) find a significant correlation of the escape fraction detected in the UV with O32, although with large scatter. Nevertheless, the O32-$f_{\rm esc}$ relation is not straightforward; Stasińska et al. (2015) have examined galaxies with enhanced O32 ratio and find that a large fraction of them exhibit no sign of leakage. Additionally, recent simulations from Barrow et al. (2020) find that the O32 ratio fluctuates over timescales different than the dynamical time associated with escape fractions. They conclude that although high O32 tend to be associated with high $f_{\rm esc}$, no systematic correlation is expected. Although less classically used than the O32 ratio, the S32 ratio has been shown to be a good tracer of the ionization parameter (Kewley & Dopita 2002; Mingozzi et al. 2020). Because it involves the [S II] $\lambda\lambda$6716,6731 Å lines emission, it should also be sensitive to the presence of density-bounded regions with reduced [S II] $\lambda\lambda$6716,6731 Å emission (Wang et al. 2019, 2021; Ramambason et al. 2020). This ratio has been used, for example, in Zastrow et al. (2011, 2013) as a diagnostic of H II region optical depths. Still, this trend has not yet been confirmed by any observational study since [S III] $\lambda\lambda$9068,9532 Å emission has only been detected in a few LyC-leaking galaxies. Based on our models constrained by IR emission lines, we find that $f_{\rm esc,HII}$ correlates well with the predicted O32 and S32. In particular, the fit we derive for the O32-$f_{\rm esc,HII}$ relation is remarkably close to the relation from Izotov et al. (2018b) derived from fitting 11 LyC-leaking galaxies, and compatible with the new sample of 89 galaxies (detection and nondetection) observed in the Low-z Lyman Continuum Survey (Flury et al. 2022b). We further discuss the reasons that may explain those trends, and why they might be associated with a large scatter in observational studies, in the following section.

7. Discussion

7.1. From gas covering factors to H II region escape fraction

While it has been shown that the escape fraction of ionizing photons is tightly related to the structure of the neutral surrounding gas (e.g., Gazagnes et al. 2018, 2020; Chisholm et al. 2018; Saldana-Lopez et al. 2022), quantitatively linking the gas distribution to the amount of escaping photons remains a complex task. Absorption lines are powerful proxies to probe the properties of the gas on a given line of sight. In particular, H I absorption lines that saturate above a given column density have been used to derive H I covering fractions, which describe the distribution of gas on global galactic scales. Due to the gas kinematics, the covering fractions that are derived are always a lower limit to the real geometric covering fraction (e.g., Gazagnes et al. 2020). Part of this neutral gas component can also be probed in emission where specific tracers emit in dense and irradiated PDRs.

Such tracers are mostly accessible in the IR domain. (e.g., [O I], [C II], [Si II]).

Previous IR studies have focused on quantifying a PDR covering factor (Cormier et al. 2012, 2019; Harikane et al. 2020), which has been used to describe the apparent porosity to ionizing radiation at galactic scales. This PDR covering factor was calculated using models where all the lines arising from the PDR were simultaneously scaled by a common factor between 0 and 1. This grouping is somewhat arbitrary as some lines emit throughout both H II regions and the PDRs. Additionally, these models did not account for possible density-bounded regions since the PDR zone (either total or partial) was always combined with a fully radiation-bounded H II region. Hence, the interpretation of the PDR covering factors is difficult as the simplistic geometry does not account well for complex gas distribution and cannot be linked to potentially escaping photons. In particular, this quantity is very sensitive to modeling assumptions such as the stopping criteria for models (e.g., depth or A_V). Among other factors, accounting for a nonunity filling factor of the ISM or for the physical depth of the photodissociated layers changes the values derived for the PDR covering factor: a very thin PDR with a large covering factor can produce the same emission lines as a thick PDR with a lower one.

In our topological models, we relax the assumption of a full PDR model scaled by a covering factor in favor of an agnostic combination of sectors (as described in Sect. 3) where both the weights and the depths of each sector are free parameters. The Bayesian framework described in Sect. 4.2 allows us to derive PDF and credibility intervals for secondary parameters and to calculate, for each sector, the amount of escaping photons. We then derive a global averaged $f_{\rm esc,HII}$ (see Sect. 3.2.4) corresponding to the population of H II regions present in a galaxy. This integrated quantity directly relates to the ISM gas distribution as opposed to the previously used PDR covering factor.

Nevertheless, as mentioned in Sect. 4.1, our approach bears its own caveats. In particular, the interdependence between total cluster luminosity and escape fractions is not easy to solve. Nevertheless, although scaling the total cluster luminosity affects the absolute lines fluxes, it does not modify the line ratios that are tracing density-bounded regions (see Sect. 6.3). Constraining the cluster luminosity through $L_{\rm TIR}$ is a crucial step that allows us to lift this degeneracy. Still, since we do not use any continuum emission, the UV luminosity we derive for our model depends on our modeling assumptions for the stellar population. The values we derive for the total bolometric luminosities and the escape fractions incorporate uncertainties on the stellar population. Another limit of our work comes from the assumed topology of galaxies; although the multisector combination introduces a layer of complexity that is usually unaccounted for when using single-component photoionization models, this representation remains simplistic when compared to the actual ISM morphology.

7.2. From H II regions to galaxies

The H II region escape fractions that we derive in this study are hardly comparable with the direct UV detection of LyC. Such detections are, first of all, very challenging at low-redshift because the sensitivity of current observatories (e.g., HST/COS) drops at short wavelengths. The few detections available are remarkably small in the local universe and only one galaxy in our sample has a LyC detection (Haro 11, $f_{\rm esc}$ = 3.2%). We note that this value is compatible with the prediction we derive for Haro 11 of $f_{\rm esc,HII} = 9.59^{+9.80}_{-9.51}\%$[6] Similarly, other observations of a few galaxies in the Local Universe detect little to no LyC leakage (e.g., Bergvall et al. 2006; Leitet et al. 2013; Borthakur et al. 2014; Leitherer et al. 2016). Those observations remain very sparse and suffer uncertainties due to line-of-sight variations. They are hard to reconcile with predictions from models (Fujita et al. 2003) and resolved obervations of local starbursts, which suggest that they should be leaking a substantial amount of photons. Quantifying the amount of ionizing photons that are actually escaping at galactic scales remains an elusive question to which we cannot answer within the scope of this study.

7.2.1. Resolution effects

Due to the simplified topology we consider (i.e., a galaxy being a weighted sum of star-forming regions), our predictions for $f_{\rm esc,HII}$ represent an averaged value of the number of photons escaping from the H II regions that populate one galaxy. Although this observable is not easily linked to direct galactic-scale observations, it is representative of the ISM structure within star-forming regions and provides a metric to quantify how many density-bounded regions contribute to the integrated emission. This approach is an attempt to generalize previous analysis performed at local scales with resolved studies in which escape fractions could directly be mapped at the scale of H II regions.

Focusing on resolved H II regions in the local starburst galaxy IC10, Polles et al. (2019) introduced a method to take into account density-bounded regions by considering the cloud depth as a free parameter. While their approach is limited to single-sector models, they find that the observed spatial scale influences the predicted depth, with regions being almost all density-bounded at small scales (~200pc) but shifting to a radiation-bounded regime at large scales. At even smaller scales, hints of substantial escape fractions (above 80%) have been reported in Chevance et al. (2022) where they find very low feedback coupling efficiencies in GMCs from 9 nearby star-forming galaxies, meaning that most of the feedback energy injected by young stars is not dissipated and may escape.

Another technique based on ionization parameter mapping (IPM) using the [O III] $\lambda\lambda$4959,5007 Å/[S II] $\lambda\lambda$6716,6731 Å line ratio was introduced by Pellegrini et al. (2012) to study the optical depth of H II regions in the LMC and SMC. This method has allowed them to classify the population of H II regions of both galaxies and derive rough estimates of the average luminosity weighted $f_{\rm esc,HII}$ (≥40% in both galaxies). These estimates were then corrected using DIG measurements to derive an global average escape fraction of 4–9% in the SMC and 11–17% in the LMC. A similar technique using IPM based on S32 was used on local starbursts (including three galaxies in our sample) by Zastrow et al. (2011, 2013) to identify ionizing cones and gas features that could be associated with enhanced escape fractions. Despite the very good resolution of their data (~60 to 200 pc arcsec^{-1}), the analysis could not provide any quantitative estimate of the escape fractions since individual H II regions were not resolved.

Recent studies have been able to overcome this resolution limit and have provided quantitative estimates of the escape fraction for a few local galaxies. Using ~83 000 resolved stars, Choi et al. (2020) find that the escape fractions of star-forming regions within NGC 4214 exhibit a wide range of

[6] The interval corresponds to the highest density interval at 94%.

values going up to 40%. Their results yield a global escape fraction of about ∼25% for the whole galaxy. Although our study focuses on two different regions of this galaxy, the values we derive ($34.67^{+12.29}_{-13.27}$% for NGC 4214-c and $32.55^{+16.44}_{-15.40}$% for NGC 4214-s) are in good agreement with their predictions. Another approach based on H II region mapping was also made possible by the resolution of MUSE on the ELT. Weilbacher et al. (2018) mapped 386 H II regions from the central field of the Antennae nebula (having approximately a solar metallicity; Bastian et al. 2006; Lardo et al. 2015) and found 38 of them to exhibit large $f_{\rm esc,HII}$ with a mean value of 72%. They estimate the overall escape fraction of the galaxy to be around 7%. Recently, Della Bruna et al. (2021) found an average escape fraction of 67% for the population of H II regions observed with MUSE in NGC 7793 (which has an averaged metallicity close to the solar value; Pilyugin et al. 2014). They find that the radiation leaking from H II regions is sufficient to explain by itself the measured fraction of DIG in this galaxy.

Our results corroborate the picture drawn by these recent studies that find that a substantial amount of ionizing photons might be leaking from H II regions. Connecting these escaping photons to the surrounding diffuse gas is a key element to understand whether or not photons might be leaking out at larger scales.

7.2.2. Missing diffuse component

In our models, a galaxy is represented as a combination of H II regions connected to PDR and molecular zone (see Fig. 3). Additionally, the potential presence of DIG is accounted for by allowing the code to add a sector with very low hydrogen density (as low as 1 $\rm cm^{-3}$). This component is only added by our code if it is needed to reproduce the suite of constraints. In our solutions, however, such diffuse sectors are usually absent. Instead, the code favors configurations with hydrogen density typical of H II regions (around 100 $\rm cm^{-3}$). This is understandable as the IR tracers that we use to constrain our models predominantly trace H II regions and PDRs. Additionally, our treatment of this DIG component is limited by the independent radiative transfers within each Cloudy model considered as a sector. Although the photon transfer is consistently computed throughout each phase, the different sectors are independent from one another and do not overlap, meaning that photons escaping from a given sector cannot be reabsorbed by a second component along the line of sight.

This unaccounted for DIG could be problematic to reproduce the optical emission lines (e.g., Hα in Sect. 5 and optical line ratios in Sect. 6.3) we derive in our analysis. Indeed, contamination by DIG has been extensively studied in the optical domain and is known to significantly contribute to some line ratios such as [S II]/Hα, [N II]/Hα and [O III]/Hβ (Oey et al. 2007; Zhang et al. 2017; Kreckel et al. 2016; Vale Asari et al. 2019, 2020; Espinosa-Ponce et al. 2020; Della Bruna et al. 2021; Belfiore et al. 2022). Nevertheless, we have seen in Sect. 5.2 that our models predict Hα fluxes in agreement with observations. This agreement is expected since our code reproduces both the metallicity (see Fig. 5) and the recombination line Huα in the galaxies in which it is detected. Nevertheless, it appears at odds with the large $f_{\rm esc,HII}$ we derive and could be explained by different scenarios. A first possibility is that the measured Hα emission in the DGS galaxies is dominated by H II regions with a negligible contribution from DIG. This would be in line with observational studies that find that, as opposed to nonstarbursting galaxies, the Hα emission in starbursts is largely dominated by H II regions in which the emission is boosted (Hanish et al. 2010). In that case, both our models and Hα observations would be fairly independent of DIG. This would suggest that the set of observables used in this study is not sensitive to DIG and that additional tracers are needed to probe this diffuse component.

Alternatively, if part of the detected Hα emission is powered by DIG, the fact that our models reproduce it without including a diffuse component means that our code somehow compensates for the missing DIG contribution, although assuming a simplified DIG-free geometry. As opposed to the previous scenario, this would only be possible if some of the IR tracers we use are sensitive to the DIG contribution. Although, to our knowledge, no systematic studies are available regarding DIG contamination in the IR domain, we do have some observational evidence that it may marginally contribute to the emission of some of the tracers considered in our study (Cormier et al. 2012, 2019). All in all, we find that our models based on combination of H II regions are sufficient to reproduce the observed Hα emission, although they may be either missing or compensating for the DIG contribution. To disentangle those scenarios, further progress is needed to either correct the emission lines for DIG contamination before using them as constraints of MULTIGRIS or properly including in the models an additional DIG component. To that purpose, high spatial-resolution and sensitive telescopes in the far-IR would be useful to study the properties of the DIG.

Most importantly, we also do not account for the actual distribution of diffuse neutral gas. This is a major caveat considering that escaping photons has been shown to be very sensitive to the distribution of neutral hydrogen (e.g., Gazagnes et al. 2020). Although our models make the most of the few IR constraints tracing the PDRs, we are only accounting for the dense, highly irradiated neutral gas, which is seen in emission. This might partly explain why our estimates of the total H I masses (M(H I)), shown in Fig. A.2, tend to be significantly smaller that the measurements from Rémy-Ruyer et al. (2014) even though we find that those values agree within 0.5dex for 16 out of 40 galaxies. However, comparing the H I column densities of our sectors, we see that even the deepest sector of each model has a column density much lower than that measured in absorption (see Table A.3). This is one of the main limitations of our models: even when our models successfully recover the total H I gas mass, the distribution of sectors remains too simplistic. The actual distribution of the ISM likely resembles more the randomly distributed version presented in Fig. 3. In particular, introducing holes or very low column-density sight-lines in our topological models would naturally push the code to predict deeper sectors with higher column densities. Because of the aforementioned limitations regarding the diffuse ionized and neutral gas, we cannot provide estimates for the global galactic escape fraction in our sample. Instead, we focus on analysing the trends with various parameters and their physical interpretation.

7.3. Dependences of $f_{\rm esc,HII}$

7.3.1. Correlation with galactic properties and observational limitations

We find a clear anti-correlation of $f_{\rm esc,HII}$ with metallicity and a positive correlation with sSFR. Such trends are reminiscent of those observed in the local sample of few direct measurements from Leitet et al. (2013). They are, however, not clearly seen in larger surveys using UV detection of the LyC (e.g., Saxena et al. 2022; Pahl et al. 2022), although the recently

observed Low-redshift Lyman Continuum Survey (89 galaxies with LyC measurements) has revealed a similar trend of f_{esc} with sSFR (Flury et al. 2022b).

The lack of observed correlation between galactic properties and measured LyC escape fractions might be due to a dilution effect of the localized H II regions from where photons escape into the galactic ISM (Saxena et al. 2022). This dilution may explain why we find clear trends of $f_{esc,HII}$ with several galactic parameters (see Sect. 6.2) while such correlations are unseen at galactic scales. Another crucial parameter that might strongly limit the observation of empirical correlation is the unknown inclination of unresolved galaxies. For example, Bassett et al. (2019) emphasize that different viewing angles can affect both the measured LyC values and the observed optical lines ratios (such as O32). Under the hypothesis of an anisotropic LyC leakage, Nakajima et al. (2020) suggest that the viewing angle of porous H II regions might be the main parameter controlling the detection or nondetection of LyC leakage in a sample of Lyα emitting galaxies from the LymAn Continuum Escape Survey (LACES). Although the dependence of emission line ratios on the viewing angle is somewhat mitigated by the use of optically thin tracers (e.g., IR lines), the exact impact of inclination on the observed emission from galaxies remains to be quantitatively examined. This effect was not examined in the current study since we cannot extract any information on the inclination in our sample of galaxies.

Further progress on this issue requires the use of complex 3D radiative transfer codes as this effect is not accounted for when using 1D radiative transfer codes such as Cloudy. For example, Mauerhofer et al. (2021) performed a detailed investigation of the viewing angle impact on simulated galaxies and find that the measured projected escape fractions are strongly impacted by orientation (with variations from 0 to 47%). This lack of correlation might also be due to propagation effects through the DIG, which are not taken into account in our H II region modeling. Those effects remain largely unconstrained since they depend on the H I gas 3D-distribution, which is unknown.

7.3.2. Potential impact from nonstellar feedback

Although the high $f_{esc,HII}$ in our sample are predominantly associated with low-metallicity, low mass, and high sSFR, a few objects in our sample exhibit surprisingly high $f_{esc,HII}$ with somewhat larger metallicities and moderate sSFR. We now discuss additional mechanisms that could explain the position of such objects on the $f_{esc,HII}$-metallicity relation. In particular, we examine if other feedback mechanisms such as the potential presence of an ULX or outflows could be responsible for the position of outliers in our sample. In Fig. 18 (first panel), we mark the galaxies having a high measured sSFR (log sSFR > -9.5). This criterion is common to all the galaxies with inferred $f_{esc,HII} > 40\%$ and shared by 10 out of 14 galaxies with $f_{esc,HII} > 20\%$. Nevertheless, 4 galaxies with $f_{esc,HII} > 20\%$ exhibit only a moderate sSFR. We now examine other potential mechanisms that may be responsible for their enhanced $f_{esc,HII}$.

On the middle panel from Fig. 18, we examine the impact of the X-ray-to-stellar luminosity ratio. Although it has not been shown that the presence of an X-ray source have a direct impact on escape fractions, high L_X tends to be associated with of intense star-formation rates (Grimm et al. 2003) and the emission of high-energy photons, which travel deep into the ISM may enhance the global escape fraction of ionizing photons. We find X-ray-to-stellar luminosity ratio above 0.5% in 8 out of 40 sources. Two out of the five reported potential ULX in our sample are associated with $L_X/L_* > 0.5\%$. We note that the two galaxies (13: I Zw 18 and 25: SBS 0335-052) with the highest $f_{esc,HII}$ in our sample (>50%) are associated with a known ULX. However, similar levels of X-ray contribution are also found in galaxies for which we infer little to no leakage. All in all, we find that large X-ray-to-stellar luminosities are not necessarily associated with enhanced leakage, although X-ray sources may contribute in some specific cases. Additionally, we note that a recent study from Marques-Chaves et al. (2022) finds no correlation between the LyC escape fractions measured in the Low-z Lyman Continuum Survey and the spectral hardness of the ionizing sources.

Finally, we expect dynamical effects disturbing the gas kinematics to also impact the H II region escape fractions. Outflows may remove gas from the galactic disk and create low-density channel from where photons can freely travel (Fujita et al. 2003). Recent studies have also examined how gas fragmentation induced by turbulence might play a role in producing the patchy structure of the ISM through which photons escape (e.g., Kakiichi & Gronke 2021; Hogarth et al. 2020). Although outflow signatures are only detected in a few objects (11/40), they offer a plausible explanation to interpret the position of some outliers having large predicted $f_{esc,HII}$ but modest sSFR and little to no contribution from an additional X-ray source. Detailed studies of the gas kinematics are available for a few objects in our sample (see Sect. 2.1 for references) based on either resolved observations or analysis of emission line profiles. Since photoionization and photodissociation models do not account for dynamical effects, we only provide a qualitative discussion on the potential role of gas outflows. In Fig. 18 (third panel), we show that the presence of outflows may explain the position of several galaxies with relatively high metallicities (12+log(O/H) > 8) but $f_{esc,HII} > 20\%$ (e.g., 2: Haro 3, 23: NGC 625, and 39–40: NGC 4214).

In summary, we find that although the presence of nonstellar feedback mechanisms such as a strong contribution of an X-ray source or outflowing gas may contribute to power photon leakage, such mechanisms seem to only have a secondary influence on the inferred $f_{esc,HII}$. Instead, we find that variations of $f_{esc,HII}$ appear to be mainly controlled by global galactic properties such as the metallicity, stellar mass and sSFR. However, once photons have escaped from their H II region, the additional feedback mechanisms that we have discussed in this section might play a role to convert $f_{esc,HII}$ into global galactic escape fractions by removing or ionizing the surrounding neutral gas.

7.4. Prospects for future observations

We provide in this study an observational framework to pinpoint galaxies with a substantial amount of $f_{esc,HII}$. The presence of leaking H II regions providing a substantial amount of ionizing photons is a necessary condition for a galaxy to be a potential LyC-leaker at larger scales. Hence, the tracers that we propose here could serve as a first basis to select galaxy samples with potential leakage.

We identify line ratios that could be used as tracers of the escape fraction. First, we find that among the classically used optical line ratio, O32 and S32 lines ratio correlate well with our predicted $f_{esc,HII}$. This finding is at odds with observational studies finding little correlations with such diagnostics, and large scatters. In particular, the clear O32-$f_{esc,HII}$ correlation is unexpected in the view of several recent studies that reported

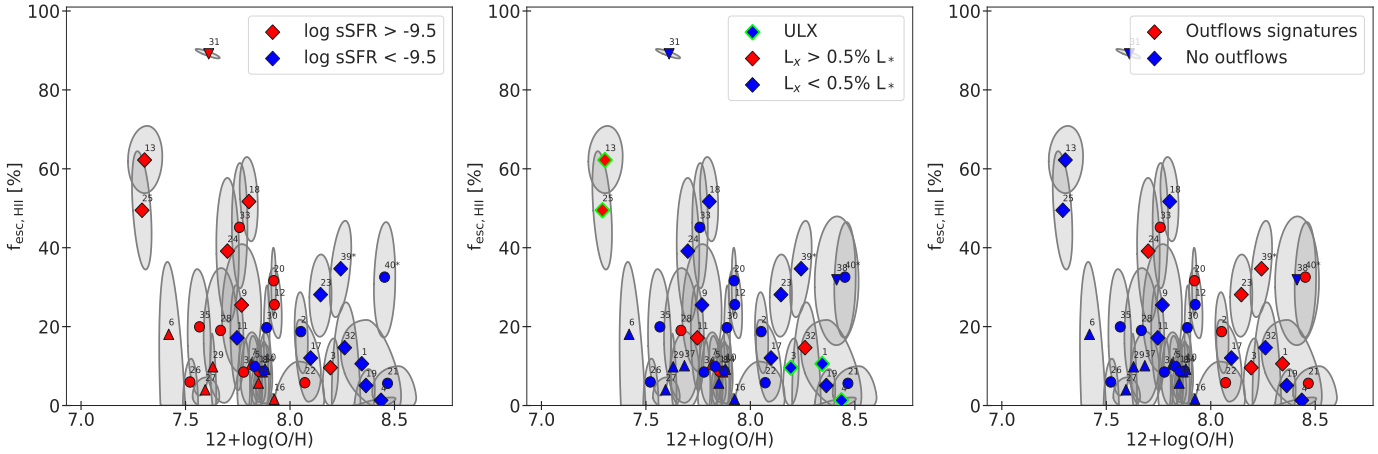

Fig. 18. Impact of sSFR, X-ray sources and outflows on the escape fraction. *Left*: impact of the measured sSFR from Rémy-Ruyer et al. (2015) on the $f_{esc,HII}$ -metallicity relation. No sSFR measurement are available for galaxies 37: HS 0017+1055 and 38: HS 2352+2733. *Middle*: impact of the X-ray-to-stellar luminosity ratio. The green annuli mark galaxies in which a potential ULX ($L_X > 10^{39}$ erg s^{-1}) have been reported. 1: Haro 2 (Otí-Floranes et al. 2012), 3: Haro 11 (Prestwich et al. 2015; Gross et al. 2021), 4: He 2–10 (Ott et al. 2005; Reines et al. 2011), 13: I Zw 18(Thuan et al. 2004; Ott et al. 2005; Kaaret et al. 2011; Kaaret & Feng 2013; Brorby et al. 2014) and 25:SBS 0335-052 (Thuan et al. 2004; Prestwich et al. 2013). *Right*: impact of outflowing gas signatures. 1: Haro 2 Otí-Floranes et al. (2012), 3: Haro 11 (Menacho et al. 2019), 20: NGC 1569 (Sánchez-Cruces et al. 2015), 21: NGC 1705 (Zastrow et al. 2013), 22: NGC 5253 Zastrow et al. (2011), 23: NGC 625 (Cannon et al. 2004; McQuinn et al. 2018), 33: UM 461 (Carvalho & Plana 2018), 39-40: NGC 4214 (McQuinn et al. 2018).

numerous caveats and limits in the use the O32 ratio for studying f_{esc} (e.g., Stasińska et al. 2015; Plat et al. 2019; Bassett et al. 2019; Flury et al. 2022a,b). One possible explanation is that the $f_{esc,HII}$ we derive is less sensitive to viewing angle dependences than the measured escape fractions. Alternatively, this result might also be due to our sample selection, which focuses on highly star-forming galaxies. Additionally, both O32 and S32 ratios have been shown to correlate with the ionization parameter (Nakajima & Ouchi 2014; Kewley & Dopita 2002). Disentangling the effect of metallicity, ionization parameter, and escape fractions remains a difficult task, even within the Bayesian framework that we developed here.

Among the IR tracer, the well-know O3C2 ratio has been used in the analysis of high-z galaxies with ALMA (Harikane et al. 2020). Nevertheless, the interpretation of high O3C2 ratio is still uncertain and debated. Recently, Katz et al. (2022b) have shown that this ratio is sensitive to a wide variety of other galactic parameters including the C/O ratio, IMF, ionization parameter, gas density and CMB attenuation. In our study, assuming that the C/O in the local universe is observationally well constrained, we do find a trend of increasing H II region escape fraction associated with higher O3C2 ratios. However, we find that ratios involving the [N II] lines (122 and 205 µm) provide better diagnostics to trace $f_{esc,HII}$ since [N II] is less contaminated than [C II] by PDR emission. Although fainter, these lines also fall in the detection range of ALMA and have already been detected in a few high-redshift galaxies above $z \sim 4$. (e.g., De Breuck et al. 2019). Even for nondetections, upper limits on [N II] could also be useful to set lower limits on the resulting escape fractions.

For tracers such as [O IV] and [Ne V], associated with ions of high ionization potentials, our interpretation is limited by the little number of constraints that we have on the X-ray component. The JWST will open a new window for the observation of galaxies with signatures of high energy sources and their associated spectral signatures (e.g., [Mg IV]λ4.49 µm, [Ne VI]λ7.64 µm, and [Ne V]λ14.32 µm; Satyapal et al. 2021; Richardson et al. 2022). The combination of such lines with the emission of neutral gas lines (e.g., [Fe II]λ1.64 µm) boosted by X-ray photons may help disentangle the feedback from potential compact objects from the one induced by turbulence and shocks (Cresci et al. 2010, 2017). Those IR tracers could be used in combination with other tracers of high energy photons available in the optical range (e.g., [Ne V], [Fe V]) and in the UV range (e.g., C IV λ5808, He II λ1640, Wofford et al. 2021; Berg et al. 2019, 2021). A first example of the combination of optical and IR data is presented in LR22. Looking for counter-parts in the radio and X-ray domain with sensitive instruments (e.g., with the forthcoming *Athena* telescope), may also allow progress on the identification of the exact nature of the different sources producing high energy photons.

Finally, the numerous caveats mentioned in this section regarding the distribution of neutral gas might be lifted by incorporating into our models tracers of the neutral gas (e.g., the Lyα line, the hydrogen 21 cm line) or by using priors calibrated on absorption lines studies. Such observations, both in the local and high-redshift universe, will provide a more complex picture of the interplay between low-metallicity ISM and the intense sources of radiation expected in the early universe. In a subsequent publication we will explore and model the optical emission lines of galaxies from the Low-redshift Lyman Continuum Survey with MULTIGRIS (Marques-Chaves et al., in prep.).

8. Conclusion

This paper presents the first application of a new Bayesian code, MULTIGRIS (LR22), which opens new possibilities in terms of multiphase modeling of the ISM. We developed a framework to automatically combine photoionization and photodissociation models as different sectors representative of different physical and chemical ISM conditions, and including the presence of density-bounded regions from where ionizing photons can escape. We applied this new code to a sample of low-metallicity dwarf galaxies from the DGS (Madden et al. 2013) with extensive sets of IR emission line measurements and whose ISM properties may resemble those of galaxies in the early Universe. We examined the impact of using multisector models on

the predictions obtained for well-constrained physical quantities such as the metallicity and SFR. We then derived predictions for the H II region escape fraction of ionizing photons in the DGS and studied how this observable varies with various galactic parameters. Our main results are summarized below:

- Accounting for several, possibly density-bounded, sectors does not significantly affect the metallicity and SFR we derive. This confirms that empirical calibrations based on emission lines provide robust results that hold at galactic scales, even when accounting for a slightly more complex topology. We find, however, that not accounting for escaping photons may result in a moderate underestimation of the SFR. The correction factor we derive by estimating SFR based on $Q(H^0)$ tightly follows the analytical curve $1/(1-f_{esc,HII})$.
- We confirm previous results from Cormier et al. (2019) that show that the ISM structure and porosity strongly varies with metallicity. Specifically, we find that $f_{esc,HII}$ tends to increase when metallicity decreases. In the global KDE derived for the whole DGS sample, we find that the high-metallicity and high-$f_{esc,HII}$ region of the parameter space is completely unpopulated, meaning that high-metallicity galaxies have a low probability of reaching high values of $f_{esc,HII}$. Despite a somewhat large scatter around this relation, especially at low-metallicity, this trend is robust to changes in our modeling assumptions on the number of sectors and holds when forcing a fixed number of sectors. Our findings back up the picture of an increasingly porous ISM at low-metallicity, in which energetic photons can easily leak out of H II regions and ionize their surrounding gas.
- We also find that the fraction of [C II] emitted in the ionized phase drops by ∼2 orders of magnitude when going from a solar metallicity down to a metallicity of $\sim 1/35 Z_\odot$. We explain this effect by the presence of enhanced radiation fields in low-metallicity environments, which favor the ionization of C^{2+} ions with respect to C^+. Hence, we find that [C II] emission arises predominantly from the neutral atomic and molecular gas in low-metallicity environments. A more detailed study of the neutral and molecular gas mass distribution in the DGS, including the CO-dark component, will be carried out in a future work (Ramambason et al., in prep.).
- We examine secondary dependences of $f_{esc,HII}$ with galactic properties and find a weak anti-correlation with stellar mass and a clear correlation with sSFR. These findings fit in the picture of the escape fraction being a feedback-dominated process with complex dependences on the gas mass distribution and gas chemistry.
- Finally, we provide several line ratios that correlate with $f_{esc,HII}$, both in the IR domain (e.g., O3N2, O3C2), including some tracers with high ionization potential (e.g., Ne32, O4C2, and Ne52) and in the optical (e.g., O32, S32), as well as mixed tracers such as L_{TIR}/L_{bol}.

The modeling framework presented in this paper is highly sensitive to the wealth of tracers that are used as inputs, especially when they trace gas in different physical and chemical conditions. With that in mind, potentially more complex and realistic models could be constrained by making use of multiwavelength data combining emission from X-ray to IR.

Acknowledgements. The authors thank the anonymous referee for constructive feedback and useful comments. V. L., C. R. and L. R. thank the FACE Foundation Thomas Jefferson Fund (grant number TJF21_053). C. M. thanks the grant UNAM/PAPIIT – IN101220. M. C. gratefully acknowledges funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through an Emmy Noether Research Group (grant numbers KR4801/1-1 and CH2137/1-1), as well as from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program via the ERC Starting Grant MUSTANG (grant agreement number 714907).

References

Allen, M. G., Groves, B. A., Dopita, M. A., Sutherland, R. S., & Kewley, L. J. 2008, ApJS, 178, 20
Amorín, R. O., Pérez-Montero, E., & Vílchez, J. M. 2010, ApJ, 715, L128
Asplund, M., Grevesse, N., Sauval, A. J., & Scott, P. 2009, ARA&A, 47, 481
Bakx, T. J. L. C., Tamura, Y., Hashimoto, T., et al. 2020, MNRAS, 493, 4294
Barrow, K. S. S., Robertson, B. E., Ellis, R. S., et al. 2020, ApJ, 902, L39
Bassett, R., Ryan-Weber, E. V., Cooke, J., et al. 2019, MNRAS, 483, 5223
Bastian, N., Emsellem, E., Kissler-Patig, M., & Maraston, C. 2006, A&A, 445, 471
Belfiore, F., Santoro, F., Groves, B., et al. 2022, A&A, 659, A26
Berg, D. A., Erb, D. K., Henry, R. B. C., Skillman, E. D., & McQuinn, K. B. W. 2019, ApJ, 874, 93
Berg, D. A., Chisholm, J., Erb, D. K., et al. 2021, ApJ, 922, 170
Bergvall, N., Zackrisson, E., Andersson, B. G., et al. 2006, A&A, 448, 513
Bian, F., Fan, X., McGreer, I., Cai, Z., & Jiang, L. 2017, ApJ, 837, L12
Bik, A., Östlin, G., Menacho, V., et al. 2018, A&A, 619, A131
Binder, B., Williams, B. F., Eracleous, M., et al. 2015, AJ, 150, 94
Binette, L., Wilson, A. S., & Storchi-Bergmann, T. 1996, A&A, 312, 365
Borthakur, S., Heckman, T. M., Leitherer, C., & Overzier, R. A. 2014, Science, 346, 216
Bradley, T. R., Knapen, J. H., Beckman, J. E., & Folkes, S. L. 2006, A&A, 459, L13
Brorby, M., Kaaret, P., & Prestwich, A. 2014, MNRAS, 441, 2346
Calzetti, D., Liu, G., & Koda, J. 2012, ApJ, 752, 98
Cannon, J. M., McClure-Griffiths, N. M., Skillman, E. D., & Côté, S. 2004, ApJ, 607, 274
Carniani, S., Maiolino, R., Pallottini, A., et al. 2017, A&A, 605, A42
Carvalho, M. S., & Plana, H. 2018, MNRAS, 481, 122
Chastenet, J., Sandstrom, K., Chiang, I.-D., et al. 2019, ApJ, 876, 62
Chevance, M., Madden, S. C., Lebouteiller, V., et al. 2016, A&A, 590, A36
Chevance, M., Kruijssen, J. M. D., Vazquez-Semadeni, E., et al. 2020a, Space Sci. Rev., 216, 50
Chevance, M., Madden, S. C., Fischer, C., et al. 2020b, MNRAS, 494, 5279
Chevance, M., Kruijssen, J. M. D., Krumholz, M. R., et al. 2022, MNRAS, 509, 272
Chisholm, J., Gazagnes, S., Schaerer, D., et al. 2018, A&A, 616, A30
Chisholm, J., Rigby, J. R., Bayliss, M., et al. 2019, ApJ, 882, 182
Chisholm, J., Prochaska, J. X., Schaerer, D., Gazagnes, S., & Henry, A. 2020, MNRAS, 498, 2554
Choi, Y., Dalcanton, J. J., Williams, B. F., et al. 2020, ApJ, 902, 54
Cormier, D., Lebouteiller, V., Madden, S. C., et al. 2012, A&A, 548, A20
Cormier, D., Madden, S. C., Lebouteiller, V., et al. 2015, A&A, 578, A53
Cormier, D., Abel, N. P., Hony, S., et al. 2019, A&A, 626, A23
Cresci, G., Vanzi, L., Sauvage, M., Santangelo, G., & van der Werf, P. 2010, A&A, 520, A82
Cresci, G., Vanzi, L., Telles, E., et al. 2017, A&A, 604, A101
Crowther, P. A., Schnurr, O., Hirschi, R., et al. 2010, MNRAS, 408, 731
Croxall, K. V., Smith, J. D., Pellegrini, E., et al. 2017, ApJ, 845, 96
de Barros, S., Vanzella, E., Amorín, R., et al. 2016, A&A, 585, A51
De Breuck, C., Weiß, A., Béthermin, M., et al. 2019, A&A, 631, A167
De Looze, I., Cormier, D., Lebouteiller, V., et al. 2014, A&A, 568, A62
Della Bruna, L., Adamo, A., Bik, A., et al. 2020, A&A, 635, A134
Della Bruna, L., Adamo, A., Lee, J. C., et al. 2021, A&A, 650, A103
Duarte Puertas, S., Vilchez, J. M., Iglesias-Páramo, J., et al. 2022, A&A, 666, A186
Eggen, N. R., Scarlata, C., Skillman, E., & Jaskot, A. 2021, ApJ, 912, 12
Eldridge, J. J., Stanway, E. R., Xiao, L., et al. 2017, PASA, 34, e058
Espinosa-Ponce, C., Sánchez, S. F., Morisset, C., et al. 2020, MNRAS, 494, 1622
Falkendal, T., Lehnert, M. D., Vernet, J., De Breuck, C., & Wang, W. 2021, A&A, 645, A120
Ferland, G. J., Chatzikos, M., Guzmán, F., et al. 2017, Rev. Mex. Astron. Astrofis., 53, 385
Finkelstein, S. L., D'Aloisio, A., Paardekooper, J.-P., et al. 2019, ApJ, 879, 36
Fletcher, T. J., Tang, M., Robertson, B. E., et al. 2019, ApJ, 878, 87
Flury, S. R., Jaskot, A. E., Ferguson, H. C., et al. 2022a, ApJS, 260, 1
Flury, S. R., Jaskot, A. E., Ferguson, H. C., et al. 2022b, ApJ, 930, 126
Fujita, A., Martin, C. L., Mac Low, M.-M., & Abel, T. 2003, ApJ, 599, 50
Galliano, F. 2018, MNRAS, 476, 1445
Galliano, F., Nersesian, A., Bianchi, S., et al. 2021, A&A, 649, A18
Gazagnes, S., Chisholm, J., Schaerer, D., et al. 2018, A&A, 616, A29

Gazagnes, S., Chisholm, J., Schaerer, D., Verhamme, A., & Izotov, Y. 2020, A&A, 639, A85
Grimm, H. J., Gilfanov, M., & Sunyaev, R. 2003, MNRAS, 339, 793
Gross, A. C., Prestwich, A., & Kaaret, P. 2021, MNRAS, 505, 610
Hanish, D. J., Oey, M. S., Rigby, J. R., de Mello, D. F., & Lee, J. C. 2010, ApJ, 725, 2029
Harikane, Y., Ouchi, M., Inoue, A. K., et al. 2020, ApJ, 896, 93
Hashimoto, T., Inoue, A. K., Mawatari, K., et al. 2019, PASJ, 71, 71
Henry, A., Scarlata, C., Martin, C. L., & Erb, D. 2015, ApJ, 809, 19
Henry, A., Berg, D. A., Scarlata, C., Verhamme, A., & Erb, D. 2018, ApJ, 855, 96
Herenz, E. C., Hayes, M., Papaderos, P., et al. 2017, A&A, 606, L11
Hogarth, L., Amorín, R., Vílchez, J. M., et al. 2020, MNRAS, 494, 3541
Hosokawa, T., & Inutsuka, S.-I. 2005, ApJ, 623, 917
Indriolo, N., Geballe, T. R., Oka, T., & McCall, B. J. 2007, ApJ, 671, 1736
Inoue, A. K., Hirashita, H., & Kamaya, H. 2001, ApJ, 555, 613
Inoue, A. K., Tamura, Y., Matsuo, H., et al. 2016, Science, 352, 1559
Izotov, Y. I., & Thuan, T. X. 1999, ApJ, 511, 639
Izotov, Y. I., Thuan, T. X., & Lipovetsky, V. A. 1994, ApJ, 435, 647
Izotov, Y. I., Chaffee, F. H., & Green, R. F. 2001, ApJ, 562, 727
Izotov, Y. I., Noeske, K. G., Guseva, N. G., et al. 2004a, A&A, 415, L27
Izotov, Y. I., Papaderos, P., Guseva, N. G., Fricke, K. J., & Thuan, T. X. 2004b, A&A, 421, 539
Izotov, Y. I., Stasińska, G., Meynet, G., Guseva, N. G., & Thuan, T. X. 2006, A&A, 448, 955
Izotov, Y. I., Thuan, T. X., & Privon, G. 2012, MNRAS, 427, 1229
Izotov, Y. I., Orlitová, I., Schaerer, D., et al. 2016a, Nature, 529, 178
Izotov, Y. I., Schaerer, D., Thuan, T. X., et al. 2016b, MNRAS, 461, 3683
Izotov, Y. I., Thuan, T. X., & Guseva, N. G. 2017, MNRAS, 471, 548
Izotov, Y. I., Schaerer, D., Worseck, G., et al. 2018a, MNRAS, 474, 4514
Izotov, Y. I., Worseck, G., Schaerer, D., et al. 2018b, MNRAS, 478, 4851
Izotov, Y. I., Schaerer, D., Worseck, G., et al. 2020, MNRAS, 491, 468
Izotov, Y. I., Worseck, G., Schaerer, D., et al. 2021, MNRAS, 503, 1734
James, B. L., Aloisi, A., Heckman, T., Sohn, S. T., & Wolfe, M. A. 2014, ApJ, 795, 109
Jaskot, A. E., & Oey, M. S. 2013, ApJ, 766, 91
Jiang, T., Malhotra, S., Rhoads, J. E., & Yang, H. 2019, ApJ, 872, 145
Jones, A. P., Köhler, M., Ysard, N., Bocchio, M., & Verstraete, L. 2017, A&A, 602, A46
Kaaret, P., & Feng, H. 2013, AAS/High Energy Astrophys. Div., 13, 402.06
Kaaret, P., Schmitt, J., & Gorski, M. 2011, ApJ, 741, 10
Kakiichi, K., & Gronke, M. 2021, ApJ, 908, 30
Katz, H., Ďurovčíková, D., Kimm, T., et al. 2020, MNRAS, 498, 164
Katz, H., Garel, T., Rosdahl, J., et al. 2022a, MNRAS, 515, 4265
Katz, H., Rosdahl, J., Kimm, T., et al. 2022b, MNRAS, 510, 5603
Kehrig, C., Vílchez, J. M., Guerrero, M. A., et al. 2018, MNRAS, 480, 1081
Kennicutt, R. C., Jr., Bresolin, F., Bomans, D. J., Bothun, G. D., & Thompson, I. B. 1995, AJ, 109, 594
Kennicutt, R. C., Jr, Hao, C.-N., Calzetti, D., et al. 2009, ApJ, 703, 1672
Kewley, L. J., & Dopita, M. A. 2002, ApJS, 142, 35
Kim, J., Chevance, M., Kruijssen, J. M. D., et al. 2022, MNRAS, 516, 3006
Kim, J.-G., Kim, W.-T., & Ostriker, E. C. 2018, ApJ, 859, 68
Kim, J.-G., Kim, W.-T., & Ostriker, E. C. 2019, ApJ, 883, 102
Kim, J.-G., Ostriker, E. C., & Filippova, N. 2021, ApJ, 911, 128
Kimm, T., Katz, H., Haehnelt, M., et al. 2017, MNRAS, 466, 4826
Kimm, T., Blaizot, J., Garel, T., et al. 2019, MNRAS, 486, 2215
Kreckel, K., Blanc, G. A., Schinnerer, E., et al. 2016, ApJ, 827, 103
Kunth, D., Mas-Hesse, J. M., Terlevich, E., et al. 1998, A&A, 334, 11
Lacerda, E. A. D., Cid Fernandes, R., Couto, G. S., et al. 2018, MNRAS, 474, 3727
Lambert-Huyghe, A., Madden, S. C., Lebouteiller, V., et al. 2022, A&A, 666, A112
Lardo, C., Davies, B., Kudritzki, R. P., et al. 2015, ApJ, 812, 160
Le Petit, F., Nehmé, C., Le Bourlot, J., & Roueff, E. 2006, ApJS, 164, 506
Lebouteiller, V., & Ramambason, L. 2022, A&A, 667, A34
Lebouteiller, V., Péquignot, D., Cormier, D., et al. 2017, A&A, 602, A45
Lebouteiller, V., Cormier, D., Madden, S. C., et al. 2019, A&A, 632, A106
Lee, M. Y., Madden, S. C., Lebouteiller, V., et al. 2016, A&A, 596, A85
Lee, M. Y., Madden, S. C., Le Petit, F., et al. 2019, A&A, 628, A113
Leitherer, C., Schaerer, D., Goldader, J. D., et al. 1999, ApJS, 123, 3
Leitet, E., Bergvall, N., Hayes, M., Linné, S., & Zackrisson, E. 2013, A&A, 553, A106
Leitherer, C., Hernandez, S., Lee, J. C., & Oey, M. S. 2016, ApJ, 823, 64
Ma, X., Hopkins, P. F., Kasen, D., et al. 2016, MNRAS, 459, 3614
Ma, X., Quataert, E., Wetzel, A., et al. 2020, MNRAS, 498, 2001
Madden, S. C., Rémy-Ruyer, A., Galametz, M., et al. 2013, PASP, 125, 600
Madden, S. C., Cormier, D., Hony, S., et al. 2020, A&A, 643, A141
Maiolino, R., & Mannucci, F. 2019, A&ARv, 27, 3
Maji, M., Verhamme, A., Rosdahl, J., et al. 2022, A&A, 663, A66
Marques-Chaves, R., Schaerer, D., Amorín, R. O., et al. 2022, A&A, 663, L1
Martin, C. L. 1998, ApJ, 506, 222
Mathis, J. S. 1986, ApJ, 301, 423
Mauerhofer, V., Verhamme, A., Blaizot, J., et al. 2021, A&A, 646, A80
McQuinn, K. B. W., Skillman, E. D., Heilman, T. N., Mitchell, N. P., & Kelley, T. 2018, MNRAS, 477, 3164
Menacho, V., Östlin, G., Bik, A., et al. 2019, MNRAS, 487, 3183
Menacho, V., Östlin, G., Bik, A., et al. 2021, MNRAS, 506, 1777
Meyer, R. A., Walter, F., Cicone, C., et al. 2022, ApJ, 927, 152
Mingozzi, M., Belfiore, F., Cresci, G., et al. 2020, A&A, 636, A42
Mirabel, I. F., Dijkstra, M., Laurent, P., Loeb, A., & Pritchard, J. R. 2011, A&A, 528, A149
Mitsuda, K., Inoue, H., Koyama, K., et al. 1984, PASJ, 36, 741
Naidu, R. P., Oesch, P. A., Reddy, N., et al. 2017, ApJ, 847, 12
Naidu, R. P., Forrest, B., Oesch, P. A., Tran, K.-V. H., & Holden, B. P. 2018, MNRAS, 478, 791
Naidu, R. P., Tacchella, S., Mason, C. A., et al. 2020, ApJ, 892, 109
Nakajima, K., & Ouchi, M. 2014, MNRAS, 442, 900
Nakajima, K., Ellis, R. S., Robertson, B. E., Tang, M., & Stark, D. P. 2020, ApJ, 889, 161
Nicholls, D. C., Sutherland, R. S., Dopita, M. A., Kewley, L. J., & Groves, B. A. 2017, MNRAS, 466, 4403
Oberst, T. E., Parshley, S. C., Nikola, T., et al. 2011, ApJ, 739, 100
Oey, M. S., Meurer, G. R., Yelda, S., et al. 2007, ApJ, 661, 801
Olivier, G. M., Berg, D. A., Chisholm, J., et al. 2021, ApJ, submitted [arXiv:2109.06725]
Otí-Floranes, H., Mas-Hesse, J. M., Jiménez-Bailón, E., et al. 2012, A&A, 546, A65
Ott, J., Walter, F., & Brinks, E. 2005, MNRAS, 358, 1453
Pahl, A. J., Shapley, A., Steidel, C. C., Chen, Y., & Reddy, N. A. 2021, MNRAS, 505, 2447
Pahl, A. J., Shapley, A., Steidel, C. C., Reddy, N. A., & Chen, Y. 2022, MNRAS, 516, 2062
Papaderos, P., Fricke, K. J., Thuan, T. X., & Loose, H. H. 1994, A&A, 291, L13
Pellegrini, E. W., Oey, M. S., Winkler, P. F., et al. 2012, ApJ, 755, 40
Péquignot, D. 2008, A&A, 478, 371
Pilyugin, L. S., & Thuan, T. X. 2005, ApJ, 631, 231
Pilyugin, L. S., Grebel, E. K., Zinchenko, I. A., & Kniazev, A. Y. 2014, AJ, 148, 134
Plat, A., Charlot, S., Bruzual, G., et al. 2019, MNRAS, 490, 978
Polles, F. L., Madden, S. C., Lebouteiller, V., et al. 2019, A&A, 622, A119
Prestwich, A. H., Tsantaki, M., Zezas, A., et al. 2013, ApJ, 769, 92
Prestwich, A. H., Jackson, F., Kaaret, P., et al. 2015, ApJ, 812, 166
Ramambason, L., Schaerer, D., Stasińska, G., et al. 2020, A&A, 644, A21
Ramos Padilla, A. F., Wang, L., Ploeckinger, S., van der Tak, F. F. S., & Trager, S. C. 2021, A&A, 645, A133
Ramos Padilla, A. F., Wang, L., van der Tak, F. F. S., & Trager, S. 2022, A&A, submitted [arXiv:2205.11955]
Reddy, N. A., Steidel, C. C., Pettini, M., Bogosavljević, M., & Shapley, A. E. 2016, ApJ, 828, 108
Reines, A. E., Sivakoff, G. R., Johnson, K. E., & Brogan, C. L. 2011, Nature, 470, 66
Relaño, M., De Looze, I., Kennicutt, R. C., et al. 2018, A&A, 613, A43
Rémy-Ruyer, A. 2013, Ph.D. Thesis
Richardson, C. T., Simpson, C., Polimera, M. S., et al. 2022, ApJ, 927, 165
Rivera-Thorsen, T. E., Dahle, H., Chisholm, J., et al. 2019, Science, 366, 738
Robertson, B. E., Furlanetto, S. R., Schneider, E., et al. 2013, ApJ, 768, 71
Robertson, B. E., Ellis, R. S., Furlanetto, S. R., & Dunlop, J. S. 2015, ApJ, 802, L19
Rosdahl, J., Katz, H., Blaizot, J., et al. 2018, MNRAS, 479, 994
Rémy-Ruyer, A., Madden, S. C., Galliano, F., et al. 2014, A&A, 563, A31
Rémy-Ruyer, A., Madden, S. C., Galliano, F., et al. 2015, A&A, 582, A121
Saldana-Lopez, A., Schaerer, D., Chisholm, J., et al. 2022, A&A, 663, A59
Sánchez-Cruces, M., Rosado, M., Rodríguez-González, A., & Reyes-Iturbide, J. 2015, ApJ, 799, 231
Satyapal, S., Kamal, L., Cann, J. M., Secrest, N. J., & Abel, N. P. 2021, ApJ, 906, 35
Saxena, A., Pentericci, L., Ellis, R. S., et al. 2022, MNRAS, 511, 120
Schaerer, D., Contini, T., & Pindao, M. 1999, A&AS, 136, 35
Schaerer, D., Fragos, T., & Izotov, Y. I. 2019, A&A, 622, L10
Sembach, K. R., Howk, J. C., Ryans, R. S. I., & Keenan, F. P. 2000, ApJ, 528, 310
Senchyna, P., Stark, D. P., Mirocha, J., et al. 2020, MNRAS, 494, 941
Senchyna, P., Stark, D. P., Charlot, S., et al. 2021, MNRAS, 503, 6112
Senchyna, P., Stark, D. P., Charlot, S., et al. 2022, ApJ, 930, 105

Shapley, A. E., Steidel, C. C., Strom, A. L., et al. 2016, ApJ, 826, L24
Simmonds, C., Schaerer, D., & Verhamme, A. 2021, A&A, 656, A127
Stanway, E. R., & Eldridge, J. J. 2019, A&A, 621, A105
Stark, D. P., Walth, G., Charlot, S., et al. 2015, MNRAS, 454, 1393
Stasińska, G., Izotov, Y., Morisset, C., & Guseva, N. 2015, A&A, 576, A83
Steidel, C. C., Bogosavljević, M., Shapley, A. E., et al. 2018, ApJ, 869, 123
Sutherland, R., Dopita, M., Binette, L., & Groves, B. 2018, Astrophysics Source Code Library [record ascl:1807.005]
Thuan, T. X., Bauer, F. E., Papaderos, P., & Izotov, Y. I. 2004, ApJ, 606, 213
Trebitsch, M., Blaizot, J., Rosdahl, J., Devriendt, J., & Slyz, A. 2017, MNRAS, 470, 224
Ugryumov, A. V., Engels, D., Pustilnik, S. A., et al. 2003, A&A, 397, 463
Umeda, H., Ouchi, M., Nakajima, K., et al. 2022, ApJ, 930, 37
Vale Asari, N., Couto, G. S., Cid Fernandes, R., et al. 2019, MNRAS, 489, 4721
Vale Asari, N., Wild, V., de Amorim, A. L., et al. 2020, MNRAS, 498, 4205
Vallini, L., Ferrara, A., Pallottini, A., Carniani, S., & Gallerani, S. 2021, MNRAS, 505, 5543
Vanzella, E., de Barros, S., Castellano, M., et al. 2015, A&A, 576, A116
Vanzella, E., de Barros, S., Vasei, K., et al. 2016, ApJ, 825, 41
Vanzella, E., Nonino, M., Cupani, G., et al. 2018, MNRAS, 476, L15
Vanzella, E., Caminha, G. B., Calura, F., et al. 2020, MNRAS, 491, 1093
Verhamme, A., Orlitová, I., Schaerer, D., & Hayes, M. 2015, A&A, 578, A7
Verhamme, A., Orlitová, I., Schaerer, D., et al. 2017, A&A, 597, A13
Walter, F., Riechers, D., Novak, M., et al. 2018, ApJ, 869, L22
Wang, B., Heckman, T. M., Leitherer, C., et al. 2019, ApJ, 885, 57
Wang, B., Heckman, T. M., Amorín, R., et al. 2021, ApJ, 916, 3
Weilbacher, P. M., Monreal-Ibero, A., Verhamme, A., et al. 2018, A&A, 611, A95
Weingartner, J. C., & Draine, B. T. 2001, ApJ, 548, 296
Westmoquette, M. S., Smith, L. J., & Gallagher, J. S. 2008, MNRAS, 383, 864
Wofford, A., Vidal-García, A., Feltre, A., et al. 2021, MNRAS, 500, 2908
Wolfire, M. G., Hollenbach, D., & McKee, C. F. 2010, ApJ, 716, 1191
Wood, K., Hill, A. S., Joung, M. R., et al. 2010, ApJ, 721, 1397
Xiao, L., Stanway, E. R., & Eldridge, J. J. 2018, MNRAS, 477, 904
Xu, X., Henry, A., Heckman, T., et al. 2022, ApJ, 933, 202
Yoo, T., Kimm, T., & Rosdahl, J. 2020, MNRAS, 499, 5175
Youngblood, A. J., & Hunter, D. A. 1999, ApJ, 519, 55
Zastrow, J., Oey, M. S., Veilleux, S., McDonald, M., & Martin, C. L. 2011, ApJ, 741, L17
Zastrow, J., Oey, M. S., Veilleux, S., & McDonald, M. 2013, ApJ, 779, 76
Zhang, K., Yan, R., Bundy, K., et al. 2017, MNRAS, 466, 3217
Zurita, A., Beckman, J. E., Rozas, M., & Ryder, S. 2002, A&A, 386, 801

Appendix A: Additional notes and figures

In this section we gather additional figures that complement the main plots presented in this article. Those figures are mentioned throughout the article to illustrate some of the discussions but do not display essential results. We recall in the captions the corresponding sections to which they are related.

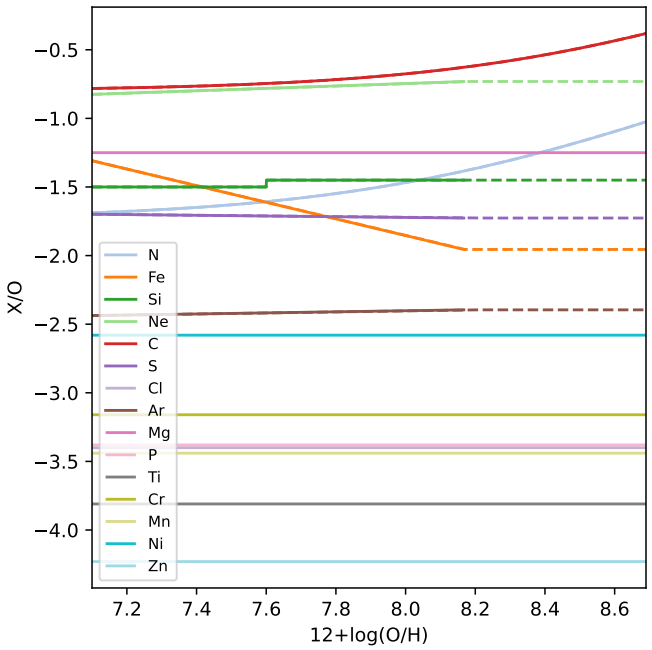

Fig. A.1. Abundance profile used to compute the grid of Cloudy models. This plot complements the description of the grid in Sect. 3 and illustrate the prescription described in Sect. 3.2.4.

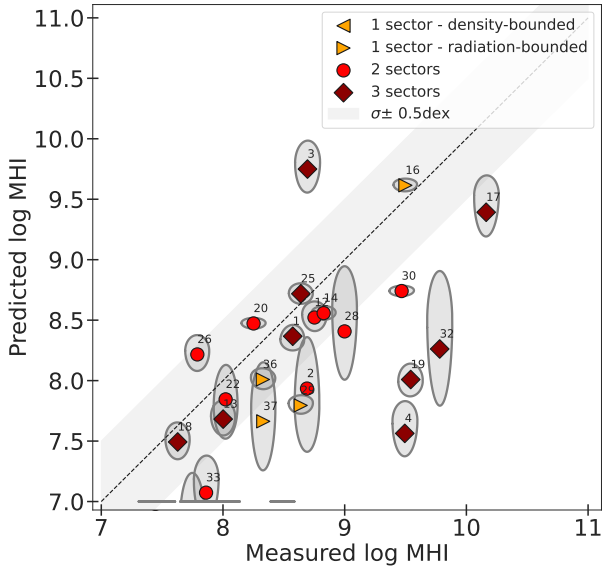

Fig. A.2. Predicted vs. measured H I masses. As discussed in Sect. 7.2.2, the predicted H I masses for the galaxies in our sample are consistent with the observed H I masses from Rémy-Ruyer et al. (2014). This agreement is at odds with the underprediction of H I column densities (see Table A.3) and is discussed in Sect. 7.

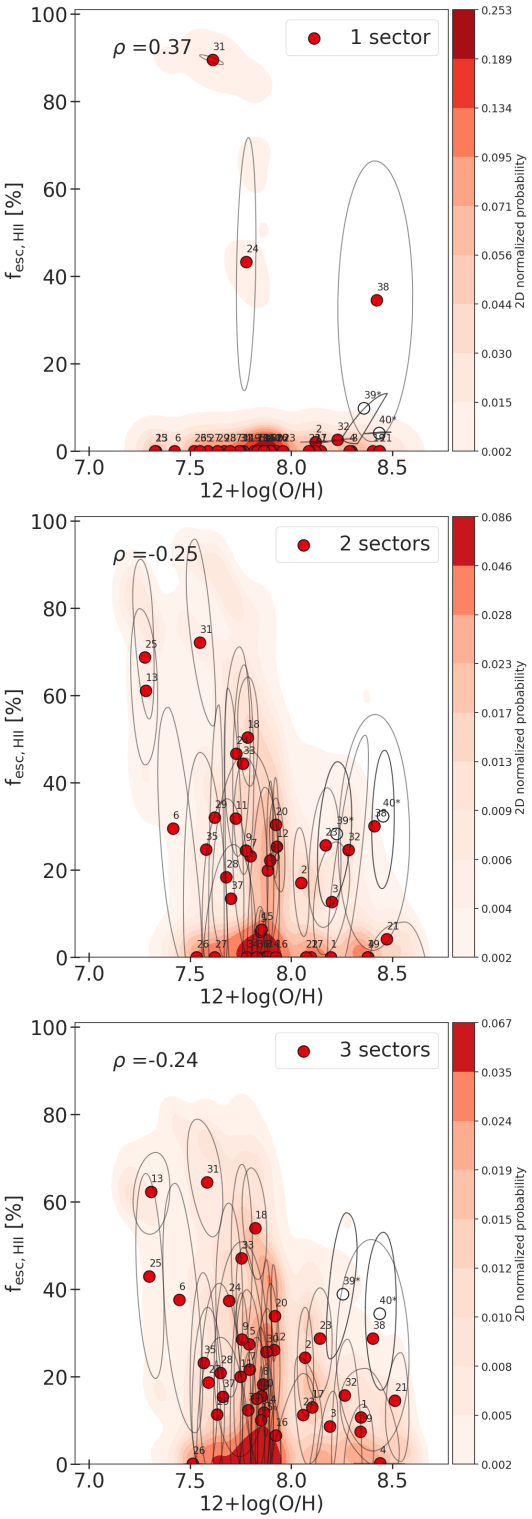

Fig. A.3. Metallicity-$f_{esc,HII}$ relation for a fixed number of sectors for all galaxies. These plots complement the Fig. 9 presented in Sect. 6.1 where the optimal number of sectors for a galaxy is chosen based on the score criteria presented in Sect. 4. We show here that the $f_{esc,HII}$-metallicity trend presented in Sect. 6.1 is robust and holds when we force an arbitrary number of sectors. We note that for single sector models, most solutions are completely radiation-bounded leading to escape fractions of 0%.

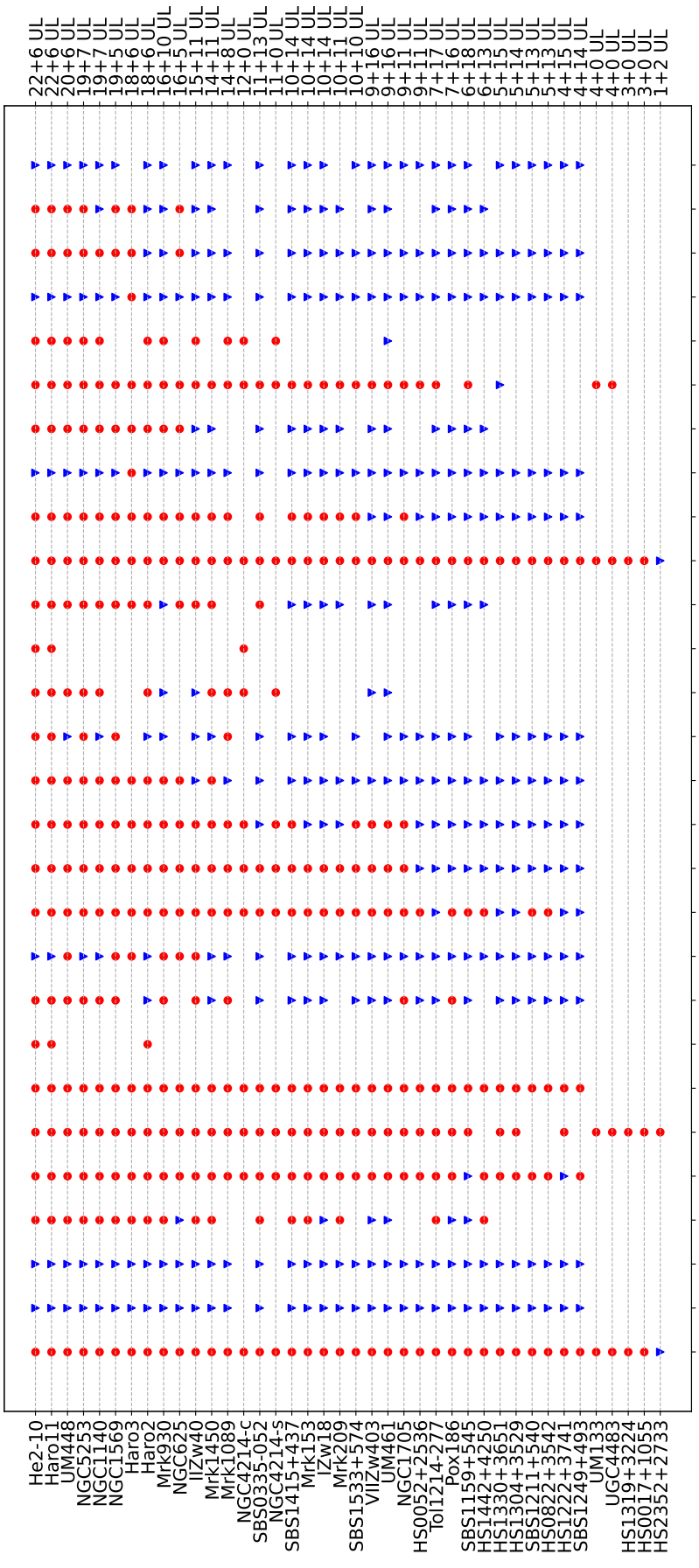

Fig. A.4. Detected lines (red dots) and upper limits (blue triangle) used in this paper. This figure illustrates the description of the available constraints for each galaxy in our sample that is presented in Sect. 2. The total number of detections and upper limits used for each object are reported on the right-hand side y-axis.

Table A.1. Best configuration selected for each galaxy and corresponding values of some indicators used to check model accuracy.

#	Galaxy	$N_{sectors,best}$	p-3σ [%][1]	$\mathcal{L}_{M,best}$ [2]	$\Delta\mathcal{L}^3_{M,1}$	$\Delta\mathcal{L}^3_{M,2}$	$\Delta\mathcal{L}^3_{M,3}$
1	Haro2	3.0	87.01	-45.49	-12.98	-4.41	0.0
2	Haro3	2.0	74.67	-60.51	-14.58	0.0	-1.28
3	Haro11	3.0	86.35	-51.95	-16.09	-1.32	0.0
4	He2-10	3.0	75.03	-93.15	-17.21	-2.59	0.0
5	HS0052+2536	2.0	86.21	-31.15	-0.17	0.0	-0.48
6	HS0822+3542	1.0	84.76	-13.19	0.0	-1.4	-3.62
7	HS1222+3741	1.0	88.61	-10.64	0.0	-0.96	-1.14
8	HS1304+3529	1.0	78.53	-18.49	0.0	-0.36	-0.75
9	HS1319+3224	3.0	94.67	-3.02	-1.25	-0.88	0.0
10	HS1330+3651	1.0	89.63	-24.35	0.0	-2.07	-0.64
11	HS1442+4250	3.0	74.18	-21.04	-0.23	-0.38	0.0
12	IIZw40	2.0	96.07	-39.4	-13.23	0.0	-1.6
13	IZw18	3.0	87.91	-31.75	-10.68	-2.24	0.0
14	Mrk153	2.0	97.54	-17.36	-1.42	0.0	-0.72
15	Mrk209	2.0	88.04	-24.83	-2.81	0.0	-0.64
16	Mrk930	1.0	98.09	-26.55	0.0	-0.67	-2.18
17	Mrk1089	3.0	80.25	-43.25	-10.14	-1.52	0.0
18	Mrk1450	3.0	87.57	-45.83	-9.31	-0.07	0.0
19	NGC1140	3.0	76.21	-73.47	-0.72	-3.14	0.0
20	NGC1569	2.0	74.77	-62.13	-5.98	0.0	-0.5
21	NGC1705	2.0	74.48	-51.53	-1.65	0.0	-1.45
22	NGC5253	2.0	76.91	-51.4	-4.36	0.0	-0.84
23	NGC625	3.0	74.99	-57.43	-4.6	-1.31	0.0
24	Pox186	3.0	92.09	-12.49	-3.46	-1.45	0.0
25	SBS0335-052	3.0	80.58	-34.49	-2.05	-0.34	0.0
26	SBS1159+545	2.0	78.39	-22.8	-0.14	0.0	-1.57
27	SBS1211+540	1.0	89.23	-16.74	0.0	-1.5	-1.94
28	SBS1249+493	2.0	89.72	-8.14	-2.76	0.0	-0.66
29	SBS1415+437	1.0	77.33	-36.14	0.0	-1.13	-0.01
30	SBS1533+574	2.0	77.95	-32.51	-0.83	0.0	-0.05
31	Tol1214-277	1.0	61.86	-21.46	0.0	-4.99	-4.56
32	UM448	3.0	75.5	-67.58	-2.34	-4.63	0.0
33	UM461	2.0	84.32	-40.5	-10.49	0.0	-0.47
34	VIIZw403	2.0	83.39	-34.71	-1.59	0.0	-0.05
35	UGC4483	2.0	75.98	-11.14	-0.81	0.0	-0.14
36	UM133	1.0	99.15	-2.47	0.0	-1.09	-0.99
37	HS0017+1055	1.0	96.12	-2.77	0.0	-0.05	-0.09
38	HS2352+2733	1.0	96.6	-1.08	0.0	-0.17	-0.27
39	NGC4214-c	3.0	90.66	-17.95	-9.85	-0.39	0.0
40	NGC4214-s	2.0	95.49	-11.49	-19.3	0.0	-2.45

(1): Probability that the best model predictions are within 3σ of the observations, for all the lines used as constraints. (2): Absolute value of the logarithm of the marginal likelihood for the best configuration corresponding to $N_{sectors,best}$. (3): $\Delta\mathcal{L}_{M,i} = \ln(\mathcal{L}_{M,i}/\mathcal{L}_{M,best})$

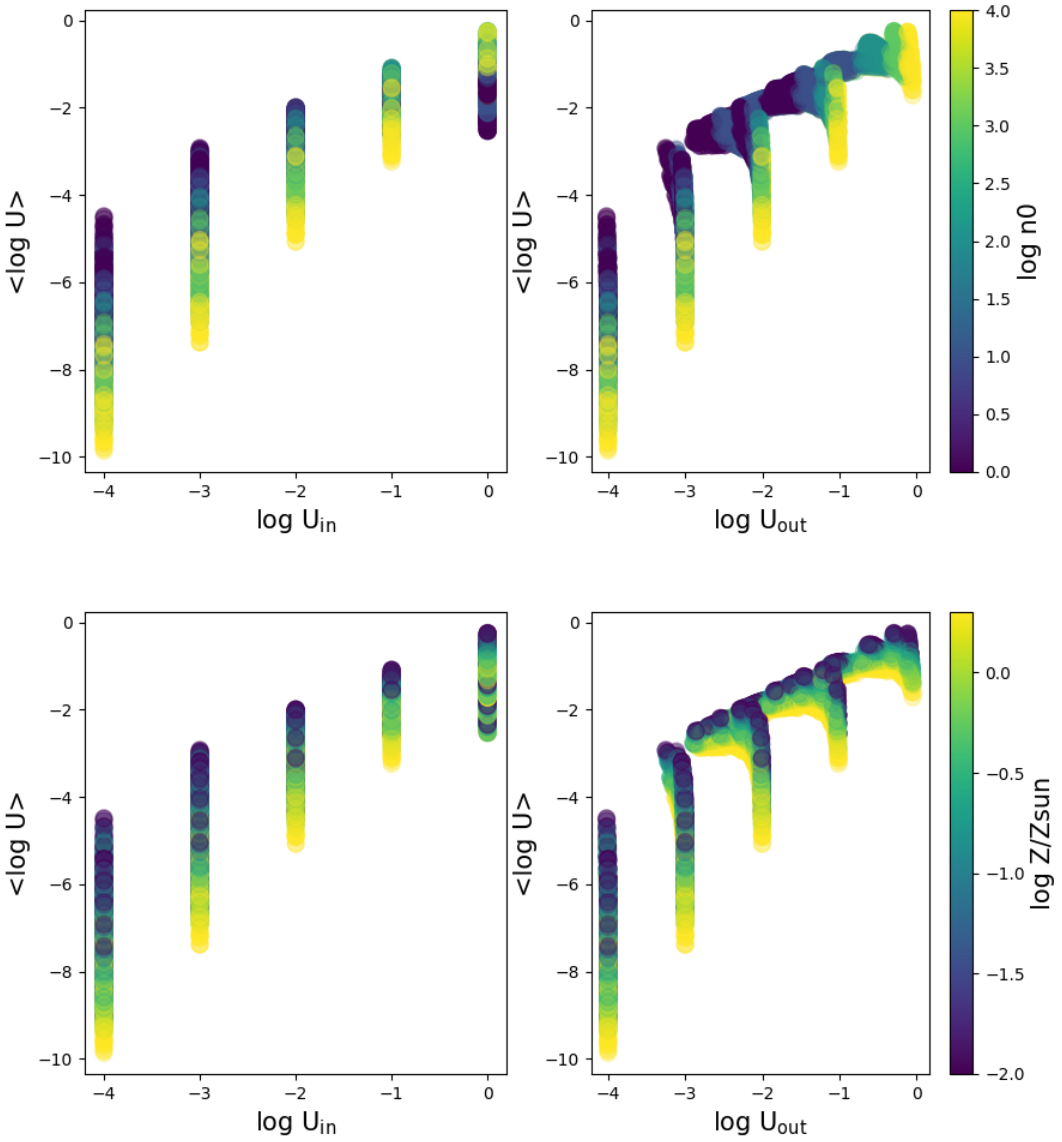

Fig. A.5. Evolution of the volume-averaged ionization parameter log $\langle U \rangle$ as a function log U_{in} and log U_{out} in the model grid. We show only a subsample corresponding to models with no X-ray source and stopping at the ionization front (cut=1). This figure complements the discussion on the possible caveats induced by the use of U_{in} as a primary parameter in our grid (see Sect. 3.2.3). We note that U_{in} only sets an upper limit on the actual volume averaged log $\langle U \rangle$. Variations in terms of metallicity and initial density n_0 yield different geometries, which cover a large range of volume averaged log $\langle U \rangle$ for the same U_{in}.

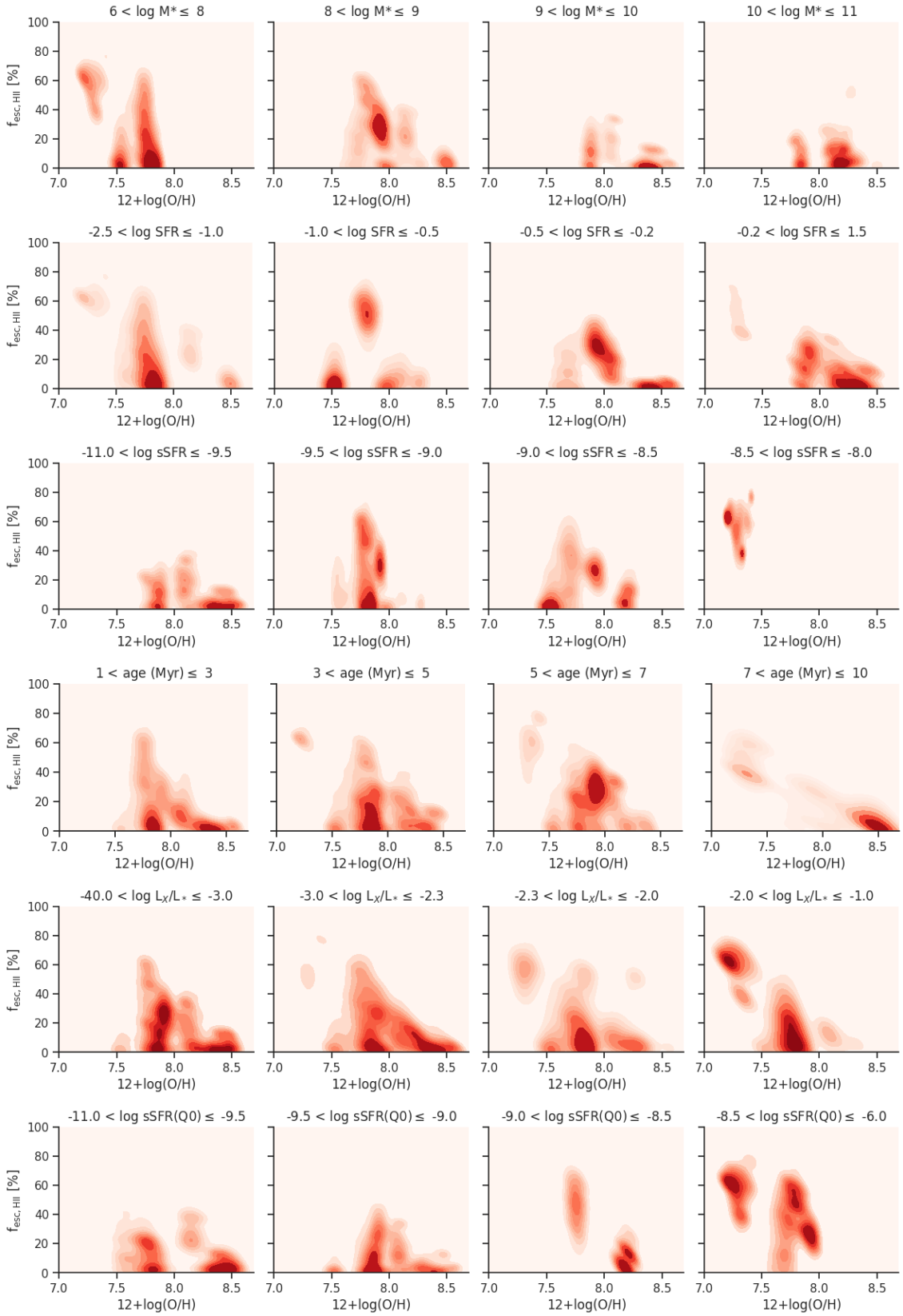

Fig. A.6. Impact of measured and inferred galactic parameters on the escape fraction vs metallicity relation. **Panels 1 to 3**:Impact of measured galactic parameters on the escape fraction vs metallicity relation: M_*, SFR, sSFR. **Panels 3 to 6**:Impact of inferred galactic parameters on the escape fraction vs metallicity relation: stellar age, Lx, sSFR(Q0). This figure complements the results presented in Sect. 6.1. It provides a reinterpretation of Fig. 9 but accounting for the additional dependences on galactic parameters by splitting the sample by bins of values for secondary parameters. Those secondary parameters are further examined in Sect. 6.2.

Table A.2. Spearman correlation and values of the fit for the best IR line ratio tracing $f_{esc,HII}$ ($\rho \geq 0.3$).

Line ratio	ρ	A	B
H1_12.3684m / Si2_34.8046m	0.527	130.42	185.563
S4_10.5076m / LTIR500.500m	0.519	149.899	456.264
H1_12.3684m / Fe2_25.9811m	0.486	98.124	30.459
H1_12.3684m / LTIR500.500m	0.485	98.586	456.934
S4_10.5076m / C2_157.636m	0.482	354.963	-6.656
H1_12.3684m / H2_17.0300m	0.471	35.689	22.384
H1_12.3684m / Fe2_17.9314m	0.47	114.504	-61.946
O3_88.3323m / N2_121.767m(1)	0.464	74.78	-163.043
H1_12.3684m / O1_63.1679m	0.464	672.149	1004.348
S4_10.5076m / Si2_34.8046m	0.464	183.689	-4.094
O3_88.3323m / N2_205.244m(1)	0.461	70.567	-176.577
S4_10.5076m / H2_17.0300m	0.461	40.522	-39.822
H1_12.3684m / N2_121.767m	0.46	52.458	-3.56
H1_12.3684m / C2_157.636m	0.459	162.834	258.149
H1_12.3684m / N2_205.244m	0.449	50.224	-24.644
S3_33.4704m / N2_121.767m	0.448	88.792	-155.97
Ar3_21.8253m / N2_121.767m	0.447	100.214	9.157
H1_12.3684m / Fe3_22.9190m	0.446	115.686	29.741
S4_10.5076m / N2_121.767m	0.445	65.545	-110.542
S4_10.5076m / N2_205.244m	0.444	57.369	-108.528
S4_10.5076m / H2_28.2130m	0.442	26.491	-28.109
N3_57.3238m / N2_121.767m	0.442	122.575	-157.575
O4_25.8832m / H2_17.0300m	0.439	58.881	-4.892
O4_25.8832m / N2_205.244m	0.439	28.946	-3.069
Ne3_15.5509m / N2_121.767m	0.433	86.891	-156.961
S4_10.5076m / Ne3_15.5509m	0.433	390.741	12.354
Ar3_8.98898m / N2_121.767m	0.432	99.306	-104.741
O4_25.8832m / C2_157.636m(2)	0.432	55.734	90.425
O4_25.8832m / H2_28.2130m	0.431	45.988	-12.765
O4_25.8832m / N2_121.767m	0.43	31.011	1.593
H1_12.3684m / Ar3_8.98898m	0.43	114.229	109.369
S3_18.7078m / N2_121.767m	0.428	85.685	-135.148
N3_57.3238m / N2_205.244m	0.428	99.587	-149.68
H1_12.3684m / S3_18.7078m	0.425	126.548	194.561
S3_33.4704m / N2_205.244m	0.424	76.793	-153.13
H1_12.3684m / Ar3_21.8253m	0.423	110.31	-18.305
Ne3_15.5509m / N2_205.244m	0.423	74.743	-152.017
O3_88.3323m / N3_57.3238m	0.422	297.947	-316.997
Ar3_21.8253m / N2_205.244m	0.419	83.176	-10.571
Fe3_22.9190m / N2_121.767m	0.418	90.861	-28.354
O3_88.3323m / S3_33.4704m	0.416	593.081	-270.111
S4_10.5076m / Ar3_8.98898m	0.414	206.794	-133.089
O4_25.8832m / Si2_34.8046m	0.41	45.209	78.238
S4_10.5076m / S3_18.7078m	0.41	320.263	-37.828
H1_12.3684m / S3_33.4704m	0.409	118.66	205.23
H1_12.3684m / O1_145.495m	0.408	-1864.77	-718.005
H2_9.66228m / H2_17.0300m	0.406	82.504	30.48
Ar3_8.98898m / N2_205.244m	0.406	84.64	-108.995
S4_10.5076m / S3_33.4704m	0.406	285.169	17.998
S3_18.7078m / N2_205.244m	0.406	73.539	-133.295
O3_88.3323m / Ar2_6.98337m	0.405	110.162	-273.012
H2_12.2752m / H2_17.0300m	0.405	101.622	69.297
H1_12.3684m / Ne2_12.8101m	0.405	60.832	47.995
S4_10.5076m / O1_63.1679m	0.404	-1237.19	-50.8
S4_10.5076m / Fe2_25.9811m	0.403	135.003	-158.851
O4_25.8832m / Fe2_25.9811m	0.403	41.974	22.719
S4_10.5076m / Ar2_6.98337m	0.403	91.98	-184.136
O3_88.3323m / Si2_34.8046m	0.403	227.883	-107.856
O4_25.8832m / Fe2_17.9314m	0.4	42.076	-9.965
S4_10.5076m / Fe2_17.9314m	0.4	187.512	-392.413
O3_88.3323m / H2_17.0300m	0.399	766.153	-1578.728
S4_10.5076m / Ar3_21.8253m	0.399	203.557	-364.099
S4_10.5076m / N3_57.3238m	0.399	164.224	-77.437
S4_10.5076m / Fe3_22.9190m	0.392	185.443	-234.334
O4_25.8832m / Ar2_6.98337m	0.392	36.116	-9.663
O3_88.3323m / Ar3_8.98898m	0.387	284.905	-318.54
S4_10.5076m / Ne2_12.8101m	0.385	78.214	-62.4
S3_33.4704m / Ar2_6.98337m	0.384	126.295	-249.556
Ne3_15.5509m / H2_17.0300m	0.38	-1741.288	2919.958
S3_18.7078m / Ar2_6.98337m	0.378	124.451	-226.135
O4_25.8832m / H2_12.2752m	0.377	244.608	-264.916
O3_88.3323m / Ne2_12.8101m	0.376	104.286	-149.435
O3_88.3323m / Ar3_21.8253m	0.376	293.241	-666.594
H1_12.3684m / N3_57.3238m	0.375	86.295	108.161
Fe3_22.9190m / N2_205.244m	0.374	83.949	-49.791
Ne3_15.5509m / Ar2_6.98337m	0.372	124.621	-254.015
O4_25.8832m / Ne2_12.8101m	0.37	34.232	31.129
S4_10.5076m / O3_88.3323m	0.368	532.903	278.691
Ne3_15.5509m / Si2_34.8046m	0.368	400.122	-32.15
O3_88.3323m / Fe3_22.9190m	0.368	310.557	-561.035
Ne5_24.2065m / H2_28.2130m	0.365	81.471	104.607
O4_25.8832m / Fe3_22.9190m	0.365	45.404	20.622
Ne3_15.5509m / Ar3_8.98898m	0.364	572.285	-415.788
Ne5_24.2065m / H2_17.0300m	0.363	86.667	152.264
Ne5_14.3228m / H2_28.2130m	0.362	77.986	107.966
O4_25.8832m / O1_63.1679m	0.362	82.694	120.192
Ne5_14.3228m / H2_17.0300m	0.36	80.987	151.057
O3_88.3323m / LTIR500.500m	0.36	224.753	564.26
O3_88.3323m / H2_28.2130m	0.359	31.542	-53.397
O3_88.3323m / C2_157.636m(1)	0.356	663.323	-302.962
S4_10.5076m / H2_12.2752m	0.355	-120.152	293.134
S3_33.4704m / Ne2_12.8101m	0.352	100.569	-78.356
O3_88.3323m / Fe2_17.9314m	0.351	212.062	-530.383
LTIR500.500m / N2_121.767m	0.351	110.802	-517.864
Ne3_15.5509m / H2_28.2130m	0.35	32.395	-40.951
Ne3_15.5509m / Ar3_21.8253m	0.35	613.279	-1152.203
O4_25.8832m / O1_145.495m	0.349	96.075	35.9
S3_18.7078m / Ne2_12.8101m	0.347	100.916	-64.499
H2_9.66228m / H2_28.2130m	0.346	64.117	9.046
Ar3_21.8253m / Ne2_12.8101m	0.346	137.09	130.134
Ne_315.5509m / Ne2_12.8101m(2)	0.344	104.82	-92.819
Ar3_21.8253m / Ar3_8.98898m	0.344	4298.962	5038.792
Ne3_15.5509m / Fe2_17.9314m	0.342	436.726	-948.463
O4_25.8832m / Ar3_8.98898m	0.34	45.886	51.343
O3_88.3323m / S3_18.7078m	0.337	591.496	-356.255
Ne3_15.5509m / S3_33.4704m	0.337	5021.612	-209.836
O4_25.8832m / N3_57.3238m	0.334	42.652	57.234
Ne3_15.5509m / LTIR500.500m	0.334	385.81	1115.446
O4_25.8832m / LTIR500.500m	0.334	47.755	219.676
Ar3_8.98898m / Ne2_12.8101m	0.334	124.763	-16.237
S4_10.5076m / O1_145.495m	0.333	-291.079	323.948
O4_25.8832m / Ar3_21.8253m	0.331	46.423	-1.869
Ne3_15.5509m / Fe3_22.9190m	0.33	496.239	-680.146
Ne3_15.5509m / N3_57.3238m	0.33	334.664	-182.823
O4_25.8832m / H2_9.66228m	0.328	490.167	-363.316
O4_25.8832m / S3_33.4704m	0.328	50.857	89.695
Ne5_24.2065m / N2_205.244m	0.326	33.235	58.843
Ne5_24.2065m / N2_205.244m	0.326	33.235	58.843
O4_25.8832m / S3_18.7078m	0.326	52.269	80.948
N3_57.3238m / Ar2_6.98337m	0.324	211.817	-319.099
Ne5_14.3228m / N2_205.244m	0.324	32.284	60.815
Ne5_24.2065m / N2_121.767m	0.322	35.963	69.431
Ne5_14.3228m / N2_121.767m	0.32	35.07	71.342
N3_57.3238m / Ne2_12.8101m	0.32	145.588	-44.665
O3_88.3323m / Fe2_25.9811m	0.318	171.32	-287.367
S3_18.7078m / Ar3_8.98898m	0.318	572.867	-293.951
Si2_34.8046m / N2_205.244m	0.316	75.291	-127.086
Si2_34.8046m / N2_121.767m	0.313	89.742	-135.462
H1_12.3684m / Ne3_15.5509m	0.312	131.48	227.804
H2_12.2752m / H2_28.2130m	0.312	77.74	32.316
Ne5_24.2065m / Ar2_6.98337m	0.311	38.156	71.118
Ne5_24.2065m / H2_12.2752m	0.311	266.128	297.313
S3_18.7078m / Fe3_22.9190m	0.311	541.475	-634.385
S3_18.7078m / H2_17.0300m	0.31	-15169.963	20996.747

Table A.2. continued.

Line ratio	ρ	A	B
Ne5_14.3228m / Ar2_6.98337m	0.308	37.532	73.659
S3_18.7078m / Fe2_17.9314m	0.307	357.808	-684.135
Ne5_14.3228m / H2_12.2752m	0.307	226.448	275.349
S3_33.4704m / Fe3_22.9190m	0.299	531.155	-699.26
Ne5_24.2065m / Ne2_12.8101m	0.299	34.38	107.357
Ne3_15.5509m / S3_18.7078m	0.298	-20086.043	4185.186
S3_33.4704m / H2_17.0300m	0.298	1494.699	-2272.294
Ne5_14.3228m / Ne2_12.8101m(2)	0.297	33.683	108.606
Ne3_15.5509m / C2_157.636m	0.296	-1935.617	158.812
O4_25.8832m / Ne3_15.5509m	0.295	54.356	92.73
Lbol / LTIR500.500m(3)	0.744	94.186	-7.617
Lbol / O2_3726+9A	0.657	83.845	-199.221
Lbol / S2_6716+30A	0.654	89.291	-289.073
Lbol / S3_9069+532A	0.643	166.175	-442.796
Lbol / N2_6548+84A	0.633	56.068	-182.48
Lbol / O1_6300+63A	0.626	90.88	-388.582
Lbol / O2_7320+30A	0.561	99.197	-416.007
He1_4471.49A / LTIR500.500m	0.485	101.138	389.835
O3_5007+4959A / LTIR500.500m	0.482	129.647	218.446
He2_4685.64A / O1_6300+63A	0.474	30.351	37.215
O3_4363.21A / LTIR500.500m	0.474	89.804	330.476
H1_6562.81A / LTIR500.500m	0.467	90.408	185.749
He1_4471.49A / S2_6716+30A	0.46	98.497	63.539
He1_4471.49A / O1_6300+63A	0.459	100.364	-38.5
He1_4471.49A / O2_3726+9A	0.456	85.291	133.629
He1_7065.22A / LTIR500.500m	0.455	91.203	365.114
He2_4685.64A / S2_6716+30A	0.452	31.702	66.998
He1_4471.49A / N2_6548+84A	0.448	61.434	35.734
H1_6562.81A / O1_6300+63A	0.446	89.915	-196.585
He2_4685.64A / N2_6548+84A	0.445	26.832	54.478
H1_6562.81A / S2_6716+30A	0.441	76.735	-77.414
O3_4363.21A / O1_6300+63A	0.441	78.211	-35.368
H1_6562.81A / O2_3726+9A	0.439	73.687	-12.733
O3_4363.21A / N2_6548+84A	0.435	56.741	23.599
H1_6562.81A / N2_6548+84A	0.433	55.108	-64.184
O3_5007+4959A / N2_6548+84A	0.432	74.729	-121.723
He2_4685.64A / O2_3726+9A	0.431	31.665	92.547
He1_7065.22A / O1_6300+63A	0.431	84.558	-15.205
S3_9069+532A / N2_6548+84A	0.43	90.149	-58.788

Table A.2. continued.

Line ratio	ρ	A	B
O3_4363.21A / S2_6716+30A	0.43	104.103	26.316
He1_7065.22A / N2_6548+84A	0.43	53.519	44.727
He1_7065.22A / S2_6716+30A	0.429	85.351	71.304
O3_5007+4959A / S2_6716+30A	0.428	130.992	-201.584
O3_4363.21A / O2_3726+9A	0.428	83.402	108.7
O3_5007+4959A / O1_6300+63A	0.428	122.712	-307.378
He1_7065.22A / O2_3726+9A	0.428	74.339	130.894
O3_4363.21A / O2_7320+30A	0.426	94.94	-46.28
O3_5007+4959A / O2_3726+9A(3)	0.425	116.705	-78.682
He2_4685.64A / O2_7320+30A	0.417	34.159	36.552
He1_4471.49A / S3_9069+532A	0.417	165.225	216.826
O2_3726+9A / N2_6548+84A	0.415	201.998	-192.16
O2_7320+30A / N2_6548+84A	0.414	162.477	132.045
O3_5007+4959A / S3_9069+532A	0.41	371.586	-352.117
He2_4685.64A / LTIR500.500m	0.409	33.221	177.064
O3_4363.21A / S3_9069+532A	0.406	144.841	156.083
He1_4471.49A / O2_7320+30A	0.404	112.029	-6.214
S3_9069+532A / S2_6716+30A(3)	0.402	211.532	-128.321
S3_9069+532A / O2_3726+9A	0.396	185.958	44.773
He1_7065.22A / O2_7320+30A	0.392	89.482	-7.637
He1_7065.22A / S3_9069+532A	0.39	121.28	187.334
H1_6562.81A / S3_9069+532A	0.39	136.392	-66.812
O3_5007+4959A / O2_7320+30A	0.385	154.476	-389.291
O3_4363.21A / O3_5006.84A	0.382	254.49	482.617
O3_4363.21A / O3_5007+4959A	0.382	249.591	505.516
He2_4685.64A / S3_9069+532A	0.378	36.926	103.908
H1_6562.81A / O2_7320+30A	0.374	98.921	-209.783
S2_6716+30A / N2_6548+84A	0.373	157.312	-5.541
O2_7320+30A / S2_6716+30A	0.347	840.02	810.936
O2_7320+30A / O2_3726+9A	0.343	437.593	811.678
LTIR500.500m / N2_6548+84A	0.326	151.553	-489.71
S3_9069+532A / O1_6300+63A	0.319	199.35	-316.807
O3_4363.21A / He1_4471.49A	0.306	628.672	-88.671
He2_4685.64A / He1_4471.49A	0.305	51.199	68.665
He2_4685.64A / H1_6562.81A	0.304	54.035	170.776

The fits are of the form Y = AX + B where Y = $f_{\mathrm{esc,HII}}$ [%] and X = $\log_{10}$(R) for a given line ratio R. (1): See Figs. 12 and 13. (2): See Figs. 14 and 15. (3): See Figs. 16 and 17.

Table A.3. Table comparing the predicted H I column densities and H I gas masses to values from the literature.

#	Galaxy	log MH $I^{(0)}_{pred}$	log MH $I^{(1)}_{obs}$	log N(H I)$_{max,pred}$	log N(H I)$_{min,pred}$	log N(H I)$_{obs}$
1	Haro2	$8.37^{+0.09}_{-0.07}$	$8.58^{+0.06}_{-0.08}$	19.47	18.79	19.6–20.5[2]
2	Haro3	$7.93^{+0.70}_{-0.46}$	$9.05^{+0.01}_{-0.01}$	18.92	18.29	
3	Haro11	$9.75^{+0.21}_{-0.15}$	$8.70^{+0.18}_{-0.30}$	19.08	18.73	
4	He2-10	$7.56^{+0.18}_{-0.33}$	$8.49^{+0.03}_{-0.03}$	19.33	18.74	
5	HS0052+2536	$9.79^{+0.06}_{-0.06}$	<10.68	20.0	18.54	
6	HS0822+3542	$4.94^{+0.84}_{-1.71}$	$7.75^{+0.09}_{-0.11}$	21.67	21.67	
7	HS1222+3741	$8.39^{+0.17}_{-0.35}$		20.16	20.16	
8	HS1304+3529	$8.83^{+0.08}_{-0.07}$		20.42	20.42	
9	HS1319+3224	$6.4^{+1.07}_{-1.25}$		18.46	17.86	
10	HS1330+3651	$8.43^{+0.46}_{-1.05}$		20.37	20.37	
11	HS1442+4250	$5.48^{+0.59}_{-0.56}$	$8.49^{+0.01}_{-0.01}$	20.84	19.29	
12	IIZw40	$8.52^{+0.11}_{-0.11}$	$8.75^{+0.06}_{-0.08}$	19.18	18.96	
13	IZw18	$7.68^{+0.13}_{-0.11}$	$8.0^{+0.06}_{-0.08}$	20.27	19.12	21.0–21.5[2]; 21.28±0.03[3]
14	Mrk153	$8.56^{+0.06}_{-0.06}$	<8.84	19.61	18.4	
15	Mrk209	$6.87^{+0.08}_{-0.10}$	$7.44^{+0.03}_{-0.03}$	19.87	18.59	
16	Mrk930	$9.62^{+0.04}_{-0.06}$	$9.5^{+0.05}_{-0.05}$	19.54	19.54	
17	Mrk1089	$9.39^{+0.21}_{-0.26}$	$10.17^{+0.03}_{-0.04}$	18.96	18.33	
18	Mrk1450	$7.49^{+0.15}_{-0.17}$	$7.63^{+0.06}_{-0.08}$	19.27	18.81	
19	NGC1140	$8.01^{+0.15}_{-0.14}$	$9.54^{+0.10}_{-0.13}$	18.43	18.28	
20	NGC1569	$8.47^{+0.04}_{-0.04}$	$8.25^{+0.06}_{-0.08}$	19.26	18.98	
21	NGC1705	$6.63^{+0.12}_{-0.09}$	$7.88^{+0.05}_{-0.06}$	19.1	17.97	
22	NGC5253	$7.85^{+0.35}_{-0.30}$	$8.03^{+0.02}_{-0.03}$	18.83	18.79	
23	NGC625	$5.46^{+0.66}_{-0.63}$	$8.04^{+0.07}_{-0.09}$	19.16	18.72	
24	Pox186	$5.3^{+1.14}_{-1.71}$	<6.30	19.75	19.3	
25	SBS0335-052	$8.72^{+0.09}_{-0.09}$	$8.64^{+0.04}_{-0.05}$	21.29	18.85	21.4–21.7[2]; 21.7±0.05[3]
26	SBS1159+545	$8.22^{+0.11}_{-0.15}$	<7.80	21.04	20.13	
27	SBS1211+540	$6.69^{+0.40}_{-0.51}$	$7.75^{+0.06}_{-0.08}$	20.56	20.56	
28	SBS1249+493	$8.41^{+0.37}_{-0.45}$	$9.0^{+0.17}_{-0.29}$	20.11	19.65	
29	SBS1415+437	$7.79^{+0.07}_{-0.10}$	$8.64^{+0.04}_{-0.05}$	20.8	20.8	
30	SBS1533+574	$8.74^{+0.04}_{-0.04}$	$9.48^{+0.05}_{-0.05}$	19.57	19.08	
31	Tol1214-277	$2.41^{+0.73}_{-0.23}$	<8.51	19.93	19.93	
32	UM448	$8.26^{+0.31}_{-1.05}$	$9.78^{+0.11}_{-0.14}$	19.26	18.68	
33	UM461	$7.07^{+0.24}_{-0.30}$	$7.86^{+0.02}_{-0.02}$	20.01	19.48	
34	VIIZw403	$6.06^{+0.33}_{-0.35}$	$7.51^{+0.03}_{-0.04}$	19.88	19.61	
35	UGC4483	$6.0^{+0.38}_{-0.51}$	$7.40^{+0.02}_{-0.02}$	20.44	19.42	
36	UM133	$8.01^{+0.08}_{-0.09}$	$8.33^{+0.02}_{-0.02}$	20.38	20.38	
37	HS0017+1055	$7.66^{+0.39}_{-0.41}$	<8.24	20.43	20.43	
38	HS2352+2733	$4.53^{+0.70}_{-0.78}$	<10.36	18.77	18.77	
39	NGC4214-c	$7.17^{+0.36}_{-0.42}$	<8.58	18.95	17.43	21.12±0.03[3]
40	NGC4214-s	$6.61^{+0.11}_{-0.13}$	<8.58	19.02	18.34	21.12±0.03[3]

This table complements the discussion from Sect. 7.2.2 and the H I gas mass comparison presented in Fig. A.2. (0): The interval is the highest density interval at 94%. (1): Rémy-Ruyer et al. (2014) (2): Kunth et al. (1998) (3): James et al. (2014)

9.4 Developments

In the paper presented above, we have examined the escape fractions of ionizing photons, which accounts for all the photons escaping with an energy above 1 Ryd. In the following, we develop and expand some of these results by looking at predictions for $f_{\rm esc}$ in different energy ranges.

9.4.1 $f_{\rm esc}$ in different energy ranges

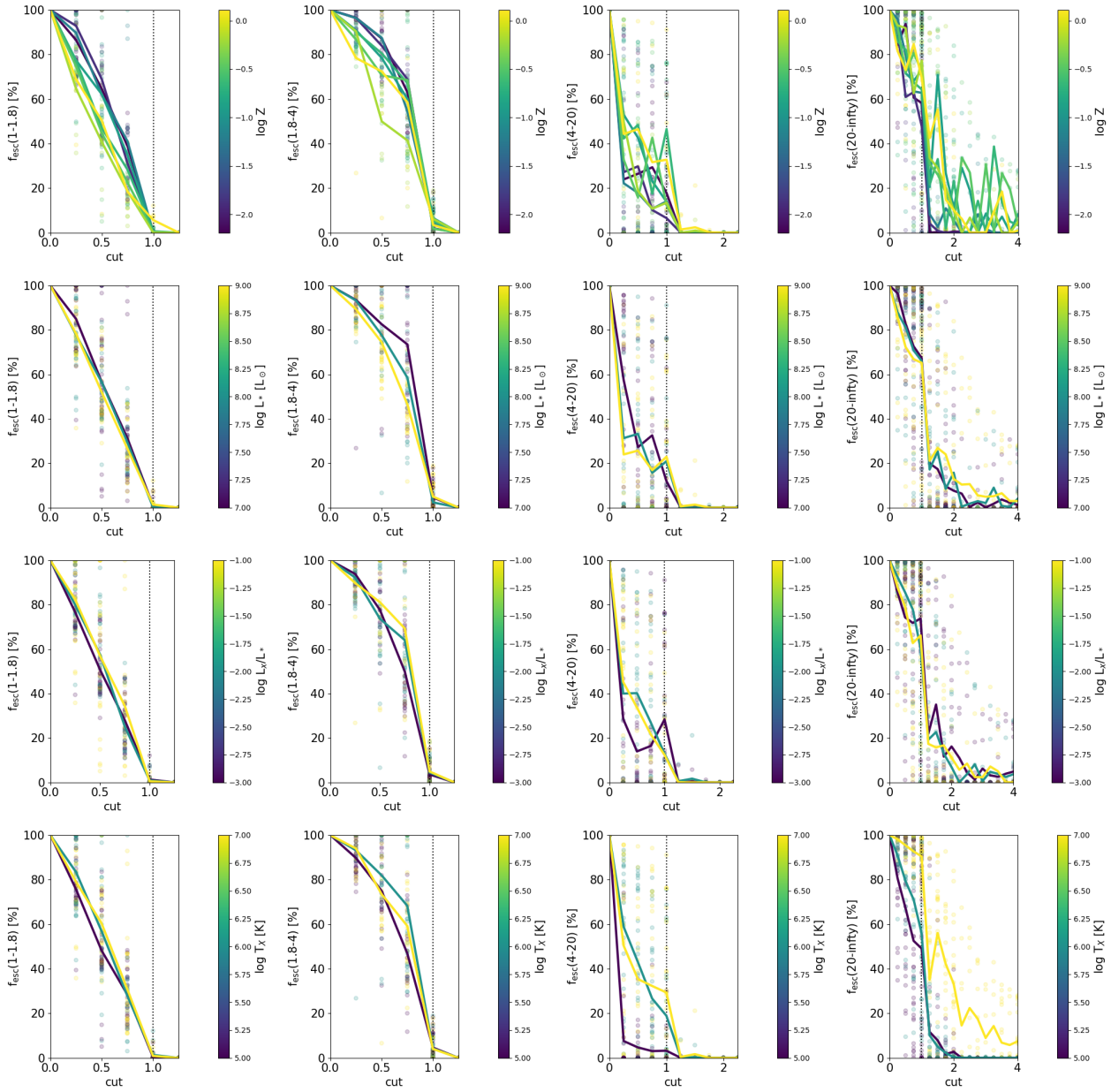

Figure 9.6: $f_{\rm esc}$ vs. cut relation in different energy ranges: 1-1.8 Ryd (1st column), 1.8-4 Ryd (2nd column), 4-20 Ryd (3rd column), and 20-∞ Ryd (4th column). The x-axis from the two last columns have been enlarged to show the escape fractions after the dissociation front (cut=2).

Among the possibilities enabled by our new modeling scheme, one can track the energy ranges of escaping photons. In particular, this option allows constraining the fraction of He-ionizing photons

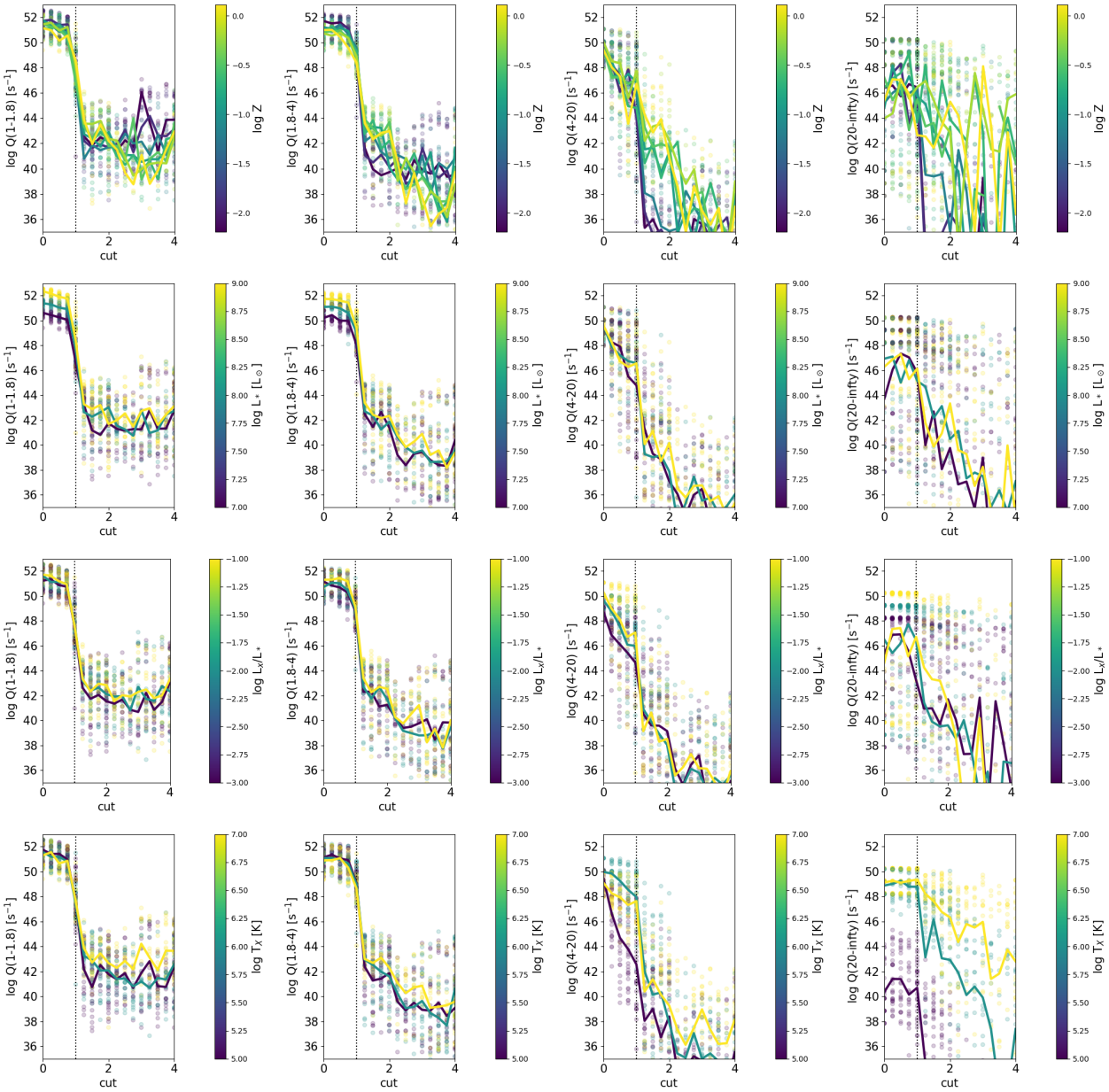

Figure 9.7: Q vs. cut relation in different energy ranges: 1-1.8 Ryd (1st column), 1.8-4 Ryd (2nd column), 4-20 Ryd (3rd column), and 20-∞ Ryd (4th column).

or to quantifying the energy range of photons that dominate the escape fractions. We define various energy ranges corresponding to ionization stages of known species: $1-\infty$, $1-1.8$, $1.8-4$ (He^0-ionizing photons), $4-20$ (He^+-ionizing photons), and $20-\infty$ Ryd. We extract the corresponding escape fractions in each range. The variation of escape fractions within those different energy bands are shown by the solid lines in Figure 9.6.

While $f_{esc}(1-1.8)$ and $f_{esc}(1.8-4)$ monotonically decrease throughout the H II region to reach approximately 0% at the IF, this is not the case of $f_{esc}(4-20)$ and $f_{esc}(20-\infty)$. In particular, $f_{esc}(4-20)$ exhibits a bump near the IF which is more prominent among models with T_X above 10^6K. Finally, photons with energy above 20 Ryd are able to penetrate deeper in the neutral gas, leading to escape fraction above 10%. Although significant escape fractions are expected at high energy, this escape fraction only represents the relative number of photons escaping per time unit, normalized by the

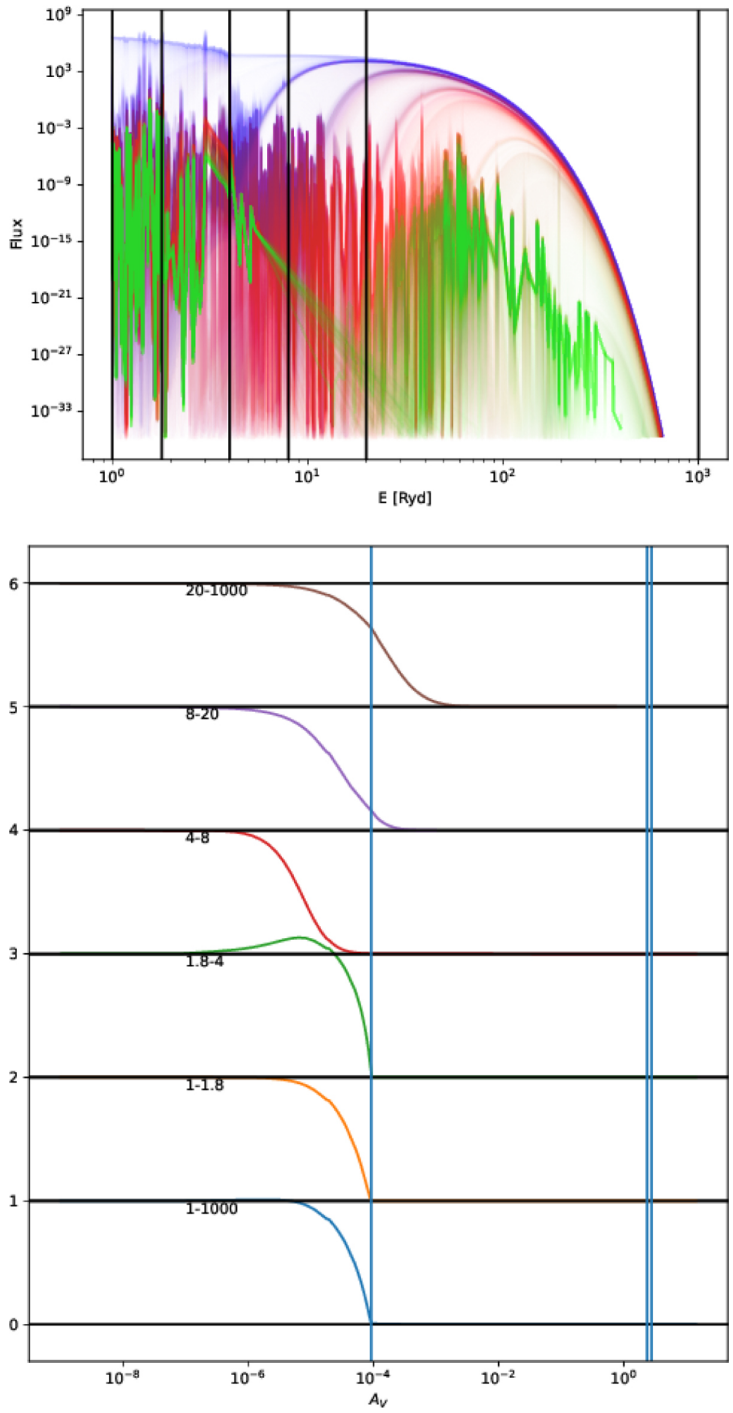

Figure 9.8: **Top**: Shape of the radiation field spectrum for each zone calculated in `Cloudy` for a radiation-bounded model, from the illuminated edge of the cloud (blue) to the last zone (green). The vertical lines show the 1, 1.8, 4, 8, 20, and 2000 Ryd energy values. The spectrum is heavily and rapidly absorbed right above 4 Ryd due to He^{2+} absorption. **Bottom**: Evolution of the integrated flux across the cloud depth (normalized to the value at the illuminated edge) for different energy ranges in Rydberg. The vertical lines show the ionization front, PDR, and C/CO transitions from left to right. The radiation field > 20 Ryd is able to penetrate in the neutral gas (top panel) but this contributes little to the total escape fraction (bottom panel). The kink for the 1.8-4 Ryd range is due to secondary ionization following He recombination. Taken from LR22.

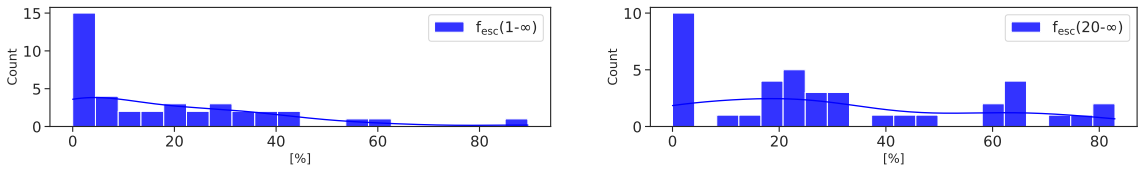

Figure 9.9: f_{esc}(1-∞) and f_{esc}(20-∞) distributions in the DGS.

number of photons produces. Hence, large escape fractions in a given energy do not necessarily translate into large absolute number of escaping photons.

In Figure 9.7, we show the total number of escaping photons per time unit (Q) with respect to the cut parameter. This allows us to disentangle which energy range contributes the most to the total escape fraction of ionizing photons. We see in Figure 9.7 that, within the H II region, the escape is clearly dominated by photons with energy between 1 and 4 Ryd. However, both Q(1-1.8) and Q(1.8-4) drop after the IF (cut= 1) by approximately 10 orders of magnitude, indicating that most photons with energy below 4 Ryd are consumed within the H II region. In the PDR (1 <cut< 2), $Q(4-20)$ is of the same order of magnitude as $Q(1-1.8)$ and $Q(1.8-4)$ and may significantly contribute to the total photon escape. While the escape fraction for photons with $E > 20$ Ryd is very large, we note that the absolute number of photons escaping in that energy range is almost always negligible, except for models with T_X=10^7 K.

While it is possible to compute such quantities for the whole DGS sample, the posterior PDFs of the escape fractions of high-energy photons (e.g., above 4 Ryd) are not well constrained since lines with high-ionization potential are not often detected. However, in the study of I Zw 18 presented in LR22, we include several lines with high ionization potential (i.e., He II 4685.64Å, [Ne III] 3869+968Å, [O IV] 25.8832μm, [Ne III] 15.5509μm). We present in Figure 9.8 the results obtained for the escape fractions in various energy ranges. When including infrared and optical emission lines from both the H II region and neutral gas, and a contribution from an X-ray source associated with the stellar cluster, the escape fraction $f_{esc}(1-\infty) \approx 59$ % lies between $f_{esc}(1-1.8) \approx 52$ % and $f_{esc}(1.8-4) \approx 78$ %, that is, in the range where most of the input radiation field energy lies. The escape fraction $f_{esc}(4-20)$, on the other hand, is remarkably lower (≈ 26 %), which is due to the absorption by He^{2+} close to the illuminated edge of the cloud (see Fig. 9.8). Nevertheless, the $f_{esc}(4-20)$ values remain significant due to the fact that the medium gradually becomes more transparent to photons significantly above 4 Ryd. As a matter of fact, $f_{esc}(20-\infty)$, which traces the escape of soft to hard X-ray photons is ≈ 87 % in the model.

Even though soft X-ray photons ($E \lesssim 1$ keV) deposit their energy in the neutral gas (Lebouteiller et al. 2017), $f_{esc}(20-\infty)$ remains large when including constraints from the neutral gas (≈ 80 %), which is due to the fact that the mixing weight of the radiation-bounded sector is relatively low. Most X-ray photons therefore escape the H II region, but the global X-ray escape fraction reaching the IGM ultimately depends on the distribution of diffuse H I gas. Since low-metallicity star-forming dwarf galaxies are known to harbor bright X-ray sources (e.g., Basu-Zych et al. 2016; Ponnada et al. 2020; Lehmer et al. 2020), this makes such sources interesting suspects for the IGM heating at $z > 10$ (Mirabel et al. 2011).

Whether these conclusions, derived for the lowest metallicity in our sample, can be generalized

for the whole DGS remains to be investigated. As shown in Figure 9.9, I Zw 18 is among the highest $f_{esc}(20-\infty)$ values predicted in the DGS. Nevertheless, when comparing the histogram of the average predictions for individual galaxies, we can see that values for $f_{esc}(20\text{-}\infty)$ are globally larger than f_{esc} in other energy ranges for the whole sample. It may also be interesting to note that among the 15 galaxies which are predicted to have average $f_{esc}(1-\infty)$ of 0%, only 10 of them are found to have $f_{esc}(20-\infty)$ of 0%. This highlights the fact that even in non-leaking galaxies (in terms of global escape fractions), part of the high energy photons with ($E \geq 20$ Ryd) may still escape.

9.4.2 Power-law models

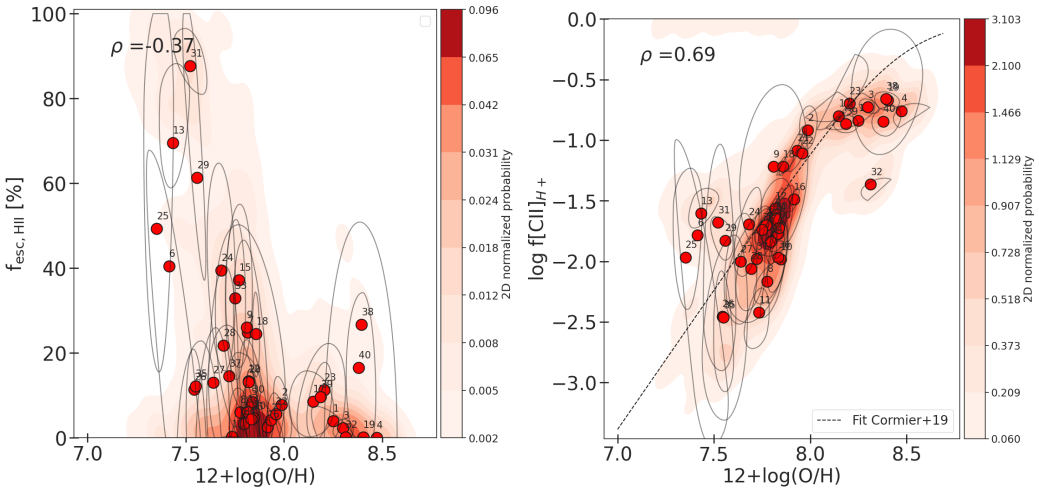

Figure 9.10: $f_{esc,HII}$ and f[C II] vs. metallicity for power-law models.

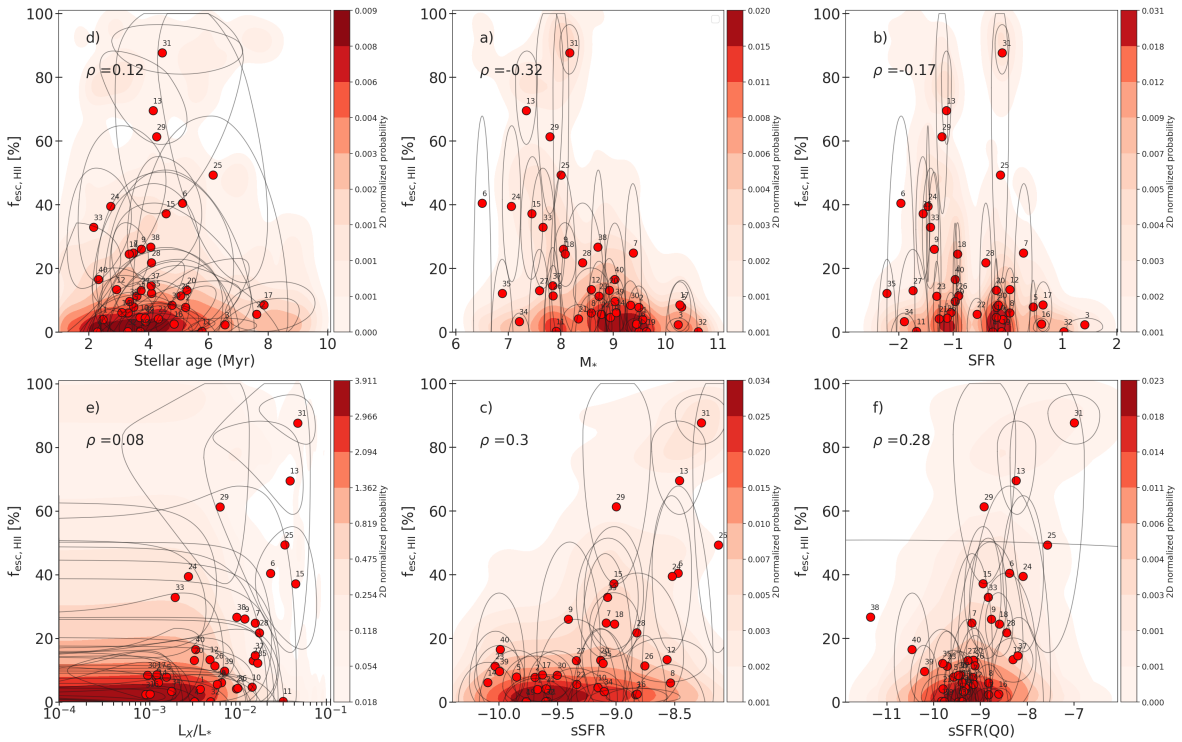

Figure 9.11: $f_{esc,HII}$ dependencies for power-law models.

While combining different components in a multisector topology works well enough to reproduce the signatures arising from the H II region and PDR, we will see in Chapter 10 that they fall short to reproduce the signatures from the molecular region, and in particular the CO lines. In the following chapter, we will hence introduce models based on the use of power-law distributions to describe the density, ionization parameters, and the depth (cut parameter). Those models will be further described in Section 10.3.5.

Although we do not need to reach that level of complexity in the study of the escape fraction, we re-examined the main trends that were derived in Ramambason et al. (2022) (presented in Section 9.3), using this time power-law distributions rather than a combination of a discrete number of sectors. In Figures 9.10 and 9.11, we show the main trends regarding the evolution of $f_{\rm esc,HII}$ and $f([\text{C II}])_{H^+}$. The prediction for $f_{\rm esc,HII}$ are globally consistent between multisector models and power-law distributions, with slightly higher values predicted for the latter configuration. We find that the trends presented throughout this chapter are robust to changes in the topology and are still clearly visible when using models based on power-law distributions. We further discuss the motivation to use this alternative topology in Section 10.3.5.

9.5 Summary

In Chapter 9, we presented a first application of `MULTIGRIS` concerning the study of the escape fraction of ionizing photons. As opposed to the predictions obtained for classically well-constrained quantities (see Chapter 8), we find that $f_{\rm esc}$ is sensitive to both the modeling assumptions made to predict $f_{\rm esc}$ within our grid and the inferred topology. Based on the definitions that we presented in Section 9.2, we derive the following results in Section 9.3:

- Our results confirm the porosity to UV photons of the H II regions at low-metallicity found in C19, with a more robust statistical method. This allows us to convert the ill-defined PDR covering factor estimates into PDFs for the escape fraction of photons from H II regions, $f_{\rm esc,HII}$.
- The inferred $f_{\rm esc,HII}$ in the DGS are high, reaching values as up as ~60% for some objects. Our results demonstrate the possibility to identify leaky-H II regions based solely on their emission lines signatures. They are, however, difficult to convert into global $f_{\rm esc}$ estimates without, among others, properly accounting for the DIG.
- We find that $f_{\rm esc,HII}$ correlates with sSFR and anti-correlate with metallicity. This fits in the picture of a feedback-regulated phenomenon with dependencies on galactic properties.
- We put to test several infrared (e.g., [O III] $\lambda 88\mu$m/[C II] $\lambda 158\,\mu$m, [O III] $\lambda 88\mu$m/[N II] $\lambda\lambda 122{,}205\mu$m, [O IV] $\lambda 25\mu$m/[C II] $\lambda 158\mu$m, or [Ne V] $\lambda\lambda 14{,}24\mu$m/[Ne II] $\lambda 12\mu$m) and optical lines ratios (e.g., [O III] $\lambda\lambda 4959{,}5007$Å/[O II] $\lambda\lambda 3726,\ 3728$Å, [S III] $\lambda\lambda 9068{,}9532$Å/[S II] $\lambda\lambda 6716{,}6731$Å) and identify those which correlate best with $f_{\rm esc,HII}$.

While such high leakage of UV photons is observed in resolved studies of H II regions, it is however not observed at galactic scale in the nearby universe. We argue that this absence of observed LyC-leakage may be due to instrumental limitations, which prevent the direct detection of LyC photons in local dwarf galaxies. However, several effects, which are not accounted for in our models,

might also affect the global escape fractions in the DGS. In particular, we discussed the potential role of a diffuse gas component that might reduce LyC-leakage at large scales. Potential improvements in the modeling framework will be discussed in Section 11.1.1.

Finally, as a development (Section 9.4) we discussed the possibility and interest to study $f_{\rm esc}$ within different energy ranges. While constraining the escape of photons with high energy requires specific signatures, this splitting in different energy ranges is especially important to understand how photons propagate in the neutral atomic and molecular phases. While the relative escape fractions of high energy (> 1.8 Ryd) photons can reach high values, the corresponding absolute ionizing fluxes remain low, except when luminous and hot X-ray sources are present. Nevertheless, even small amounts of high energy radiation can significantly affect the neutral gas heating in the PDR and molecular phase (see Section 1.2.2), especially in low-metallicity environments (see Section 1.2.3). We also revisit our results with a somewhat more complex topology based on power-law distributions of parameters and find that the results presented in this chapter globally still hold under those different assumptions on the gas distribution.

Chapter 10

Estimating CO-dark H_2 masses

In Chapter 9 we presented a detailed analysis of the escape fraction of ionizing photons from H II regions in the DGS galaxies. We found that substantial $f_{esc,HII}$ values are expected in low-metallicity environments and that the values are mainly driven by metallicity and specific star formation rate. Our main conclusions are summarized in Section 9.5.

In the current chapter, we examine how galactic properties, including the metallicity and hardness of the ionizing radiation, impact the global neutral gas reservoirs in the DGS. To do so, we perform the same kind of analysis as in Chapter 9, but focusing this time on deriving PDFs for the H II, H I, dust, and especially H_2 masses, including the ubiquitous CO-dark H_2 reservoir (see definition in Section 2.2.4). In Section 10.1, we briefly recall the many challenges of estimating the total molecular H_2 masses and the importance of accounting for the CO-dark gas reservoir to robustly derive CO-to-H_2 conversion factors. In Section 10.2, we discuss previous results from Madden et al. (2020) (hereafter; M20) predicting that this CO-dark gas component can dominate the molecular mass budget, especially in low-metallicity galaxies such as the DGS. In Section 10.3, we introduce different configurations, assuming either a single-sector, a combination of discrete sectors, or power-law distributions of parameters, in order to reproduce the CO(1-0) emission line flux. In Section 10.4, we explore how those different modeling assumptions impact the parameters controlling the geometries of CO-emitting clouds, specifically in terms of A_V (or depth). Finally, in Section 10.5, we compare and discuss the mass estimates obtained with these different configurations, as well as the CO-dark gas fractions and CO-to-H_2 conversion factors predicted by each model.

10.1 Context

10.1.1 Molecular gas traced by H_2 and CO lines

Quantifying the molecular gas content of galaxies is a complex task which requires choosing the proper proxies. Because the H_2 molecule has no permanent dipolar momentum, the cold H_2 reservoirs are not directly observable in emission. Moreover, the H_2 rovibrational lines only probe a partial reservoir in which the gas is sufficiently hot to reach the energy necessary to excite the lowest ortho and para transitions (E/k~510 K and 1015K, Dabrowski 1984). The lowest vibrational transition is

even more difficult to excite (E/k = 6471 K). Thus, H_2 emission lines are difficult to detect and do not trace the whole molecular gas reservoir.

To overcome this issue, other tracers have been proposed to trace the total molecular gas mass. Among the most classically used, the CO molecule (and especially its most abundant isotopologue $^{12}C^{16}O$) has been used to derive conversion factors that allow estimating the total H_2 molecular gas mass from the integrated CO luminosity. As opposed to H_2, the CO dipolar momentum is low and its ground transition is associated to a low excitation energy (E/k~5.53K). Additionally, CO is among the most abundant molecules in the ISM, making it an ideal tracer to probe the cold molecular reservoir.

However, changes are expected at low-metallicity regarding both the H_2 production mechanisms (e.g., due to the low dust-to-gas mass ratio, since H_2 forms predominantly on dust surfaces) and CO production mechanisms (e.g., due to low C and O abundances at low metallicity). Under such low-metallicity conditions, CO may be efficiently photodissociated by strong radiation fields while part of the H_2 reservoir remains shielded in clumps which are unseen is CO. Several studies have hence reported spatial offsets between the CO-emitting regions and the total H_2 distribution, both in the Milky-Way (e.g., Grenier et al. 2005; Ackermann et al. 2012) and in extragalactic environments (e.g., Poglitsch et al. 1995; Madden et al. 1997). This molecular gas reservoir, which is unseen in CO, is usually referred to as the "CO-dark" molecular gas (e.g., Wolfire et al. 2010).

In particular, large CO-dark gas fractions have been reported in low-metallicity dwarf galaxies (e.g., Poglitsch et al. 1995; Madden et al. 1997; Schruba et al. 2012; Amorín et al. 2016; Schruba et al. 2017; Cormier et al. 2014; Madden et al. 2020) and several proxies have been proposed to quantify the mass of this CO-dark component. Accounting for this unseen molecular mass is crucial to understand how the CO-to-H_2 conversion factor varies with metallicity (e.g. Glover & Mac Low 2011; Schruba et al. 2012; Bolatto et al. 2013; Genzel et al. 2015; Amorín et al. 2016; Accurso et al. 2017; Tacconi et al. 2018; Madden et al. 2020).

10.1.2 Deriving CO-to-H_2 conversion factors

The attempts to derive the H_2 molecular gas mass from the CO integrated luminosity have relied on several assumptions that we now briefly discuss. For simplicity, we will refer in the following only to the most common isotopologue $^{12}C^{16}O$ and specifically to the CO(1-0) rotational transition (λ=2.6 mm). For a more detailed discussion on the validity of those assumptions in different environments, we refer to Bolatto et al. (2013), on which the following section is based.

The principle of using a CO-to-H_2 conversion factor relies on the assumption that the total H_2 column density of a molecular cloud can be simply related to its CO intensity, such that:

$$N(H_2) = X_{CO} \times I_{CO}, \tag{10.1}$$

with $N(H_2)$ the H_2 column density in cm^{-2} and I_{CO} the integrated CO(1-0) line intensity in $K\,km\,s^{-1}$. Equation 10.1 can be translated into a mass-luminosity relation by integrating over the CO-emitting area of the molecular cloud and correcting for the mass contribution of heavier elements mixed into the molecular gas, leading to:

$$M_{\rm mol} = \alpha_{\rm CO} \times L_{\rm CO}, \tag{10.2}$$

where $M_{\rm mol}$ is usually expressed in $M_\odot$ and $L_{\rm CO}$ in $\rm K\,km\,s^{-1}\,pc^2$.

In theory, there is not an exactly linear relation between the molecular mass of a molecular cloud and its CO luminosity. Because of this, the estimated $\alpha_{\rm CO}$ remains weakly dependent on the mass of the cloud. However, as explained in Bolatto et al. (2013), a simple linear relation can be used as a fairly good approximation for resolved GMCs that match the following conditions:

- The GMC is virialized (i.e., its total kinetic energy due to the turbulent motions of its constituent parts (T) balances its gravitational potential energy (U) as 2T+U=0).
- Its mass is dominated by H_2 molecules.
- It follows a size-line width relation (Larson 1981): $\sigma \propto \rm R^{0.5}$ with R the projected radius in pc and σ the 1D velocity dispersion in $\rm km\,s^{-1}$.
- The GMC has an approximately constant temperature.

Even in the ideal case of a virialized GMC, the CO-to-H_2 conversion factor is expected to depend on environmental parameters such as the gas density and temperature.

Nevertheless, Equation 10.2 is not only used to describe the case of resolved molecular clouds but has also been extensively applied to galaxies in which GMCs are unresolved. In this case, the proportionality between the mass and CO luminosity comes from the assumption that the integrated CO luminosity is powered by an ensemble of non-overlapping CO-emitting clouds. This results in a proportionality between the CO intensity and the number of clouds observed within the telescope beam. Under the assumption that those clouds are at virial equilibrium, CO hence allows tracing the cumulated mass of those individual clouds.

Because the $\alpha_{\rm CO}$ conversion factor initially relates to individual molecular cloud properties, it may locally vary within galaxies. Deriving a single $\alpha_{\rm CO}$ value for a given galaxy is an obvious simplification which bears significant uncertainties. At galactic scale, several dependencies of $\alpha_{\rm CO}$ are expected with, for example, the average surface brightness of the CO-emitting gas, its temperature, and the size distribution of GMCs within galaxies.

Nevertheless, in Galactic conditions, the $\alpha_{\rm CO}$ conversion factor is relatively well constrained. In our Galaxy, independent measurements (see Section 10.1.3) all yield values in global agreement within uncertainties and close to the value of $X_{\rm CO} = 2 \times 10^{20}\,\rm cm^{-2}\,(K\,km\,s^{-1})^{-1}$ (Bolatto et al. 2013). This value corresponds to $\alpha_{\rm CO}$= 4.3 $\rm M_\odot\,(K\,km\,s^{-1}\,pc^2)^{-1}$. In the current study, however, we compare the mass estimates from our models with values excluding the contribution of helium mass mixed within molecular gas, which lead to a correction of 36%:

$$\alpha_{\rm CO,MW} = 3.2\,M_\odot(\rm K\,km\,s^{-1}\,pc^2)^{-1} \tag{10.3}$$

This value may vary in galaxies with different physical and chemical conditions. In particular, the metallicity of the ISM and the potential presence of CO-dark gas reservoirs strongly impact the $\alpha_{\rm CO}$

values that are derived. Thus, numerous efforts have been made in order to calibrate the metallicity dependencies of α_{CO} based the information collected with independent proxies that trace molecular gas.

10.1.3 Calibrating α_{CO}

Estimating how the molecular gas fractions evolves with metallicity is an important step to connect the star formation rates to the total molecular gas reservoir. In particular, constraining the evolution of the α_{CO} factor with metallicity may help us understand the apparent deviation of low-metallicity dwarf galaxies from the Schmidt-Kennicutt relation. Assuming the linear relation between CO luminosity and total molecular mass (see Equation 10.2), the calibration of the α_{CO} conversion factor relies on: (1) the possibility to detect CO emission, (2) the use of an independent proxy that allows probing the total molecular gas.

To do the latter, different proxies have been used to estimate the total molecular gas mass in different environments. At the GMCs scale, several studies have relied on estimating the virial masses (M_{vir}) based on the clouds kinematics and found a remarkably tight correlation between L_{CO} and M_{vir} (Solomon et al. 1987; Scoville et al. 1987). The analysis of kinematics is, however, only applicable in resolved GMCs and other proxies, based on galactic observables, have been proposed to probe the molecular gas content in the Milky-Way and in extra-galactic environments.

γ-ray emission

Because γ-rays interact with the hydrogen nuclei that form H atoms, H^+ ions, or H_2 molecules, they are ideal tracers to probe the total interstellar gas, including the CO-dark reservoir. They have been used in Galactic studies for comparison with CO and H I observations. By doing so, several papers have shown the presence of a molecular gas reservoir within our galaxy, unseen with CO or H I (e.g., Grenier et al. 2005; Ackermann et al. 2012; Remy et al. 2018; Hayashi et al. 2019). While this proxy is, in principle, one of the most accurate methods, its use has been so far limited to Galactic studies because of the need of high-resolution observations and sensitivity.

Dust continuum

Dust measurements can be used to estimate the total gas masses of galaxies. This proxy has been extensively used in both Galactic (Planck Collaboration XIX 2011; Reach et al. 2015) and extra-Galactic environments (e.g., Magnelli et al. 2012; Sandstrom et al. 2013; Genzel et al. 2015; Tacconi et al. 2018; Zavala et al. 2022; Tokuda et al. 2021). This approach must, however, rely on a conversion based on an assumed dust-to-gas mass ratio, which is known to vary strongly in the low-metallicity regime (e.g., Rémy-Ruyer et al. 2014; Galliano 2018; Galliano et al. 2021).

Absorption lines

Molecular lines observed in absorption, such as OH and HCO^+, have also been used as tracers of dark molecular gas in our Galaxy (e.g., Liszt & Lucas 1996; Lucas & Liszt 1996; Allen et al. 2015; Nguyen et al. 2018). However, the need for a sufficiently high sensitivity and resolution hampers their use in other galaxies. Absorption lines associated with ions that may co-exist within the CO-dark layers have also been used to estimate the column densities associated with the CO-dark component (e.g., S^+; Hernandez et al. 2021). Those estimates are nonetheless quite sensitive to variations of elemental abundances and depletion patterns. Finally, H_2 absorption features detected in the afterglow of GRBs can also serve as interesting proxies to probe the molecular gas content of distant GRB host galaxies (Heintz et al. 2021). More generally, the observation of H_2 absorption lines remains challenging due to the dust obscuration of the sightlines, which can only be performed with very sensitive instruments (such as, e.g., the LUVOIR telescope; The LUVOIR Team 2019).

Emission lines

Finally, several ionic and atomic species that may co-exist with H_2 in this CO-dark phase have been proposed as possible tracers of the molecular gas reservoir. In particular, carbon ions which partly originate from the photodissociation of CO have received a lot of attention. Specifically, the C^+ ion, which emits the bright IR fine structure line [C II]158μm was used to show the presence CO-dark gas in local dwarf galaxies (Poglitsch et al. 1995; Madden et al. 1997). This line has been extensively used to trace molecular gas in local and high-redshift galaxies (e.g., Zanella et al. 2018; Béthermin et al. 2020; Vizgan et al. 2022). The [C I] emission lines have also been investigated as a promising tracer of molecular gas (e.g., Jiao et al. 2019; Crocker et al. 2019; Heintz & Watson 2020; Dunne et al. 2021, 2022).

10.2 Evidence for a CO-dark gas reservoir in the DGS

10.2.1 Reference study

A systematic study of the "CO-dark" gas content in the DGS galaxies was motivated by their very high [C II]/CO(1-0) ratio and the possibility to probe molecular gas reservoirs down to extremely low-metallicity. A previous analysis performed in M20 has consisted in quantifying the total molecular gas content in DGS galaxies based on extensive suites of IR emission lines, including ionic and atomic tracers that may co-exist with H_2 in the CO-dark phase (e.g., C^+, C^0, O^0).

The strategy adopted in M20 uses the wealth of emission lines available for the DGS galaxies, and in particular the [C II]158μm line, to constrain the total H_2 gas masses based on predictions from models. This method allows comparing the total gas mass, predicted by models, to the CO-bright H_2 mass estimated from CO(1-0) measurements, assuming a galactic conversion factor (see Equation 10.3). The difference between the total and CO-bright H_2 mass hence provides an estimate of the CO-dark gas mass.

This work uses the numerous IR emission lines arising from the H II region, PDR, and molecular

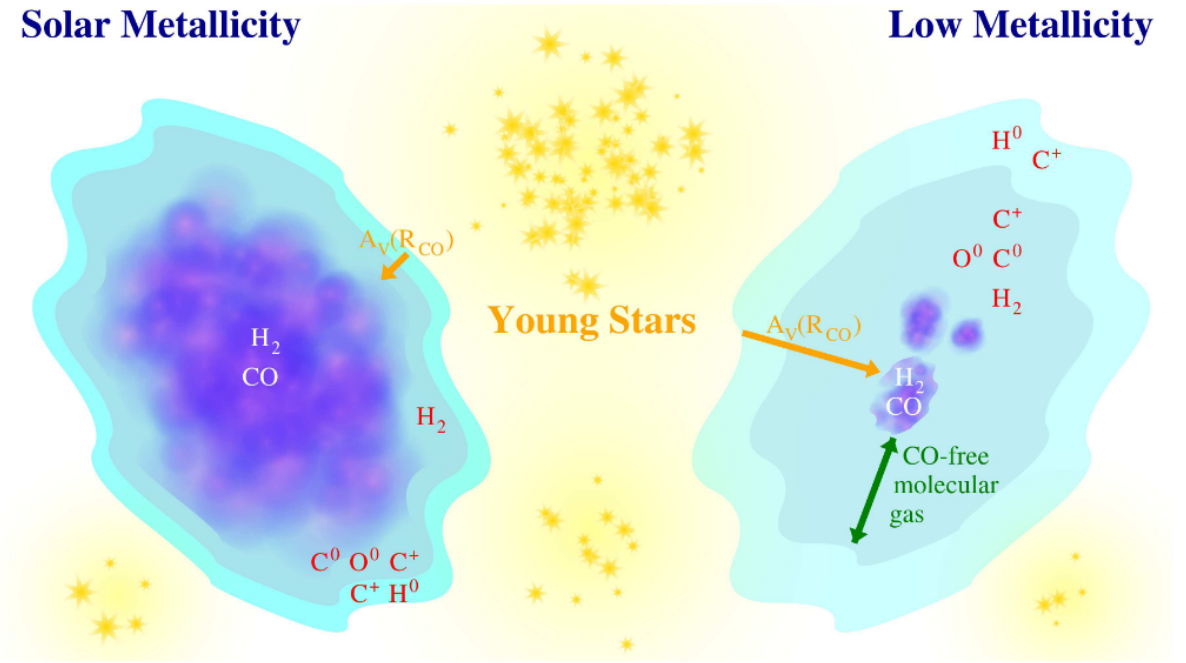

Figure 10.1: Schematics of the CO-dark gas reservoir at low- and solar-metallicity. Taken from Madden et al. (2020).

phase to constrain multiphase, single-component, `Cloudy` models and extract the total H_2 masses. We briefly describe the methodology which was adopted to derive the H_2 masses. The major differences between this approach and the one of the current study will be described in Section 10.2.2. Because of the difficulty to simultaneously fit all the lines with simple topological models (e.g., single sector), the approach to estimate H_2 gas masses in M20 relied on an iterative process:

- First, the values of the density and ionizing parameter values were selected based on a χ^2 minimization routine to reproduce all the available emission lines from the H II region and PDR, **excluding CO(1-0)**. This best fitting model corresponds to the solution reported in C19 for the single-sector models. We note that in the analysis from C19, those models were all stopped at A_V=5 by default, since no constraint was used to control the depth.

- In a second step, the A_V of the best fitting model was calculated a posteriori, while keeping all other parameters unchanged, in order to match the [C II]/CO(1-0) luminosity ratio. Based on this A_V, the total H_2 mass was calculated by integrating the mass of H_2 within the corresponding `Cloudy` model.

- When no CO detection was available, an upper limit on the maximum total H_2 mass was derived by integrating the model until $\tau_{CO} \approx 1$.

Based on this analysis, several interesting trends were reported:

- A large fraction (≥70% in all galaxies) of the H_2 gas mass in the DGS galaxies is found to be CO-dark. Hence, the CO(1-0) traces only a very small fraction of the total H_2 mass, M(H_2), in low-metallicity galaxies.

- There is a tight correlation between the [C II]158 μm and the total M(H_2) over the range of DGS metallicities.
- The effective A_V in the models is found to be anti-correlated with the [C II]/CO(1-0) luminosity ratio, while the fraction of CO-dark molecular gas correlates with the [C II]/CO(1-0) luminosity ratio.
- The α_{CO} values derived in M20 have a steeper dependency in metallicity than in most studies, although it resembles that of Schruba et al. (2012) derived assuming a constant H_2 depletion time τ_{dep} (i.e., $\alpha_{CO} = \tau_{dep} \times SFR/L_{CO}$).
- Taking into account the significant CO-dark H_2 reservoir in the star-forming dwarf galaxies shifts their position in the $\Sigma_{SFR} - \Sigma_{MH_2}$ plan. While the latter were above the Schmidt–Kennicutt relation for M(H_2) estimates based only on the observed CO(1-0), the estimates accounting for CO-dark gas yield a position in agreement with the relation derived for normal star-forming disc galaxies.
- Although it is more difficult to detect than [C II], [C I] also seems to be a reliable tracer of the M(H_2), particularly at low-metallicity where it shows little dependency in n_H and G_0.

10.2.2 Comparison with Madden et al. (2020)

We aim to reproduce and extend the results from M20 with the new statistical formalism presented in Chapter 7. One major improvement is the possibility to test different topological models (see Chapter 4) and examine the effect on the mass estimates that are derived. In order to facilitate the comparison with the results from M20 and before diving in the analysis of the new results, we first discuss some important differences between the two modelling strategies.

In M20, the choice of the best fitting models and of the stopping depth of the model are performed sequentially: first, the best model (in terms of ionization parameter, density, and metallicity) is selected based on the reduced χ^2 minimization calculated from the emission lines in the H II region and PDR. Then the stopping A_V, which was not a free parameter, is adjusted to match the [C II]/CO(1-0) line ratio. This strategy ensures that CO(1-0) is always well reproduced by design. Nevertheless, this somewhat modifies *a posteriori* the solution that was chosen from the χ^2 routine. Thus, CO(1-0) is reproduced at the expense of other lines (e.g., [O I], [C II]) which may be overestimated by the new model.

On the other hand, models run with `MULTIGRIS` are simultaneously fitting the whole set of lines used as constraints, including CO. While this approach allows a more consistent modeling, the resulting models do not always succeed in predicting CO(1-0) values in agreement with observations. One specificity of the plots that will be presented in this section is that we show only *predicted* values for the models, except in plots where we explicitly compare predictions to observations. This choice avoids us to define hybrid quantities (as in M20) which are, for example, ratios of quantities predicted by models and of observed quantities.

In M20, upper limits are used to derive the maximum masses of CO-bright, based on a Galactic α_{CO} conversion factor. While it is straightforward to convert upper limits on CO into upper limits

on the CO-bright M(H_2), the effect on the CO-dark gas estimates is less clear. In M20, these CO upper limits were not accounted for to derive the maximum CO-dark gas masses, with the latter being computed by integrating the models up to an arbitrary optical depth of $\tau_{CO} \approx 1$. Hence, it is not easy to interpret the CO-bright to CO-dark molecular gas ratio derived in galaxies in which CO is undetected. This issue arises from the mixing of a model-based estimate of the total molecular gas and an observation-based estimate for the CO-bright molecular gas.

To avoid this problem, we treat the upper limits directly within the Bayesian framework of `MULTIGRIS`. In the following sections, we show the values predicted from models, accounting for the upper limits, rather than the upper limits themselves. This difference is to be kept in mind when comparing the different plots showing the location galaxies for which only upper limits are available.

10.3 Reproducing CO(1-0) observations with topological models

We now present the available measurements of CO emission in the DGS and examine different types of model in order to reproduce the CO(1-0) detections.

10.3.1 CO measurements and unit conversions

Table 10.1: Table of CO measurements in the DGS. The upper limits correspond to 3σ level.

#	Galaxy	θ[″]	CO(1-0) [K km s^{-1}]	CO(1-0) [$10^{-20}\times$ W m^{-2}]	Reference
1	Haro 2	55	1.13±0.26	14.25±3.28	(1)
2	Haro 3	45	1.98±0.28	16.72±2.36	(2)
3	Haro 11	30	≤ 0.69	≤ 2.59	(3)
4	He2 10	55	4.85±0.24	61.18±3.03	(4)
5	II Zw 40	22	0.88±0.16	1.77±0.32	(2)
6	I Zw 18	≤ 1	≤ 0.75	≤ 0.003	(5)
7	Mrk 209	45	0.45±0.10	3.80±0.84	(6)
8	Mrk 930	22	0.14±0.07	0.28±0.14	(1)
9	Mrk 1089	22	1.53±0.30	3.10±0.61	(3)
10	NGC 1569	55	0.685±0.104	8.64±1.31	(7)
11	NGC 1705	45	≤ 0.03	≤ 0.25	(3)
12	NGC 5253	55	0.725±0.148	9.15±1.87	(8)
13	NGC 625	30	2.05±0.63	7.69±2.36	(3)
14	SBS 0335-052	≤ 1	≤ 5.60	≤ 0.023	(9)
15	UM 448	55	1.80±0.57	22.71±7.19	(2)
16	UM 461	55	≤ 0.78	≤ 9.84	(2)
17	VII Zw 403	55	≤ 0.99	≤ 12.49	(10)

(1): Thronson & Bally (1987) (conversion from Israel 2005)
(2): Sage et al. (1992)
(3): Cormier et al. (2014)
(4): Kobulnicky et al. (1995)

(5): Leroy et al. (2007)
(6): Young et al. (1995)
(7): Greve et al. (1996)
(8): Taylor et al. (1998)
(9): Dale et al. (2001) (conversion from Hunt et al. 2014)
(10): Leroy et al. (2005)

We gathered the measurements available for the DGS galaxies in the literature, with the associated uncertainties. Table 10.1 is based on quantities used in M20 but was updated with the most recent measurements, and with references for which uncertainties were available, leading to small variations for some sources. We also excluded galaxies for which only partial regions were observed. The CO measurements and corresponding references are reported in Table 10.1.

Most measurements are expressed in $\mathrm{K\,km\,s^{-1}}$ and are reported with the instrumental beam θ in arcseconds. Because most of the sources are unresolved, and due to the difficulty to derive a source size (see Cormier et al. 2014, for a more detailed discussion), we assumed that the angular size of the source is negligible compared to the telescope beam: $\Omega_S << \Omega_b$. Hence, we approximate the solid angle of the source convolved with the telescope beam $\Omega_{S*b} \approx \Omega_b \approx \theta^2$, with θ the FWHM of the telescope beam in arcsecs. This assumption is equivalent to fixing arbitrarily the size of the source, which is unknown. The impact of such an assumption on the mass estimates that are derived are further discussed in Cormier et al. (2014).

We convert measurement in $\mathrm{K\,km\,s^{-1}}$ in $\mathrm{W\,m^{-2}}$, using the following formula, valid for unresolved lines within the Rayleigh-Jeans approximation:

$$[\mathrm{W\,m^{-2}}] = \frac{10^{-26}}{13.6} \times \frac{\nu[\mathrm{Hz}]}{\mathrm{c[km\,s^{-1}]}} \times \left(\frac{\theta[''}{\lambda[\mathrm{mm}]}\right)^2 \times [\mathrm{K\,km\,s^{-1}}], \tag{10.4}$$

with ν the frequency of the unresolved line, λ its wavelength, c the speed of light and θ the instrumental beam. For $\nu_{CO(1-0)}$=115.3 GHz and $\lambda_{CO(1-0)}$=2.6 mm, Equation 10.4 becomes:

$$[\mathrm{W\,m^{-2}}] = 4.17 \times 10^{-23} \times \theta^2 \times [\mathrm{K\,km\,s^{-1}}]. \tag{10.5}$$

This intensity can be converted into L_{CO} in $\mathrm{erg\,s^{-1}}$ using:

$$L_{CO}\,[\mathrm{erg\,s^{-1}}] = 10^7 \times 4\pi D_L^2 I_{CO}[\mathrm{W\,m^{-2}}] \tag{10.6}$$

with D_L the luminosity distance converted in meters (1 pc = 3.086×10^{16} m).

To calculate the α_{CO} conversion factor, one usually uses L_{CO} given in unit of $\mathrm{K\,km\,s^{-1}pc^2}$. The latter quantity can be calculated as (see formula from Solomon et al. 1997):

$$\begin{cases} L_{CO}\,[\mathrm{K\,km\,s^{-1}pc^2}] = I_{CO}\,[\mathrm{K\,km\,s^{-1}}] \times 23.5 \times \theta^2 \times D_L^2 \times (1+z)^{-3} \\ L_{CO}\,[\mathrm{K\,km\,s^{-1}pc^2}] = \frac{23.5}{4.17\times10^{-23}} \times D_L^2 \times (1+z)^{-3} = I_{CO}\,[\mathrm{W\,m^{-2}}] \times 5.63 \times 10^{23} \times D_L^2 \times (1+z)^{-3}. \end{cases} \tag{10.7}$$

where θ is the telescope beam in arcsecs, D_L the luminosity distance in Mpc, and z the redshift of the source. Finally, it is useful to note that, when converting directly our predictions in erg s^{-1} to K km s^{-1}pc^2, the conversion factor only depends on the redshift of the source:

$$[\mathrm{K\,km\,s^{-1}pc^2}] = \frac{5.63 \times 10^{23} \times (1+z)^{-3}}{10^7 \times 4\pi \times (10^6 \times 3.086 \times 10^{16})^2} = 4.70 \times 10^{-30} \times (1+z)^{-3}\,[\mathrm{erg\,s^{-1}}]. \quad (10.8)$$

10.3.2 Tension between H_2, CO, and [C II] lines

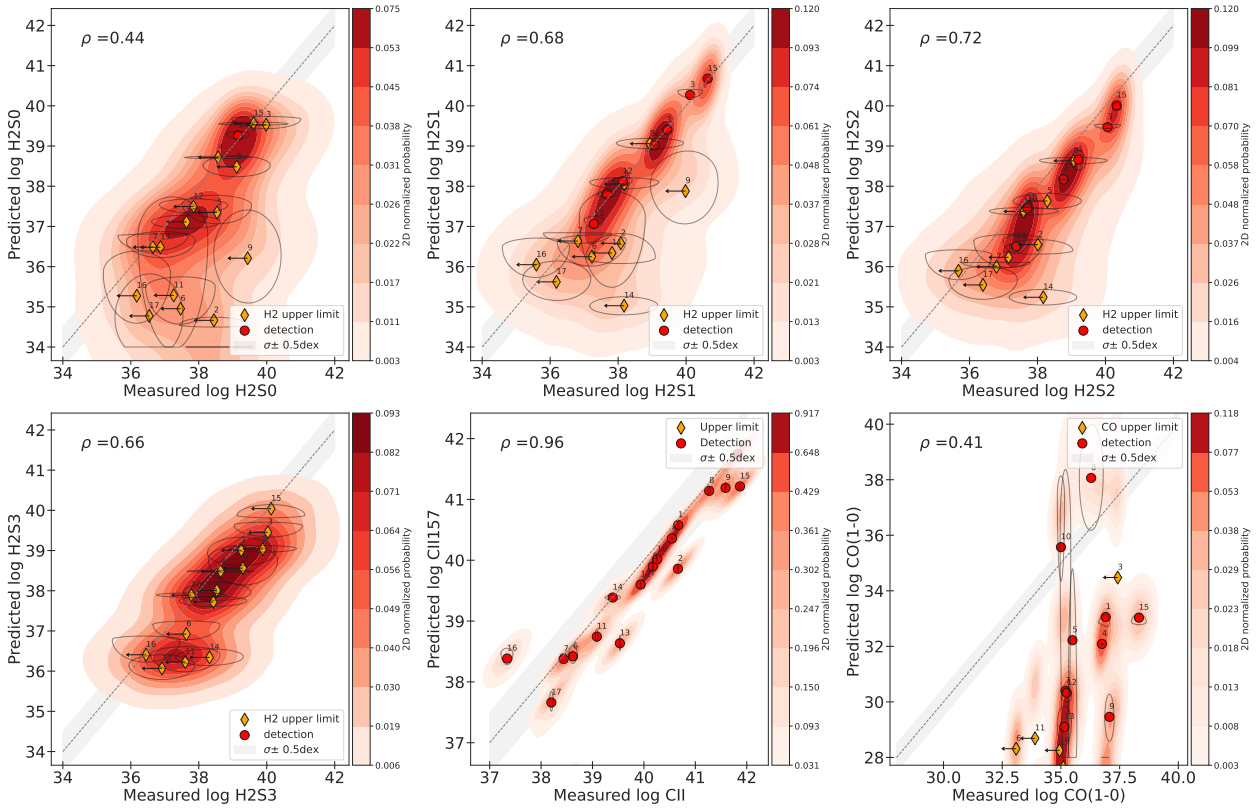

Figure 10.2: Comparison of predictions for emission lines from single-sector models including H_2 as constraint: H_2 S(0), H_2 S(1), H_2 S(2), H_2 S(3), [C II]158μm, and CO(1-0).

The list of constraints used in the models that will be presented in this section is composed of the same set of lines used to constrain models of the DGS galaxies in Section 9.3, with the addition of CO(1-0) line. One important change is that we removed the H_2 lines (which were mostly upper limits) as constraints. The removal of H_2 lines is motivated by several issues: first, H_2 lines were not used as constraints in the analysis of M20 to which we compare our results to. Second, the preliminary tests described below have shown that our code is, in fact, unable to simultaneously reproduce H_2 and CO(1-0) lines.

Indeed, as shown in Figure 10.2, when H_2 limits are respected, CO(1-0) is underpredicted by several orders of magnitude for most galaxies. To a lesser extent, this tension is also impacting [C II] predictions, which are in good agreement with observations but always slightly below. This tension persists with the various configurations that were tested (i.e., multisector and power-law, described in

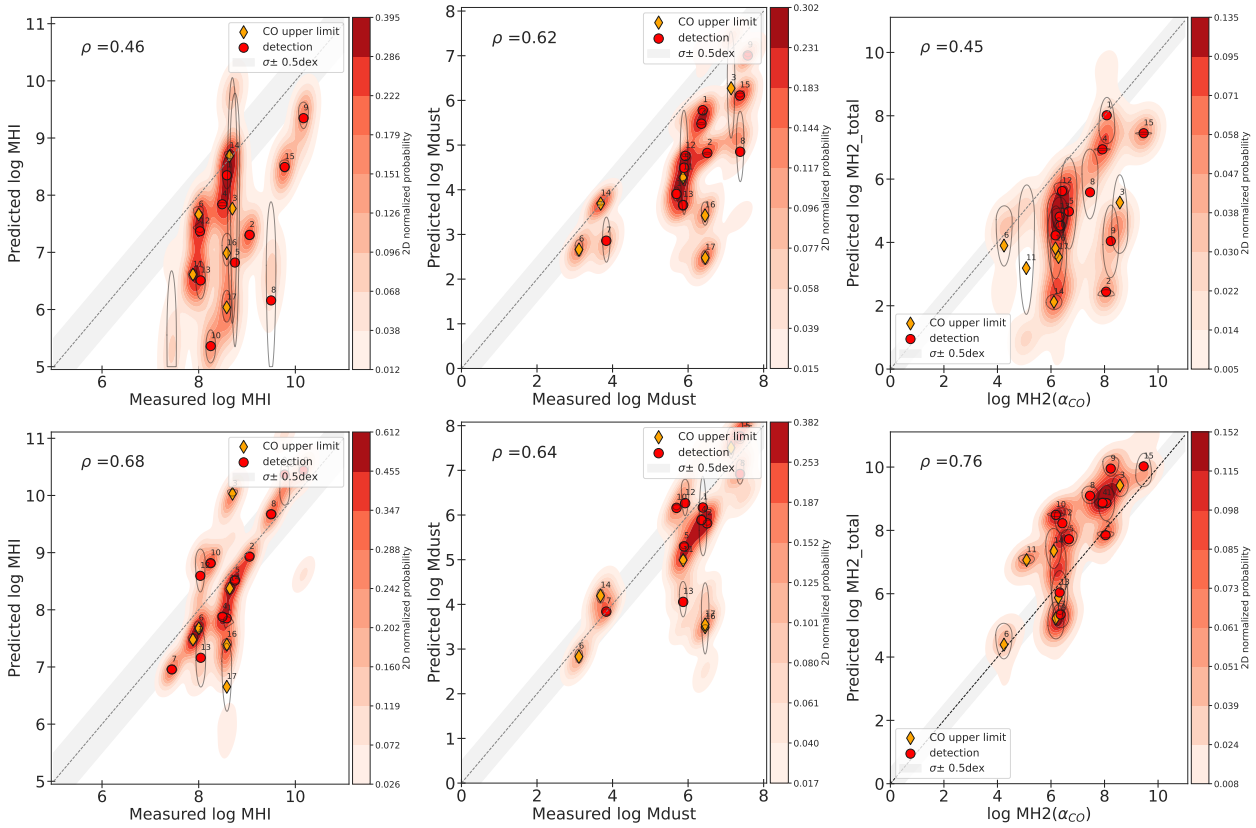

Figure 10.3: Comparison of predictions for single-sector models using either H_2 or CO as constraint. **Top row:** Comparison of the predicted M(H I), M(dust), and M(H_2) obtained when H_2 is used as constraint, with values derived from observation. For M(H_2), we plot the value obtained from CO lines assuming a Galactic α_{CO} conversion factor (Equation 10.3), which provides a lower limit on the total molecular gas mass. **Bottom row:** Same plots showing this time the predictions when CO is used as constrain instead of H_2 lines.

Sections 10.3.4 and 10.3.5) but disappears when H_2 lines are not used as constraints. As described later in Section 10.3.5, we find that our models which reproduce well CO(1-0) overestimate some H_2 lines. Specifically, the H_2 S(0) and H_2 S(1) are overestimated by up to a factor of a few tens, while H_2 S(2) and H_2 S(3) are globally in agreement with the observations.

The origin of this systematic overestimation of H_2 is yet to be investigated. It may be related to some assumptions regarding H_2 (e.g., H_2 formation pathways, chemical reaction rates) or missing physics in our models (regarding, for example, the cosmic ray rates, X-rays, shocks, dust distribution and temperature). In particular, Bron et al. (2014) have shown that accounting for the dust temperature fluctuations is important to accurately predict H_2 lines in PDR models. In their models with high radiation fields ($G_0 \geq 100$), they find that this effect tends to decrease the predicted H_2 intensities. Nevertheless, they find that this decrease remains somewhat moderate (at most 30%) which is not enough to explain the overestimation we find.

A similar overestimation issue was reported in Hu et al. (2021) which found that, at low-metallicity, H_2 deviates from the steady-state equilibrium, causing steady-state models to significantly overestimate H_2. They find that this overestimation has, however, little impact on CO, which may lead to overestimating the CO-dark fractions and α_{CO} conversion factors when using the steady-state approx-

imation. While this effect may explain the significant overestimation of H_2 lines in our models, this time-dependency is expected, according to Hu et al. (2021), to be somewhat mitigated in low surface density environments such as dwarf galaxies in which the SFR surface density is lower.

Including H_2 lines as constraints also leads to predicted H I masses and dust masses which are significantly below observed values (see Figure 10.3, top row). This issue was already pointed out for H I masses, which were systematically underpredicted, in the discussion from Ramambason et al. (2022) (Section 9.3). Similarly, we also find that the predicted H_2 masses are below the masses derived from CO when assuming a Galactic conversion factor, although this presumably provides a lower limit on the actual molecular masses at low-metallicity. We find that relaxing the constraints on H_2 yields H I and dust masses, which are globally in much better agreement with the observed values (see Figure 10.3, bottom row). It also yields, as expected, H_2 gas masses larger than the one obtained with Galactic α_{CO}, which will be further examined in Section 10.3.3.

While accounting for H_2 strongly biases the mass estimates, it has only little impact on the derived values for $f_{esc,HII}$ or f[C II] in the study presented in Chapter 9. We checked that removing H_2 lines does not affect the results that were presented in Chapter 9, which is the case because the quantities we derived were mostly sensitive to the conditions in the ionized gas. Accounting or not for H_2 lines is, however, a problem in the study of the neutral gas. In the following, we remove any constraint on H_2 lines and focus on reproducing CO(1-0).

10.3.3 Single-sector models

We adopt the same modeling strategy as in Chapter 9. The emission lines used to constrain the models are the same as the ones used to constrain the models of the DGS galaxies in Section 9.3, with the addition of CO(1-0) line and removal of H_2. First, we tried to reproduce CO using single-sector models. In such models, the molecular gas is assumed to be uniformly distributed in 4π steradians around the cluster. Those models are somewhat analogs to the ones used in the analysis from M20. In particular, we assumed the same density profile in the PDR and molecular clouds as what was used in M20 (see Section 5.2.3). Nevertheless, as discussed in Section 10.2, the way in which lines are accounted for in our analysis is different. In particular, we perform a simultaneous fitting of all lines, including CO(1-0).

The first column from Figure 10.4 shows the comparison between observations and model predictions for some interesting lines emitted throughout the PDR ([C II] and [O I]) and the molecular phase (CO) for single-sector models. Because we are using a single component to fit all the lines, the code is unable to find a solution that reproduces all the lines well. While [C II] and [O I] lines are almost always well reproduced, the CO(1-0) is well reproduced (within 0.5 dex) only in 4 out of 11 detected galaxies, but is significantly underpredicted in the other 7 galaxies, sometimes by several orders of magnitudes.

This underprediction is not surprising considering that `MULTIGRIS` favors solutions that reproduce most of the lines used as constraints. In our case, since we used mostly lines arising from the H II region and PDR, the code favors solutions corresponding to the physical conditions in those phases. Specifically, the solutions for single-sector models mostly stop at small A_V, which are not sufficient to produce enough CO emission to match the observed values.

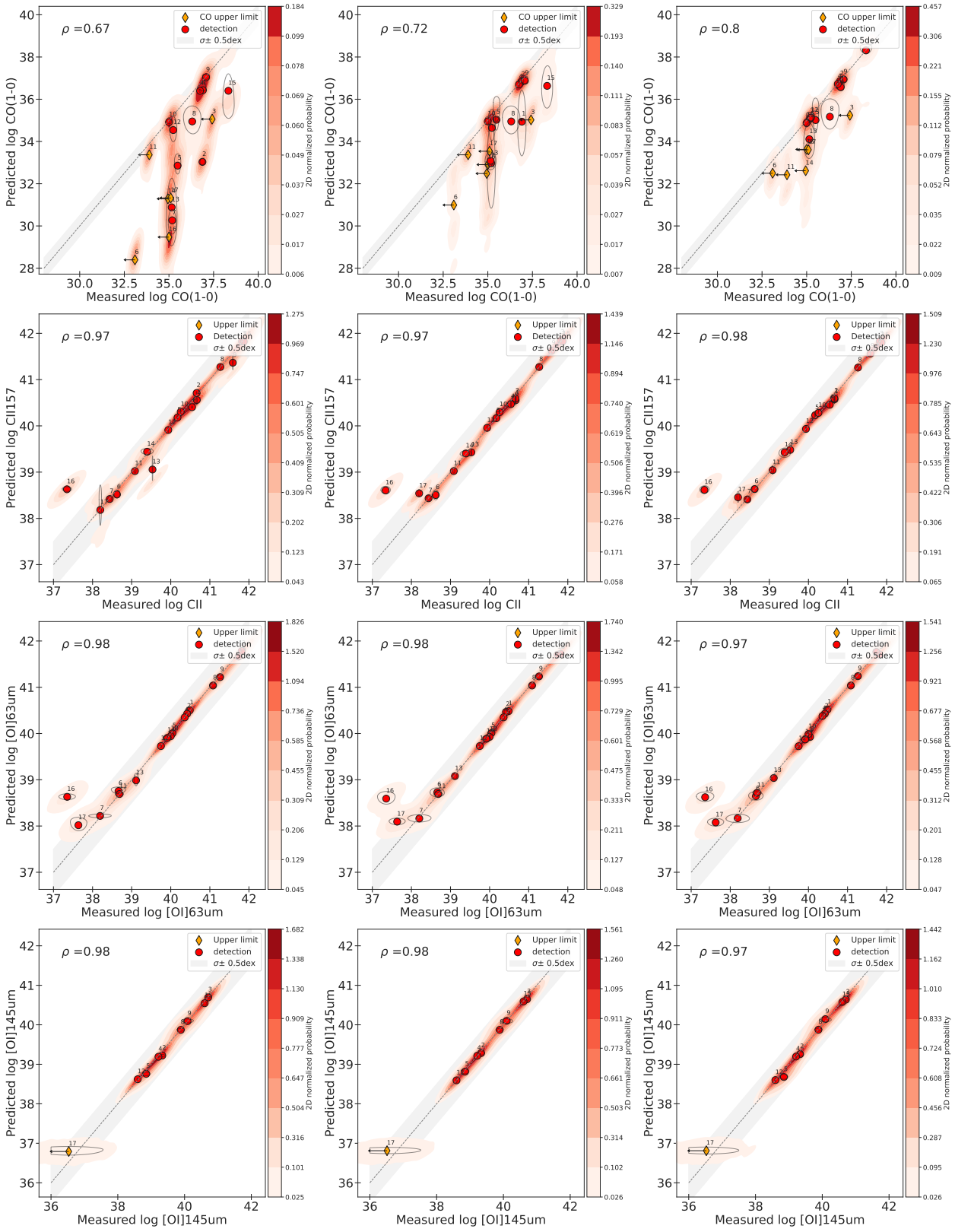

Figure 10.4: Predicted vs. observed lines for single-sector model (left column), multisector models (middle column), and broken power-law models (right column).

While in the single-sector models from M20, A_V was fixed a posteriori to reproduce the CO luminosity, we find that this agreement cannot be obtained when fitting all lines simultaneously. In other words, the CO detection is never strong enough to force a model solution to reach large enough

depths. We hence examine more complex topologies in order to reproduce CO.

10.3.4 Multisector models

A tailored model for I Zw 18

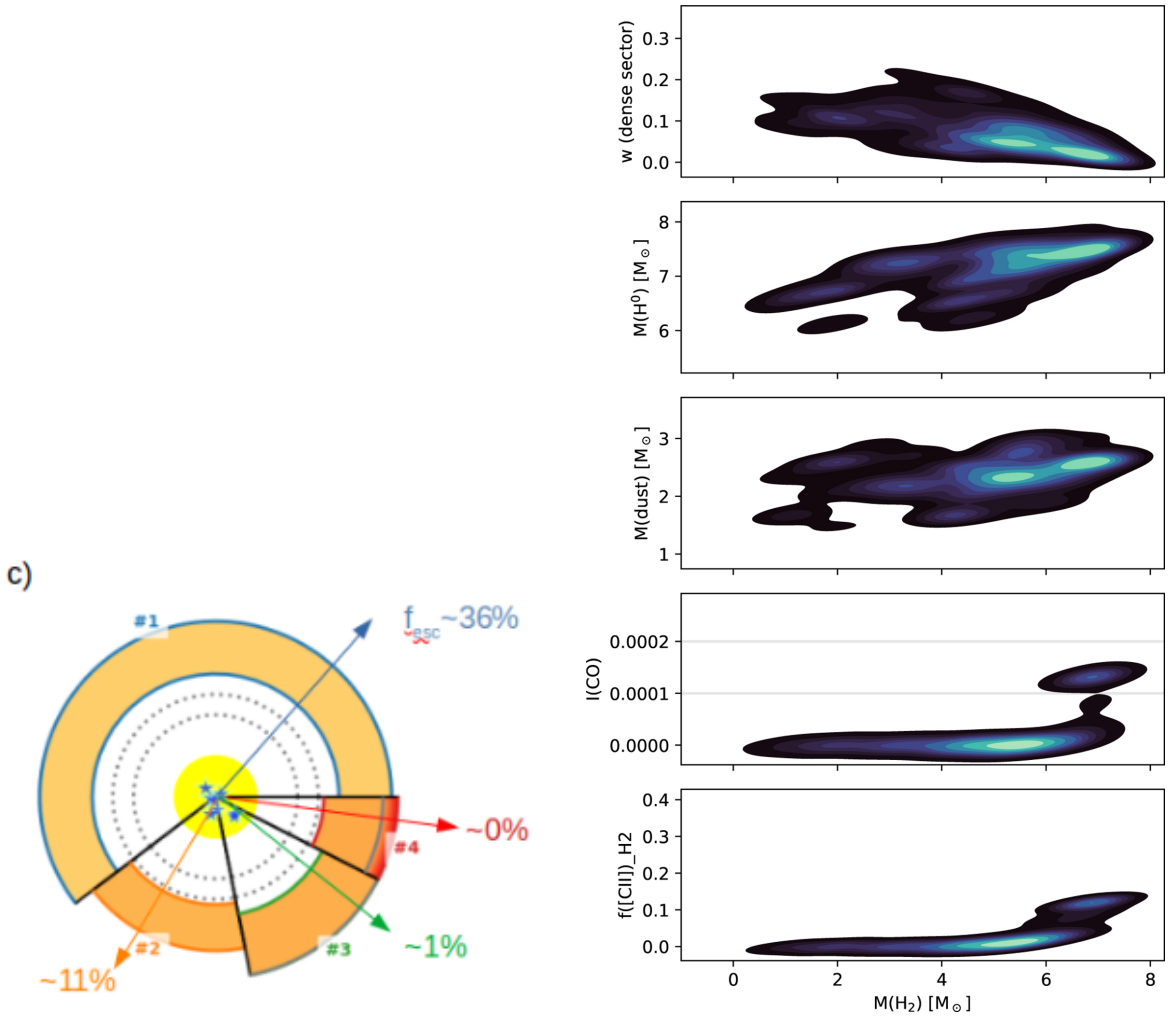

Figure 10.5: **Left:** 4-sector representative topology associated with the best solution from LR22 for I Zw 18. **Right:** Properties of the molecular sector #4. From top to bottom we show as a function of the total predicted H_2 mass, the mixing weight of the molecular sector, the mass of H^0, of dust, the CO(1-0) intensity (scaled to 4.56×10^{36} erg s^1; with the horizontal lines showing the observed range in Zhou et al. 2021), and the fraction of [C II] originating from the molecular sector. Taken from LR22.

The interest of using multisector models to reproduce the emission lines, including CO emission, was first demonstrated in LR22 with the study of the galaxy I Zw 18. We used a 4-sector model to reproduce an extensive suite of optical+IR emission lines, including a claimed CO detection (Zhou et al. 2021). In this analysis, we find that a deep and dense sector with a relatively small mixing weight was needed to reproduce the observed CO emission (see Figure 10.5, left-hand side panel).

The solution found assuming a 4-sector model and accounting for the suite of optical and emission

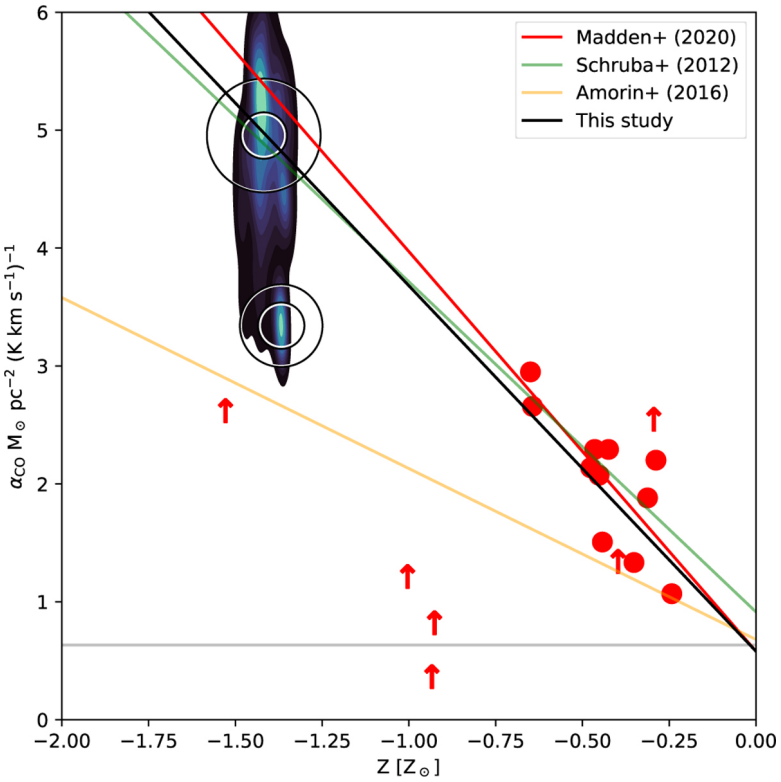

Figure 10.6: Variation of α_{CO} as a function of metallicity compared to Schruba et al. (2012); Sandstrom et al. (2013); Amorín et al. (2016); M20. The red data points and upper limits are from M20. We also show on top the most likely solution (empty black circle with the gray circle indicating the standard deviation) with M(H_2) = $10^{5.4}$ $M_\odot$ and CO 3500 times fainter than in Zhou et al. (2021) and on bottom the secondary, less likely, solution with 10^7 $M_\odot$ and CO in agreement with Zhou et al. (2021). Taken from LR22.

lines is multi-modal. Specifically, we find that the most likely solution predicts a CO value much lower than the observed one from Zhou et al. (2021) (see Figure 10.5, right-hand side panel). This solution is associated with a relatively small H_2 mass of log M(H_2) = $5.39\pm^{+1.67}_{-4.17}$ $M_\odot$ which is almost completely CO-dark. This solution is associated with an α_{CO}= 9×10^4 $M_\odot$ pc^{-2} (K km s^{-1})$^{-1}$ and to a metallicity coefficient of $y = -3.1$ with $\alpha_{CO} \propto (Z/Z_\odot)^y$, in agreement with M20 and Schruba et al. (2012) (see Figure 10.6).

We note, however, that I Zw 18 is an extreme object, with the lowest metallicity in the DGS sample. We now wish to perform the same kind of analysis for the whole DGS sample, including for galaxies in which CO detection are less uncertain. Although the analysis that will be presented in the next sections provides results for I Zw 18, they do not correspond to the same set of lines as the one used in LR22, which included optical lines. Additionally, in the following section, we do not use the CO detection from Zhou et al. (2021) but the upper limit from Leroy et al. (2005) for I Zw 18.

Multisector models for the DGS

We now examine the results obtained with multisector topological models for the whole DGS sample. This approach is similar to the one presented in Chapter 9 in which we varied the number of sectors and selected the configuration corresponding to the highest marginal likelihood. While we used up

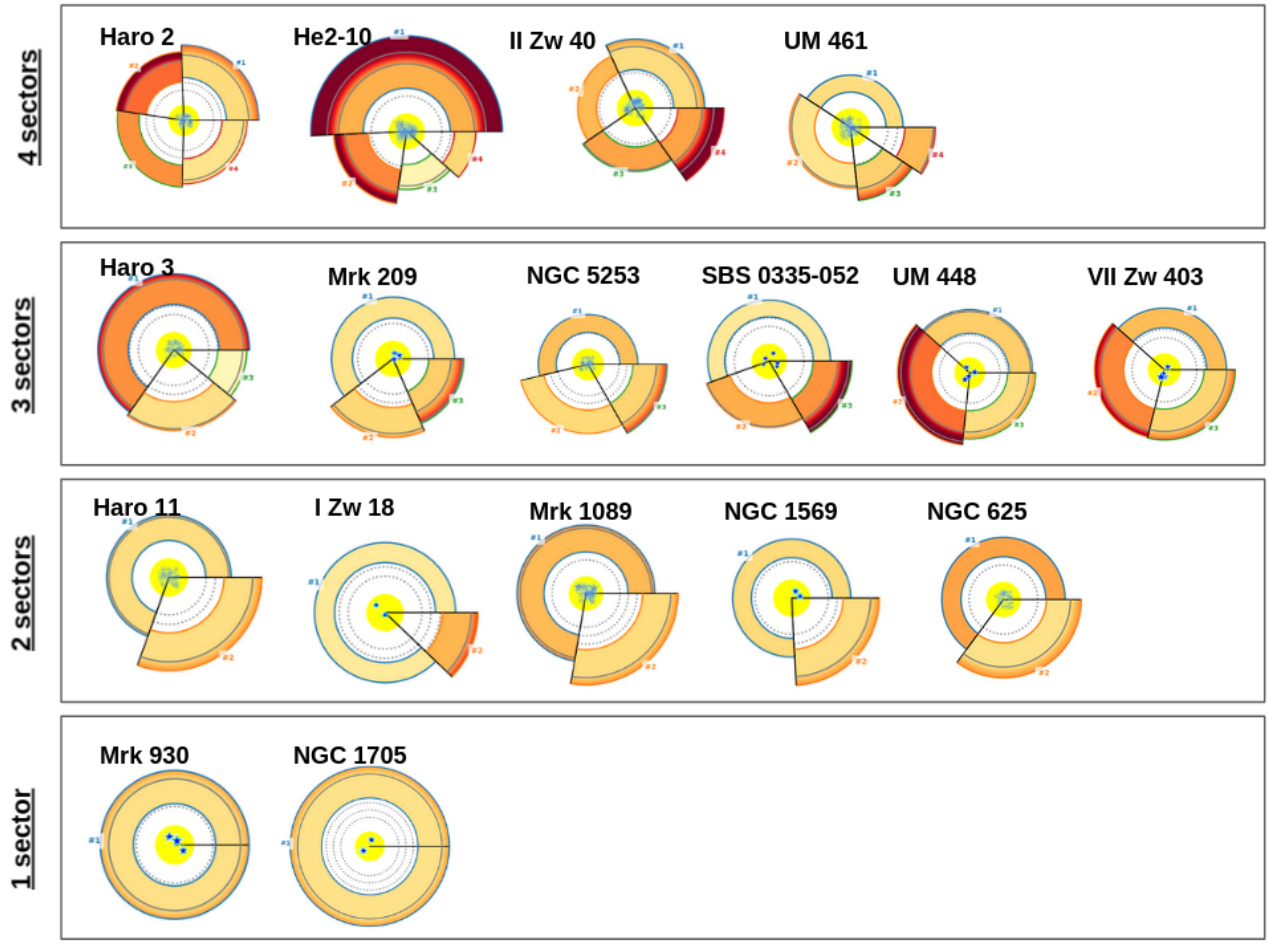

Figure 10.7: Best solutions among multisector models for the DGS galaxies (see Figure 4.1 for the plot description).

to 3 sectors in the analysis focusing on f_{esc}(see Section 9.3), we now allow the number of sectors to vary up to 4. This is motivated by the fact that, in galaxies which were best fitted by 3 sectors without including CO(1-0), an additional sector may be necessary to match the observed CO value.

The best configuration obtained for each galaxy based on the marginal likelihood criterion (see Section 7.5.1) is represented in Figure 10.7. Although the code considers up to 4 sectors, the solution with the highest marginal likelihood has only 1 sector for 2 galaxies, and only 2 sectors for 5 galaxies. This can be explained by the lack of information provided by the set of line used as input, which do not allow constraining more sectors. For example, our best configuration for I Zw 18 has only two sectors for the set of IR lines we use. This highlights the dependency of the best configuration chosen by `MULTIGRIS` on the suite of lines used inputs (see discussion in LR22 and in Section 8.1.3).

In the middle column from Figure 10.4, we show the predicted vs. observed values for [C II], [O I], and CO(1-0). As in the single-sector case, [C II] and [O I] are well reproduced by the models. Considering multisector models allows us to predict CO(1-0) values in slightly better agreement with observations than in the single-sector case. Nevertheless, CO(1-0) remains significantly underpredicted for 5 out of 11 detected galaxies, although the difference is smaller than for single-sector models. Interestingly, forcing the number of sectors (e.g., forcing 4 sectors for all galaxies) does not help in reproducing CO values in all galaxies. We conclude that even more complex topologies

which combine discrete components as sectors may not be complex enough to match the observed CO emission.

10.3.5 Broken power-law distribution

Table 10.2: Distribution of parameters adopted in the broken power-law configuration.

Random variables	Prior
Density distribution	$P(n) \propto n^{\alpha_n}$; $[n_{\rm min}, n_{\rm max}]$
α_n	$\mathcal{N}(\mu$=-1.5, σ=2)
$n_{\rm min}$	$\mathcal{N}(\mu$=0, σ=1)
$n_{\rm max}$	$\mathcal{N}(\mu$=4, σ=1)
Ionization parameter distribution	$P(U) \propto U^{\alpha_U}$; $[U_{\rm min}, U_{\rm max}]$
α_U	$\mathcal{N}(\mu$=-1.5, σ=2)
$U_{\rm min}$	$\mathcal{N}(\mu$=-4, σ=1)
$U_{\rm max}$	$\mathcal{N}(\mu$=0, σ=1)
Cut distribution	$P^{(1)}(\text{cut}) \propto \begin{cases} \text{cut}^{\alpha_{\rm cut,1}}; [\text{cut}_{\rm min}, \text{cut}_{\rm pivot}] \\ \text{cut}^{\alpha_{\rm cut,2}}; [\text{cut}_{\rm pivot}, \text{cut}_{\rm max}] \end{cases}$
$\alpha_{\rm cut,1}$	$\mathcal{N}^{(2)}(\mu$=0, σ=2)
$\alpha_{\rm cut,2}$	$\mathcal{N}(\mu$=0, σ=2)
$\text{cut}_{\rm min}$	$\mathcal{N}(\mu$=0, σ=1)
$\text{cut}_{\rm max}$	$\mathcal{N}(\mu$=2.25, σ=1)
$\text{cut}_{\rm pivot}$	fixed at the ionization front (cut=1)

(1) $P(X)$ is the probability distribution function assumed for parameter X.
(2) $\mathcal{N}(\mu, \sigma)$ is a Normal distribution with mean μ and full-width half maximum σ.

We have seen in Sections 10.3.3 and 10.3.4 that single-sector models and multisector models poorly reproduce the observed CO(1-0) emission. In both cases, we assumed that CO can be produced by a single sector provided it is deep enough.

While in multisector models the CO-emitting sector is allowed to span physical conditions that are very different from the sectors associated with the emission of H II region and PDR lines (e.g., denser, stopping at larger depth), we saw that this additional degree of freedom is not sufficient to reproduce CO in all cases. In fact, the combination of discrete sectors, even numerous, may not reproduce well the actual distribution of CO-emitting clouds. Another approach based on the use of power-law distributions to describe the density and the ionization parameter was introduced for studies of optical lines in SDSS galaxies with the LOC models Richardson et al. (2014, 2016), as described in Section 4.4. We here extend this method by using a broken power-law distribution for the cut parameter (see Section 5.2.3 for the cut definition). The broken power-law distribution consists of two joint power-law distributions with a change of slope at a given pivot point, which we fix at the ionization front. This approach was first tested with an exploratory model of I Zw 18 in LR22.

Our new topological configuration relies on the use of power-law and broken power-law distributions to described three parameters: the density, the ionization parameter, and the cut. The predictions

associated with these models result from a combination of many components with different values, within a given range, where the mixing weights are controlled by the slope α of the power-law distribution. This is, in essence, an extension of the multisector model, in which we combine a very large number of components following a continuous distribution. The priors adopted for the different parameters are described in Table 10.2.

In the right hand-side column from Figure 10.4, we show the agreement between predicted fluxes and observed values for [C II], [O I], and CO(1-0). As opposed to what was observed for single- and multisector models, we see that adopting a broken-power law configuration allows us to reproduce CO(1-0) luminosities in agreements with detections and upper limits for all the galaxies. These broken power-law models also succeed in reproducing well the other emission lines (from H II region and PDR) although with somewhat lower global marginal likelihoods than the multisector models from Chapter 9. The ability of power-law models to reproduce CO, when single- and multisector could not, is due to the assumption of a distribution of clouds which allows for the contribution of fewer and deeper CO-emitting clouds. We now further examine the gas geometry in the different configurations that were presented in this section.

10.4 Geometries of the models

10.4.1 A_V dependencies in single and multisector models

In Figure 10.8, we show the variation of A_V with various quantities for single-sector and multisector models. In the left-hand side column, we show the variation of the A_V of the CO-emitting sector (i.e., either the only sector in single-sector models or the deepest sector in multisector models) with metallicity. In both cases (single and multisector models), we find that A_V correlates well with metallicity. This correlation is a direct consequence of the dust-to-gas mass ratio that we adopted in the grid, which scales with metallicity (see Section 5.2.2).

In the middle column from Figure 10.8, we show the variations of the [C II]/CO(1-0) versus the A_V associated with the CO-emitting sector. Interestingly, the situation changes dramatically between single and multisector models. In single-sector models (top row), we find that the A_V at which models stop is anti-correlated with [C II]/CO(1-0) ratio. This result is in line with what was found in M20 using a similar topology. We note, however, that because our single-sector models sometimes drastically underestimate CO(1-0), our predicted [C II]/CO(1-0) reaches values up to 10^{10}, while the observed ratio is at most of a few 10^5. Nevertheless, the anti-correlation still holds for those high values of [C II]/CO(1-0) ratios and corresponds to small predicted A_V. The A_V determined for single-sector models should thus be taken with caution because it is entirely controlled by the [C II]/CO ratio.

On the other-hand, when several sectors are present, the A_V corresponding to the CO-emitting sector is not anymore correlated with the [C II]/CO ratio. Indeed, in multisector models, this ratio is sensitive to the relative contribution of all sectors to [C II] and CO emission, which do not necessarily arise from the same component. This means that in such models, [C II]/CO depends on the inferred topology and specifically on the mixing weight of each sector. Hence, accounting for the presence of different gas components with different physical conditions (density, depths, covering angle...)

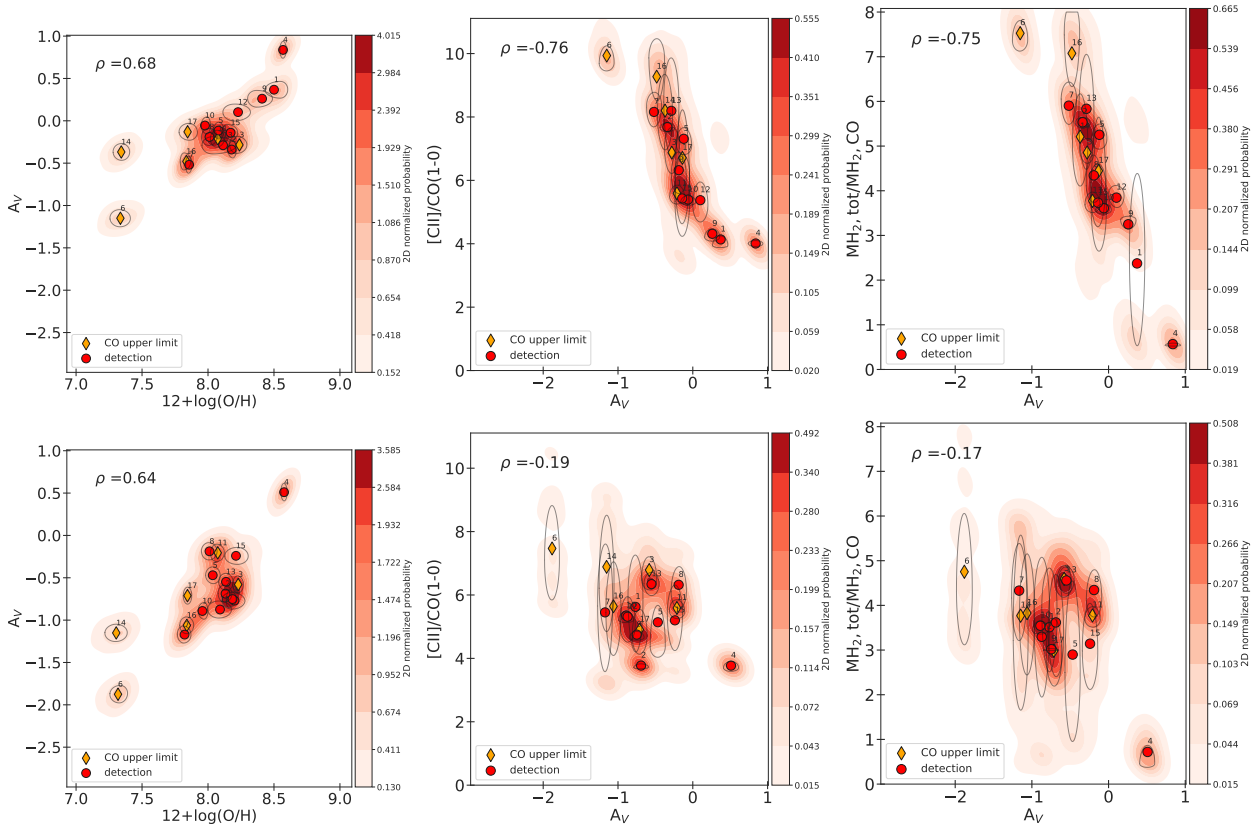

Figure 10.8: A_V dependencies in single-sector (top row) and multisector (bottom row) models. **Left:** [C II]158μm/CO(1-0) vs. A_V. **Middle:** Ratio of the total H_2 mass and CO-bright H_2 vs. A_V. **Right**: Variation of A_V as a function of metallicity.

complicates the interpretation of the A_V of the CO-emitting sector, which cannot be easily constrained by the [C II]/CO ratio only.

In the right-hand side column from Figure 10.8, we show the variation of the total H_2 mass over the CO-bright H_2 mass versus A_V. In the single-sector case (top row), we find results similar to what was reported in M20: this mass ratio anti-correlates with A_V, simply reflecting the fact that deeper models reaching larger A_V produce more CO-bright molecular gas. Again, our total-to-CO-bright H_2 mass ratios are much larger than the ones found in M20 because our models contains less CO-bright H_2 since CO is not well reproduced. Because A_V correlates with metallicity (see the right-hand side column from Figure 10.8), the anti-correlation of the [C II]/CO(1-0) and $MH_{2,total}/MH_{2,CO}$ with A_V also corresponds to an anti-correlation with metallicity.

However, we find that the anti-correlation of $M(H_2)_{tot}/M(H_2)_{CO}$ with A_V becomes less strong and comes with a much larger scatter when we consider combinations of sectors. The idealized case of single sector model, stopping at a single A_V value, controlled by the [C II]/CO(1-0) ratio, and single-handedly determining the $M(H_2)_{tot}/M(H_2)_{CO}$ is an oversimplification that may not correspond to the real A_V distribution of CO-emitting clouds. While multisector models are a first step towards a more realistic distribution of the CO-emitting clouds, we have seen in Section 10.3 that such topologies are not always complex enough to reproduce the CO emission. We now further detail the geometry which is considered when adopting power-law and broken power-law distributions of parameters.

10.4.2 PDFs of the power law parameters

Prior PDFs

For the broken power-law models, we cannot properly define A_V as we did for single- and multiple-sector models (see Figures 10.8). Instead, we examine the PDFs of the parameters that control the distributions of density, ionization parameter, and cut. Those distributions consist of power-laws (defined by a slope over a given range of values) and of a broken power-law (defined by two slopes over two ranges of values separated by a pivot points). The prior distributions set for those parameters are described in Table 10.2. They consist of relatively weak priors defined as normal distributions centered on a value μ with a relatively large σ.

Posterior PDFs

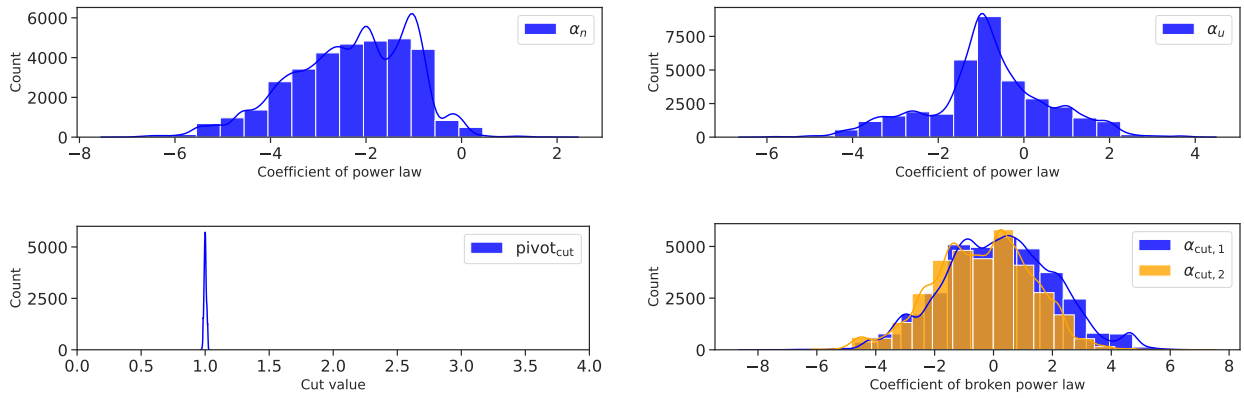

Figure 10.9: Posterior PDFs of the power-law parameters.

We note that the posterior PDFs for all galaxies are rather wide, as shown in Figure 10.9. The posterior PDFs for α_n and α_U are relatively close to the prior PDFs but show a peak respectively around -3 and -1. The pivot cut value for the broken power-law has a maximum probability at cut=1, corresponding to the location of the ionization front. It is worth noting that the posterior PDF is consistent with the prior that was set, while it could have been flat or shifted to other values, because of the relatively weak prior that we used.

As introduced in Section 4.4, the interest of using a power-law distribution lies in the possibility to account for different frequencies of physical parameters in the population of molecular that dominates the emission. Hence, the negative slopes found for U and n indicate that the emission is dominated by relatively low density gas with moderate ionization parameters. We note that, the power-laws control the distributions of densities and ionization parameter at the illuminated face of the cloud and that the distributions may be different if averaged over the volume. In Figure 10.10, we show the slopes and boundaries obtained for the density and ionization parameters for individual galaxies. The mean density at the illuminated front is varies 5 to 100 cm^{-3} and the ionization parameter log U between -3 to -1.

Interestingly, the α_{cut} parameters have clearly distinct values, with the PDF of $\alpha_{\mathrm{cut,2}}$ being shifted to lower values than $\alpha_{\mathrm{cut,1}}$. A negative α value indicates that clouds with relatively smaller depths are

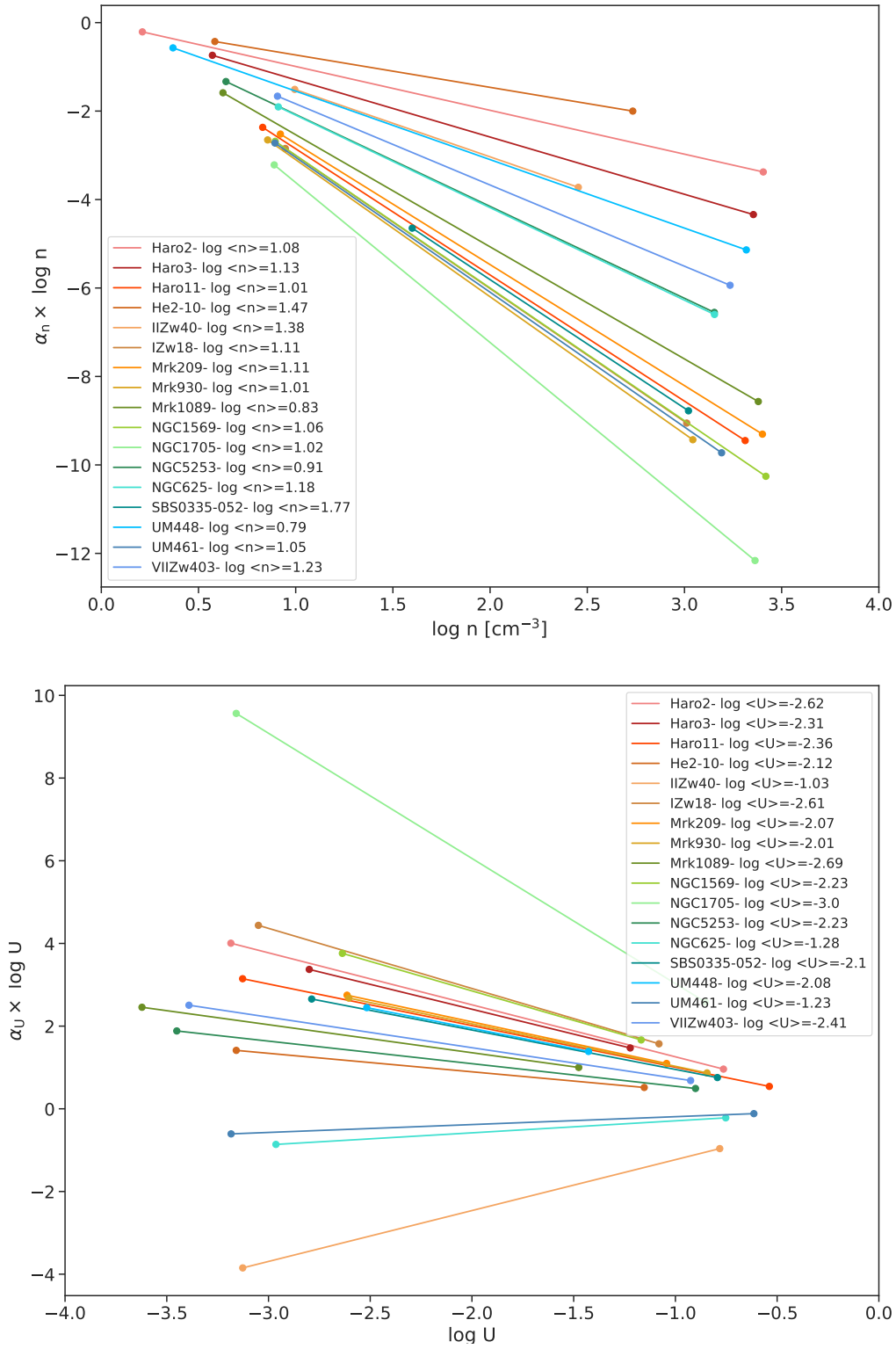

Figure 10.10: Values of the power law parameters controlling the density and ionization parameters.

more numerous, while we find fewer clumps with large depths. The exact values, however, differ for each galaxy and also depend on the pivot point. These distinct PDFs obtained for the $\alpha_{\rm cut}$ parameters reflect differences in terms of slopes of the power-law before and after the IF, as shown in Figure 10.11.

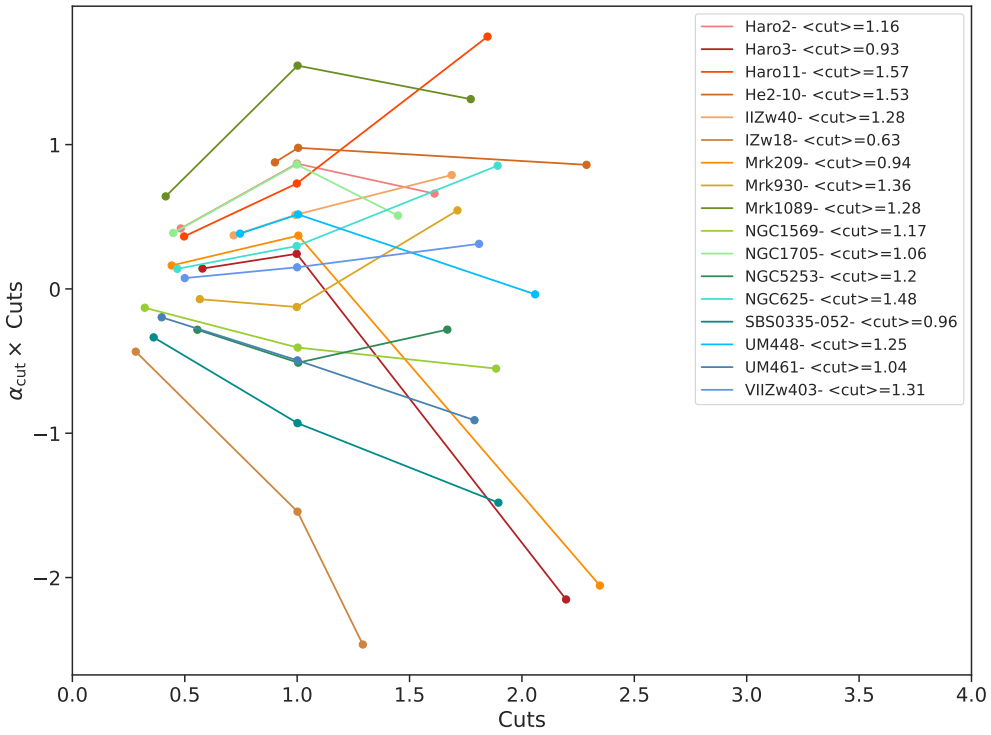

Figure 10.11: Values of the broken power law parameters controlling the cut.

10.5 Total H_2 mass, CO-dark gas fractions, and α_{CO} estimates

10.5.1 Total H_2 mass

In the left-hand side panel from Figure 10.12, we show the predicted total H_2 mass vs. the [C II] 158μm luminosity for all models (single- and multisector models as well as models with broken power-law distributions). For all models, we find values in globally good agreement with the ones found in M20, except for the single-sector models (top row) for which our predictions for the faintest [C II] values are slightly too low. Those galaxies correspond to models in which CO(1-0) is drastically underpredicted. Nevertheless, it is interesting to note that, although our single-sector models predict significantly smaller CO-bright H_2 masses than the ones from M20, the total H_2 masses remain in globally good agreement. As discussed in Section 10.6, all the configurations predict consistent total M(H_2), in good agreement with M20, regardless of whether they reproduce CO emission, which a direct consequence of the molecular being mostly CO-dark.

10.5.2 CO-dark gas fractions

In the middle panel from Figure 10.12, we show the total-to-CO-bright H_2 ratio vs. [C II]/CO(1-0). Similarly to M20, we find a tight correlation between those two quantities for the single-sector models (top row). We find a somewhat flatter slope than the one from M20 which is caused by the very small predicted CO(1-0) luminosity, and small CO-bright H_2 masses in our models. The correlation between [C II]/CO(1-0) and the total-to-CO-bright H_2 ratio still holds for multisector models (middle row), but

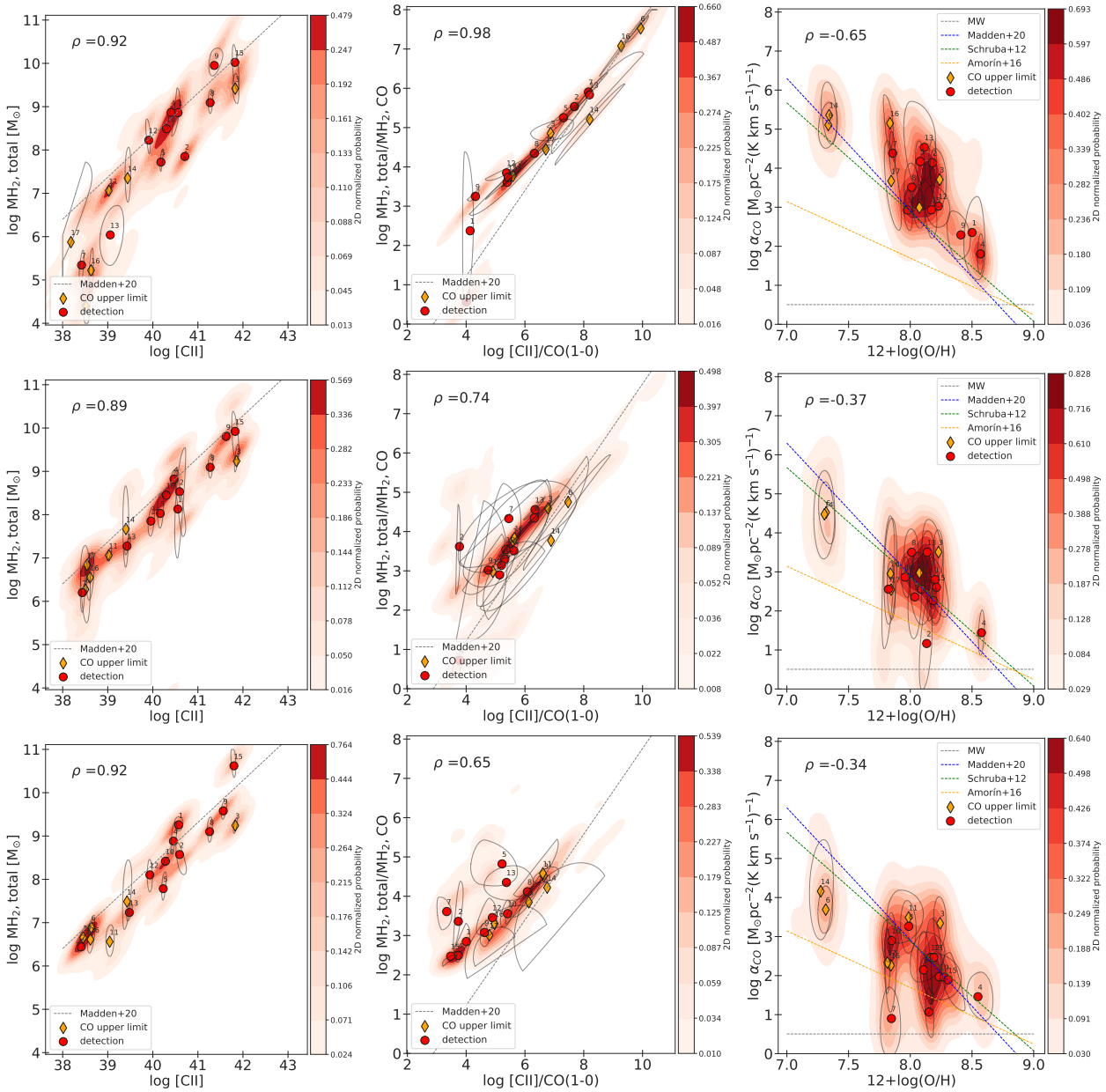

Figure 10.12: H_2 masses, CO dark fractions, and α_{CO} estimates for single-sector (top row), multisector (middle row), and power-law models (bottom row). **Left column:** [C II] vs. $MH_{2,total}$. **Middle column:** $MH_{2,total}/MH_{2,CO}$ vs. [C II]/CO(1-0). **Right column:** α_{CO} vs. metallicity.

with a larger scatter. The range of [C II]/CO(1-0) values predicted by multisector models is in better agreement with the observed range, although still shifted to larger values. Most of the predicted values for the total-to-CO-bright H_2 ratio lie above the relation from M20 because multisector models are still producing less CO-bright H_2 gas than the M20 estimates (based on Galactic α_{CO} conversion factors).

Interestingly, even for power-law models (bottom row) that successfully reproduce the CO(1-0) emission, the CO-bright H_2 mass estimates remain much smaller than those of M20, leading to higher $M(H_2)_{tot}/M(H_2)_{CO}$. Those small CO-bright masses are linked to the power-law distribution that we adopted, in which a few dense clumps can produce enough CO(1-0) emission to match the observational constraints. Nevertheless, in most cases the index of the power-law after the IF is

negative, indicating that only a few clouds with large A_V are required. Hence, power-law models are associated with smaller $M(H_2)_{CO}$ than single or multisector models. While the total mass of molecular gas is roughly constant, the predicted CO-bright molecular gas decreases when going from single-sector, to multisector, and to power-law models.

10.5.3 α_{CO} conversion factor

In this section, we examine how our results on the α_{CO} conversion factor compare to the results from M20. While other studies could also be taken as reference, the possible agreement or discrepancy with our results is less easy to interpret as they were obtained with various methods on different galaxy samples. On the other hand, the main difference between M20 and the current study is that we accounted for a more complex, multicomponent geometry while they assumed single-sector models. The sample of galaxy is, however, the same, and the prescription used to build the Cloudy grids used in both analysis are fairly similar, although not exactly.

In the right-hand side panel from Figure 10.12, we show the α_{CO} conversion factor from the model predictions vs. metallicity. The α_{CO} values we derive for single-sector models (top row) are globally larger than the relation from M20. This is due to the fact that our models are associated with similar H_2 masses as the M20 models but produce much less CO(1-0) emission. We stress that those high α_{CO} values do not reflect an actual physical trend but only reflect the inadequacy of single sector models to reproduce well the observed CO(1-0) values.

The α_{CO} values derived for multisector models (middle row) are in better agreement with the relation from M20 than single-sector models, although the scatter is very large. Specifically, for models in which CO is underpredicted, we still find large α_{CO} values, above the relation from M20. However, it is interesting to note that one galaxy (#2: Haro 3) seems to be relatively close to the Galactic α_{CO} value, despite its low-metallicity.

Finally, we show on the bottom row the α_{CO} predicted for the power-law models, which are the only ones to successfully reproduce CO in all galaxies (see Section 10.3). We find that our predictions globally lie in between the Galactic α_{CO} and the relation from M20. Only one galaxy (#3: Haro 11), for which the only CO constraint is an upper limit, has a predicted α_{CO} significantly above the M20 relation. Interestingly, 3 galaxies (#2: Haro 3, #4: He2-10, and #7: UM 461) are predicted to have α_{CO} values compatible with the Galactic value within error bars, although two of them have relatively low metallicities. One of these galaxies (Haro 3) was also predicted to have a nearly Galactic α_{CO} value with multisector models. For those three galaxies, the upper bound in cut is found to be relatively large (> 2), meaning that the models account for relatively deeper clouds, and the $\alpha_{cut,2}$ value is negative and steep. This configuration can be interpreted as CO emission being powered by a few dense clumps, while shallower slopes and lower upper bounds would correspond to a more diffuse H_2 emission.

As discussed in Bolatto et al. (2008) and Bolatto et al. (2013), if CO(1-0) emission is dominated by opaque, bright-CO emitting clouds, no dependency is expected to be found with metallicity. The increase of α_{CO} when metallicity decreases is, however, expected when H_2 comes from large regions in which the offset between CO-emitting surface and H_2 location becomes important.

10.6 Summary

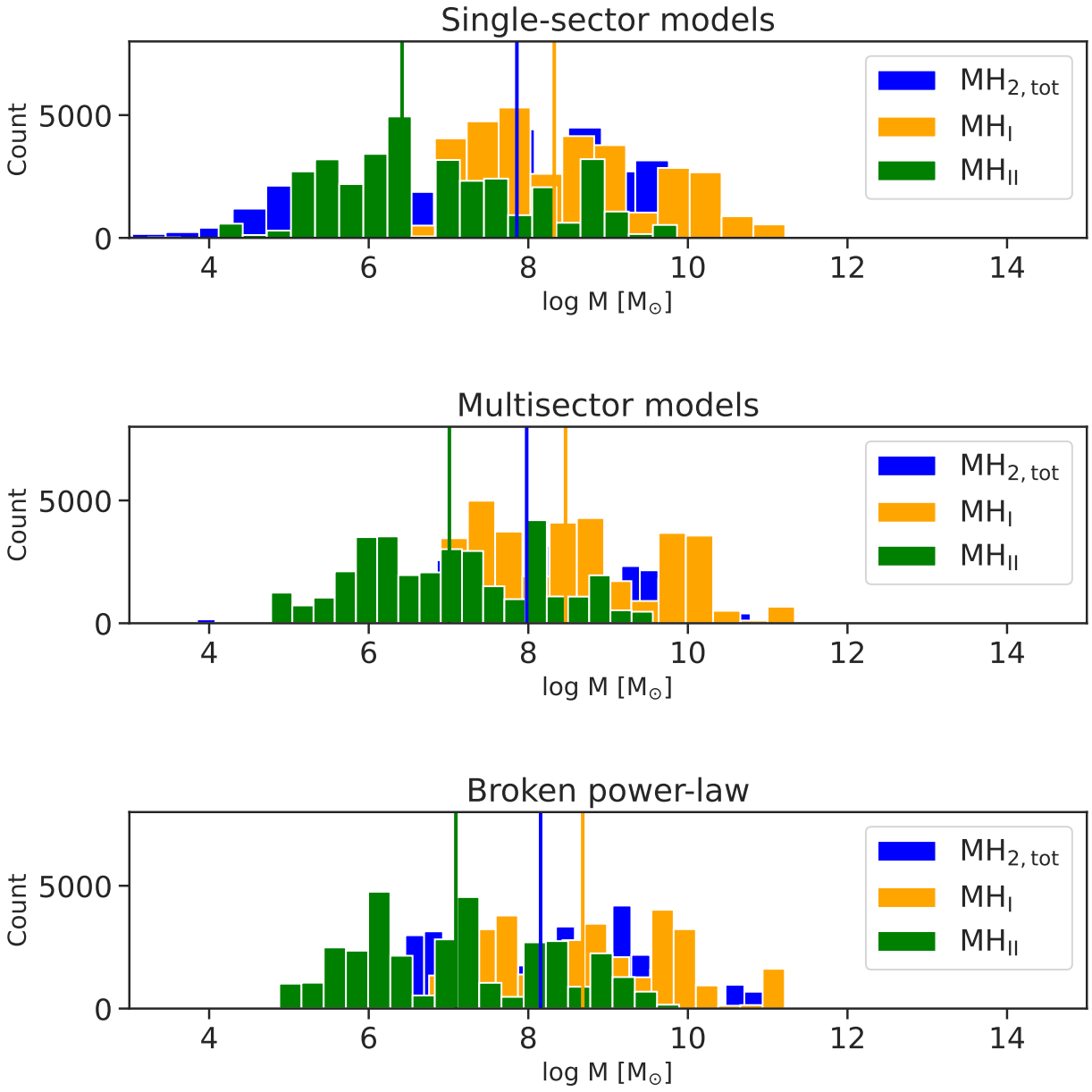

Figure 10.13: Comparison of the H II,H I, and H_2 masses inferred for the full DGS sample, assuming different topologies.

In Figures 10.13 we compare the different configurations that we considered throughout this chapter. Globally, we show that the median values of predicted H II and H I masses are not changing much when assuming different configurations. The total H_2 mass, however, is sensitive to changes in the topology: because single-sector models are unable to reproduce CO(1-0) (in a simultaneous fitting process), they predict small H_2 masses. The predicted masses are larger when accounting for the contribution of multiple sectors, although CO(1-0) remains underestimated in some galaxies. Interestingly, power-law models predict H_2 masses with the same median values as multisector models, although they predict higher CO(1-0) values, in agreement with observations. This is due to the bro-

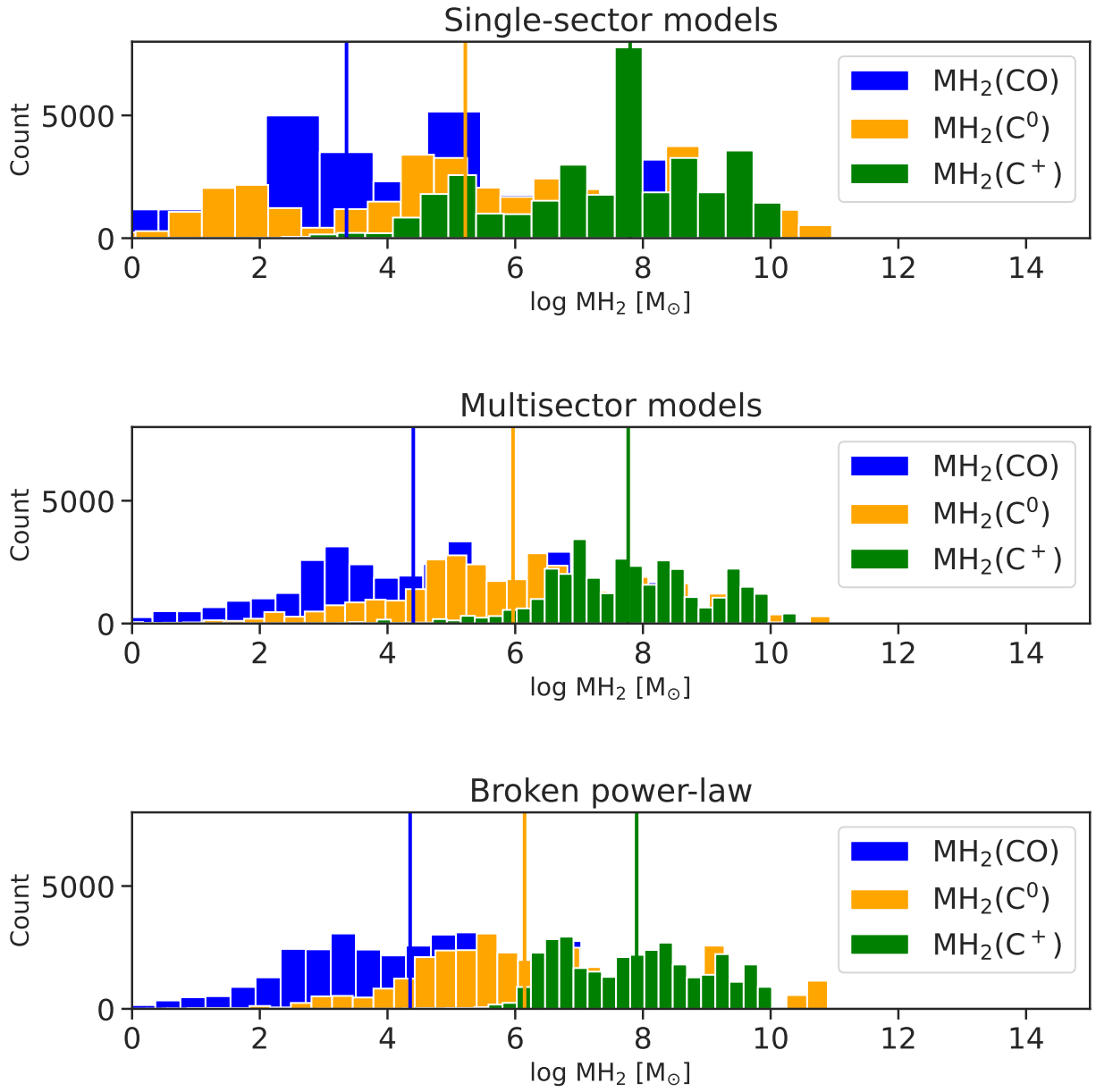

Figure 10.14: Comparison of the H_2 masses associated with C^+, C^0, and CO inferred for the full DGS sample, assuming different topologies.

ken power-law distribution of the cuts which allows reproducing CO(1-0) emission with only a few dense CO-emitting clumps, and to the fact that most H_2 is CO-dark.

In Figures 10.15, we compare the H_2 masses associated with C^+, C^0, and CO, for all the configurations. We find that in all the configurations we considered, the total H_2 mass in always dominated by CO-dark molecular gas, specifically by the component associated with C^+. Subsequently, we confirm that [C II] is a good tracer of molecular gas that tightly correlates with the gas total gas masses in all of our configurations. The relation we derive for $M(H_2)$-[C II] is robust to changes in the topology and follows that of M20 (as shown in the left-hand side column from Figure 10.12).

On the other hand, the predictions for $M(H_2)_{C^0}$ and $M(H_2)_{CO}$ are more sensitive to changes in the

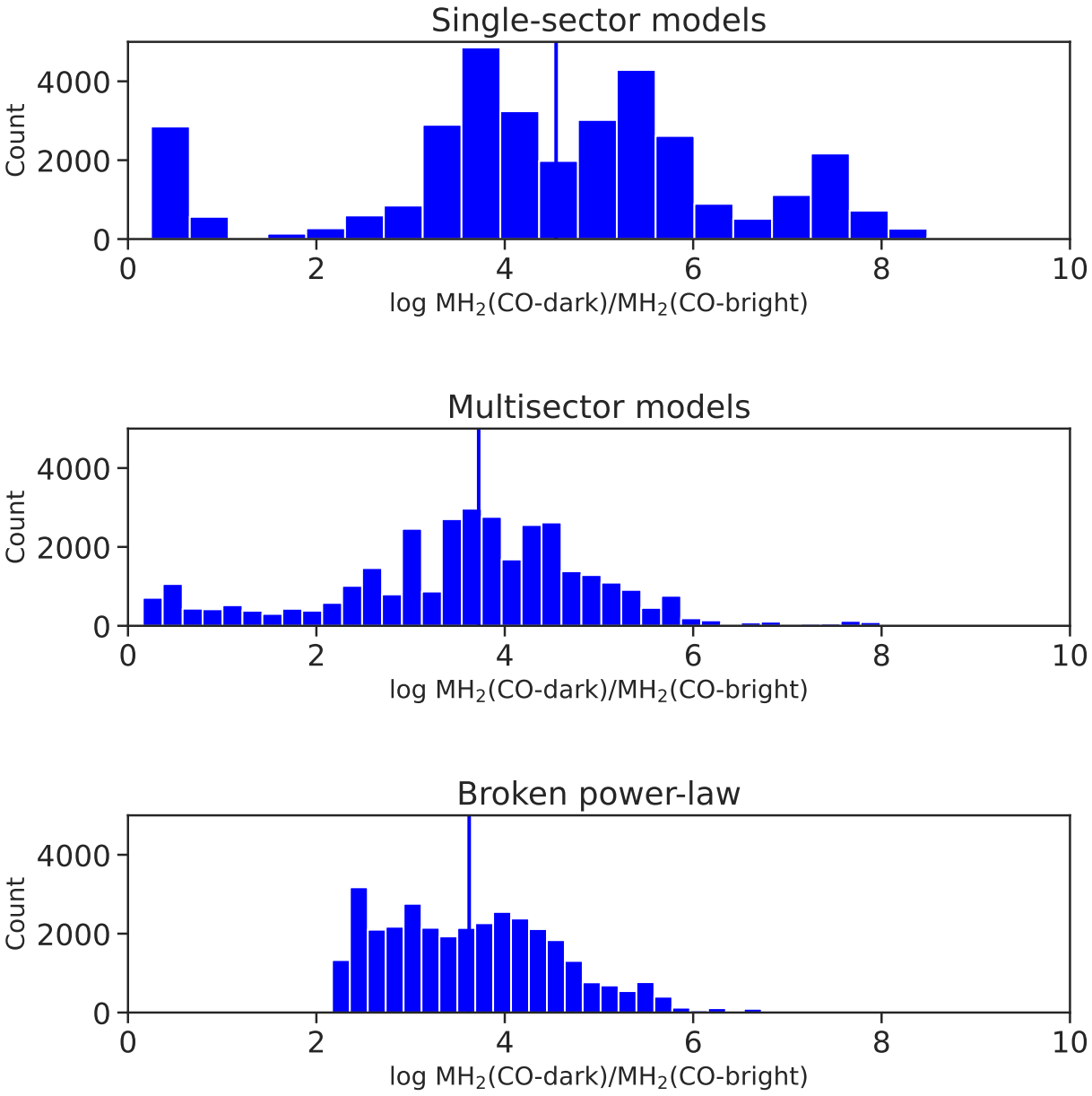

Figure 10.15: Comparison of the CO-dark to CO-bright H_2 mass ratios for the full DGS sample, assuming different topologies.

topology, although the median values obtained for both multisector models and power-law models are strikingly close. As shown in the middle column from Figure 10.12, we find an increasing scatter and globally lower masses of CO-bright H_2 when including high A_V components with small mixing weights. The relation derived for the CO-bright to CO-dark gas mass ratio and the α_{CO} in M20 are driven by simplistic models assumptions. We find that the estimates of CO-dark gas fractions and α_{CO} values include large uncertainties, which reflect different modeling assumptions. More constraints are needed on the gas distribution to properly derive such quantities.

To conclude, the main results from our analysis are the following:

- All our models predict that the molecular gas is the DGS is completely dominated by $M(H_2)_{C^+}$,

which is in average 100 times larger than $M(H_2)_{C^0}$ and 1000-10000 times larger than $M(H_2)_{CO}$. We find total molecular masses in agreement with M20 and, as a consequence, also that the DGS galaxies should lie on the Schmidt-Kennicutt relation.

- We confirm that [C II] is a good tracer of the total molecular gas. In particular, the robustness of the $M(H_2)$-[C II] relation to changes in the model topology, and the little scatter of our predictions suggest that [C II] is a very good proxy to predict H_2 in low-metallicity environments. On the contrary, the interest of CO as a tracer of molecular gas breaks down at low-metallicity (e.g., Bolatto et al. 2013), since its emission is barely associated with any molecular gas mass.

- CO(1-0) can not be easily reproduced in our statistical framework that intend to simultaneously fit all the available emission arising from different phases of the ISM. Both single-sector and multisector configurations are unable to reproduce all the CO detections in our sample. While the multisector representation could successfully describe the ionized and neutral phases analyzed in Chapter 9, we find that they fall short of reproducing emission arising from the denser CO-emitting gas.

- Instead, we find that using a broken-power distribution to control the depth of the clouds that are combined allows us to reproduce the observed CO emission. The emission is hence produced by relatively few dense regions, which contribute little to the total molecular mass. This results in large CO-dark fractions associated with larger value of $MH_{2,tot}/MH_{2,CO}$ than what was predicted by M20.

- The predicted CO emission and associated CO-bright gas, on the other hand, vary significantly with different modeling assumptions. This makes the predictions for the $M(H_2)_{total}/M(H_2)_{CO}$ ratio and for the α_{CO} conversion factor highly sensitive to modeling assumptions.

- We find that deriving a galactic value for α_{CO} is a complex task and that a significant scatter is present in our predictions. We identify two extreme cases corresponding to different distributions of the CO-emitting gas: (1) the CO-emitting gas is distributed homogeneously and can be modeled by a single CO-emitting sector associated with relatively large H_2 masses. This is the case described by single- and multisector models, leading to α_{CO} values close to the relation of M20. (2) CO emission is dominated by a few small clumps with large A_V, at the upper-end of the power-law distribution. In this case, the α_{CO} values are close to the Galactic values. We find that our predictions are spread in between the M20 relation and the Galactic value, with a relatively large scatter. The exact location in between those two relations is likely driven our modeling assumptions regarding the distribution of CO-emitting clouds. More constraints are needed on the CO-distribution if we want to accurately derive α_{CO}.

Part V

Prospectives and conclusion

Chapter 11

Prospective and conclusion

In the current chapter, we discuss possible improvements of the modeling scheme that we presented throughout the thesis. Specifically, technical improvements are possible regarding the models used to create the grid of `Cloudy` models (see Chapter 5) and the statistical tools used to explore the parameter space (see Chapter 7). They are presented in Section 11.1. In Section 11.2.1, we present the possible applications of `MULTIGRIS` to both simulated mock galaxies and to multiwavelength observations of current and future telescopes. Finally, we summarize the work that was presented in this thesis in Section 11.3.

11.1 Technical improvements

11.1.1 Improving `Cloudy` models

Changing in the parameterization of models

As mentioned in Section 5.2.3, we adopted a fixed density profile in our `Cloudy` models, which strongly affects the predictions we derive, in particular for the PDR, and the molecular phase. While the density law was physically motivated (see Sections 1.2 and 5.2.3), it remains somewhat arbitrary and other density/pressure profiles could be explored. In particular, adopting a two-pressure profile may allow a more realistic representation of the sharp pressure changes experienced between the ionized and neutral phases.

Another improvement regarding the parameterization of models that was also mentioned in Section 8.2.3 is to use the volume-averaged ionization parameter $<U>$ as a substitute for the `Cloudy` input ionization parameter, defined at the illuminated edge U_{in}. As a matter of fact, both the density and ionization primary parameters were defined at the illuminated edge of our model. In practice, we would like to control parameters describing the physical conditions throughout the different phases, in particular when using an interpolation on those primary parameters. Hence, future improvements may include changing the input parameters we used in favor of volume- or luminosity-averaged parameters, such as $<U>$ or n.

Stellar and dust continuum

While the present study has focused on exploiting the information from emission lines, we do not use so far any constraint coming from the continuum. Indeed, taking into account the full continuum would enlarge significantly the memory size occupied by our grid and would require to adapt the way the comparison is performed. Nevertheless, there is much to gain in accounting for information from the continuum, which may allow constraining the dust and stellar properties. Specifically, the continuum produced by dust gives important information on the dust composition, size, mass, and temperature distributions, which are poorly known at low-metallicity. On the other hand, the stellar continuum provides information on the SFH, although its interpretation is made more complex because of extinction effects. Combinations of photometry with spectral information are for example incorporated in state-of-art fitting codes such as `CIGALE` (Boquien et al. 2019) or `BEAGLE` (Chevallard & Charlot 2016).

In particular, while our code is able to account for the presence of stellar populations with different ages (see multicluster configurations in Section 4.3), we have not yet exploited this configuration. Although the galaxies in the DGS are dominated by young bursts of stars, it has been shown that an older population of stars is always present (e.g., Aloisi et al. 2001, 2003, 2005, 2007) and may contribute to the total line and continuum emission. Hence, models combining an old stellar population with a younger burst of star formation may provide a more realistic picture of stellar properties. Another possibility would be to run models using a different SFH (e.g., continuous SFH) and compare or combine them with our current grid using only single bursts.

In the context of the escape fraction study (see Chapter 9), we stored the ionizing continuum (i.e., above below 912Å), in order to calculate the $f_{\rm esc}$ in different energy ranges (see Section 9.4). Storing the whole continuum rather than only the ionizing continuum would also allow us to derive interesting quantities, such as the $f_{1500\text{Å}}/f_{912\text{Å}}$ flux ratio, which is used in observational studies. This may facilitate the comparison between the predicted escape fractions and the observed ones (see discussion from Section 9.3). In the context of the molecular gas study (see Chapter 10), one possible interest of using the continuum information would be to better constrain the dust content in galaxies, which is crucial to accurately model the heating and cooling mechanisms, as well as the formation mechanisms of molecules such as H_2 and CO. Specifically, using the dust SED in the mid-IR would bring constraints on the presence of PAHs which dominate the photoelectric effect, while the far-IR bears information on the dust temperature related to the geometry of regions (e.g., compactness), and the submillimeter domain on the total dust masses. All these quantities, which are for now fixed by analytical prescriptions (see Chapter 5), are important to correctly estimate the gas mass distribution in galaxies.

Finally, accounting for the information given by the full spectrum (continuum + spectral lines) may be a crucial step to take in order to extrapolate our analysis to high-redshift objects, such as primordial galaxies. The information given by the continuum may help us relax some assumptions made in the current study: the IMF, the abundance patterns, the dust content, etc (see Chapter 5), which are essential elements to understand the heating and cooling of the gas. While we based those assumptions on empirical calibrations for well-known local BCDs, such calibrations are not available for high-redshift galaxies in which the dominant physical mechanisms may be drastically different.

Additional diffuse component

Regardless of how much it contributes to the emission, the ionized and neutral diffuse gas which surrounds the star-forming regions (i.e., the H II regions + PDR + molecular gas complexes) plays a crucial role in absorbing photons that may escape from density-bounded H II regions. This shortcoming was pointed out in Section 9.3 where the lack of a proper treatment for the diffused ionized and neutral gas has prevented us from converting our $f_{\rm esc,HII}$ estimates into a global $f_{\rm esc}$. The treatment of this additional component is somewhat less important in the CO-dark gas study presented in Chapter 10, because the diffuse gas likely represents only a small contribution to the total H_2 masses that we estimate.

One possibility to properly account for the diffuse gas component would be to store the full continuum spectrum that escapes out of our topological models (i.e., at any given cut) and use it as an input of another `Cloudy` models, with parameters tailored to reproduce the diffuse gas (in particular density). While this step is not easy to include in an automated routine, and requires a two-step modeling procedure, it would allow properly accounting for the effect of the diffuse gas, assuming in can be modeled by a single additional component.

Another, probably simpler option to account for the DIG contamination, would be to correct beforehand the fluxes which are used as inputs for `MULTIGRIS` in order to remove the DIG contribution. We note, however, that this solution does not help to estimate the global $f_{\rm esc}$ estimates as opposed to $f_{\rm esc,HII}$. Studies aiming at disentangling the DIG emission and the one from H II regions have been performed in the optical on a few galaxies resolved with MUSE (e.g., Della Bruna et al. 2020; Belfiore et al. 2022) and for larger samples like CALIFA (e.g., Lacerda et al. 2018), MaNGA (e.g., Zhang et al. 2017; Vale Asari et al. 2019; Mingozzi et al. 2020; Stark et al. 2021; Mannucci et al. 2021), MOSDEF (e.g., Sanders et al. 2020) but the contribution of the DIG to IR emission line remains unknown. Some progresses may be made with the advance of high-resolution IR telescopes like the JWST.

11.1.2 Improving `MULTIGRIS`

Hierarchical modeling

One interesting possibility offered by Bayesian statistics, which has not been explored in this thesis, is the possibility to adopt a hierarchical parameter structure by defining hyperparameters. Those hyperparameters encapsulate the information given by a sample of objects which are treated simultaneously rather than independently. For example, a hyperparameter could be used to control the density distribution used as a prior for the entire sample. The density for a given galaxy would then be drawn from the hyperparameter posterior PDF, at each step of the sampling. This feature is especially interesting in the context of large samples of galaxies, in which some objects are poorly constrained and others are well-constrained.

Adopting a hierarchical structure allows exploiting the information given by the whole sample to derive meaningful priors for the less constrained objects. Such methods have already been implemented in codes like `HerBIE` (Galliano et al. 2021), and has proved to significantly improve the posteriors PDFs of poorly constrained objects. This feature is currently being implemented in the

code for future studies.

Using machine learning techniques

Another improvement consists in using machine learning techniques in order to mitigate the long timescales necessary to run grids of models. The long time needed to re-run `Cloudy` models spanning a different parameter space, or updated with the latest improvements in terms of, e.g., stellar SEDs is one of the bottleneck of our method. Hence, machine learning techniques could be used in order to refine large grids of models without having to re-run additional models. They might also allow replacing the linear interpolation, which is time-consuming in high dimensionality space and cannot capture non-linear variations. Such methods are currently being developed, for example, in the machine learning library `ai4neb` (Morisset et al., in prep) and within the framework of the Bayesian code `HerBIE` (Tedros & Galliano, in prep). A possible improvement for `MULTIGRIS` may consist in adapting such methods in order to refine our models grid.

In parallel, several recent works have used machine learning technique to directly train networks to associate spectral features with physical parameters. For example, the code `GAME` (Ucci et al. 2017, 2019) uses supervised machine learning techniques in combination with grids of single-sector `Cloudy` models to directly infer key physical properties of galaxies (density, metallicity, column density, and ionization parameter) from their emission-line spectra.

In a similar vein, Bron et al. (2021) have used Random Forest algorithms trained on a grid of astrochemical models to search for the best observable tracer of the ionization fraction in dense gas. Random Forest algorithms have also been used in Dobbels & Baes (2021) to perform a pixel-by-pixel analysis to create high-resolution dust maps based on photometric bands. Focusing on GMCs, Kang et al. (2022) coupled a conditional invertible neural network with the `WARPFIELD`-emission predictor (Pellegrini et al. 2020) to estimate the physical properties of star-forming regions from spectral observations. They find that the use of a neural network allows them to predict the posterior PDFs very quickly and efficiently. While such machine learning tools are not yet widely used, those first results are promising and motivate the use of such techniques to explore grids of models and constrain the key physical parameters directly from the observed spectra.

From line measurements to the parameters

Finally, we are also considering extending the workflow of `MULTIGRIS` (see Figure 7.1) back to the line measurements themselves. This would for example consists in including the line fitting procedure within the Bayesian framework, for example by fitting the lines at each draws while accounting for the measured uncertainties. This would allow us to consider the actual PDF of the line flux instead of just a measured value and with error bars that assumes a Normal/asymmetric Normal distribution. This would also provide a global framework to go directly from the observations to the PDF of parameters, while including all the possible sources of uncertainties consistently.

11.2 Other applications of MULTIGRIS

11.2.1 Application on other samples of galaxies

Follow-up studies in the DGS

Even without including the improvements that we presented in Section 11.1, our code can already be used in its present version to infer various parameters in different samples of galaxies. A direct continuation of the work presented in this thesis consists in applying MULTIGRIS to the exact same sample of DGS galaxies in order to put constraints on other parameters than the escape fraction or the gas mass distribution. Specifically, the study on the molecular gas masses which was presented in Chapter 10 could be extended by looking, for example, at other quantities related to the molecular gas. Tracers such as the [C I] lines which were proposed in Madden et al. (2020) as a potentially more precise (i.e., less dependent on other parameters like the density) tracers of the molecular gas than [C II] could be examined. Other interesting quantities include the study of the dominant heating mechanisms which are all saved as outputs of Cloudy models (see Section 1.2.3, for an illustration with two typical models) and will allow investigating which physical mechanisms are dominant in the DGS conditions at low-metallicity (Lebouteiller et al., in prep).

Unresolved observations in the local Universe

While no statistical trend can be derived from individual analyses of galaxies, the application of MULTIGRIS to well-studied objects is an interesting process in order to assess in more details how the results based on topological models compare to other empirical diagnostics. It also allows testing various configurations and extracting the PDFs of different parameters, with a similar level of details as in the study of I Zw 18 (LR22). Ongoing projects include the application of our code to the well-known galaxy NGC 253 (Beck et al., in prep) to provide a refine model based on the new IR emission lines obtained with the airborne observatory SOFIA, which were presented in Beck et al. (2022). In particular, we are interested to study the metallicity of the nuclear regions and the possible spectral signatures associated with the AGN.

To extend the analysis of Chapter 9, one interesting pathway would be to apply MULTIGRIS to a sample of known leaking-galaxies. In particular, the recent Low-z Lyman Continuum Survey (Flury et al. 2022a,b) gathers 89 galaxies with detections and non-detections of LyC-leakage obtained with HST/COS. An ongoing project focuses on combining MULTIGRIS with the BOND database (Asari et al. 2016) to infer the escape fractions expected in this sample based on their optical emission lines (Marques-Chaves et al. in prep). This analysis is crucial to assess whether LyC-leaking galaxies in the primordial universe could be identified uniquely based on their optical emission lines that can be observed with facilities such as the JWST.

Finally, another project focusing on the RESOLVE sample aims at combining MULTIGRIS with LOC models (described in Section 4.4; Richardson et al. 2014, 2016; Richardson et al. 2022). The RESOLVE sample (Kannappan & Wei 2008; Polimera et al. 2022) is a volume-limited survey of galaxies at $z \sim 0$, including large fractions of dwarf galaxies, which were observed in the optical

range with the Sloan Digital Sky Survey. Such a large sample is well-suited to calibrate the parameter distribution of power-law models (see Section 4.4). It is also well-suited to test the hierarchical framework (Section 3.1).

The RESOLVE sample is especially interesting as some galaxies are suspected to host significant fractions of AGN and may be used to study which spectral signatures are associated with the presence of X-ray sources. This is especially interesting in the context of up-coming data with the JWST which will be ideally suited to detect spectral signatures of high energy sources in galaxies (e.g., [Mg IV]λ4.49μm, [Ne VI]λ7.64μm, and [Ne V]λ14.32μm; Satyapal et al. 2021; Richardson et al. 2022). The combination of such lines with the emission of neutral gas lines (e.g., [Fe II]λ1.64μm) boosted by X-ray photons may help disentangle the feedback from potential compact objects from the one induced by turbulence and shocks (Cresci et al. 2010, 2017). The latter IR tracers could be used in combination with other tracers of high energy photons available in the optical range (e.g., [Ne V], [Fe V]) and in the UV range (e.g., C IV λ1550, He II λ1640, Wofford et al. 2021; Berg et al. 2019; Berg et al. 2021). Looking for counter-parts in the radio and X-ray domains with sensitive instruments (e.g., with the forthcoming *Athena* telescope), may also allow progress on the identification of the exact nature of the different sources producing high energy photons.

High-redshift galaxies

As described in the previous section, `MULTIGRIS` can be used to compare the properties derived at various spatial scales in local galaxies in order to better understand the averaged properties that are inferred when no spatial information is available. This is, for example, the case for high-redshift galaxies for which only spatially unresolved spectra are available. Hence, tools such as `MULTIGRIS`, which allows recovering information on the topology, gas mass distribution, and integrated properties (such as the escape fraction) are well adapted to make the most of the current and up-coming spectra of high-redshift galaxies.

Although promising, this project bears its own caveats:

- First, a meaningful analysis of high-redshift galaxies can only be performed by comparing the observations to models that correspond to the physics at play in the early universe. How much information can be obtained with a tool like `MULTIGRIS` ultimately depends on how reliable the underlying models are.
- Secondly, the ability of `MULTIGRIS` to constrain posterior PDFs increases with the number of lines which are given as inputs. When only a few lines are available, additional priors (e.g., on the bolometric luminosity, total mass, metallicity...) are much needed to help in reducing the parameter space.

Interpreting observations of high-redshift galaxies is a complex task, which will require a proper calibration of the models based on the physical processes which are expected in the early universe. While in the local universe we could limit the parameter space to explore, based on empirical studies of local star-bursting dwarf galaxies, this empirical data is scarce for the primordial galaxies. Refined simulations which incorporate theoretical prescriptions for the early universe may help overcome this important caveat.

Joint efforts to understand the physics at play in the primordial universe from a theoretical perspective and to develop complex tools to analyze the observations are much needed to make the most of the spectra of the primordial galaxies that have been recently detected with facilities such as ALMA and the JWST.

11.2.2 Application to other data

Mock galaxies from simulations

Comparing observation-based approaches and theoretical approaches found in simulations offers interesting possibilities, in particular to better understand the biases introduced by the modeling (e.g., assuming a simplified geometry) while getting rid of other observational biases. Simulations also allow studying the temporal evolution of galaxies, while observations only grant access to a given snapshot of the object. Nevertheless, comparing the results obtained with different methods is not straightforward.

In order to make progress on these issues, several pathways could be explored. A first approach consists in running `MULTIGRIS` on simulated galaxies which have been post-processed with radiative transfer codes. This exercise may shed light on important questions such as: how well can `MULTIGRIS` recover a given gas distribution? How many lines are needed to constrain a given observable based of the simulated galaxies?

This approach is especially well-suited to test the validity of the different configurations (i.e., combinations of discrete sector or distribution of parameters) that were presented in Chapter 4. It is an important step to disentangle the quantities which can be correctly estimated with a simplified, single-sector or multisector topology from the ones which are intrinsically sensitive to the local distribution of the gas and should not be estimated with simplified models. Moreover, additional noises and biases can be introduced in simulated data in order to test how well we can deal with systematic uncertainties.

`MULTIGRIS` could be applied to both simulated observation of unresolved galaxies, and simulated resolved maps, hence allowing to examine the effects of changing the resolution on the PDFs that are derived. Using hydrodynamical simulations in which individual H II regions are resolved may also allow us to calibrate the distribution of parameters (e.g., power-law, broken power-law, and normal distribution) that we introduced in Section 4.4. We could, for example, assess whether the distribution of a given parameter can be easily parameterized using power-laws, broken power-laws, or normal distributions, and if other continuous functions may be more appropriate to describe their variations. We may also derive informative priors on the indices of such distributions.

Finally, mock observations can provide potentially any emission lines in any wavelength range. They are hence ideal to study which lines are the most informative to constrain the PDFs of a given parameter. As such, `MULTIGRIS` can be used as a predictive tool to prepare or motivate future observations.

Resolved observations

Although MULTIGRIS could be adapted to perform a pixel-by-pixel analysis of resolved maps, the configurations that were presented in Chapter 4 may not be well-suited for a comparison with individual sightlines. The complex task of modeling individual sightlines for which the emission may result from a mix of H II region+PDR was for example introduced in Lambert-Huyghe et al. (2022). While adopting a multiphase analysis such as the one presented in this thesis may be difficult due to the presence of multiple, possibly independent ionizing sources that irradiate the gas, ongoing projects have focused on analyzing the emission arising from well-studied GMCs from the Orion B project (Palud et al. in prep). Their approach is based on a Bayesian framework, similar to the one presented in Chapter 7, but includes in addition a regularization matrix which allows connecting the physical properties of adjacent pixels.

Another option to deal with pixel-by-pixel analyses would be to use the hierarchical framework (see Section 11.1.2) to define hyperparameters constrained by the distribution of all the pixels of a given observed map. This method provides a way to regularize parameters across a map, even though it does not use a spatial regularization function per se. In other words, the MCMC draws would follow the distribution of parameters (e.g., U, n) in the map of that particular object, which should result in fewer sharp changes from one pixel to the surrounding ones.

Another possibility would be to zoom on observations of individual H II regions around stellar clusters. Such an application may be of interest to lift several of the limitations which were mentioned in Chapter 9 and 10 (e.g., the unknown distribution of DIG, the impact of the resolution of $f_{\rm esc}$ or the distribution of molecular gas at galactic scales). In particular, it could be interesting to perform a multiscale analysis in order to compare the results obtained at resolved scale to the ones performed on whole galaxies. A similar approach was, for example, used in Polles et al. (2019) in IC10, but the results were limited by their statistical approach based on the χ^2 method.

Several telescopes are well suite to provide resolved observations of H II regions and GMCs. In particular, a resolved and multiwavelength picture of local galaxies can be obtained by combining the spectral information in the UV (e.g., with HST imaging), optical (e.g., with VLT/MUSE and the upcoming ELT and EELT), and infrared (e.g., with ALMA and JWST). Samples such as the PHANGS dataset which combines both high angular resolution data and a multi-wavelength coverage are ideally suited to bridge the gap between local physical conditions and the averaged parameters derived for galaxies. Such studies are crucial to understand the biases and systematic effects that may affect studies of unresolved galaxies.

11.3 Conclusion

In this thesis, we have presented a new framework adapted to study the multiphase ISM in galactic environments, including the low-metallicity ones. In such environments, important changes of the heating and cooling mechanisms are expected (see Chapter 1), with an impact on their spectral signatures. We have tried to use those spectral signatures as probes of the physical conditions in low-metallicity dwarf galaxies, which may resemble the conditions of the early universe (see Chapter 2).

The new framework presented in this thesis (see Chapters 4 and 7) is extensively based on previous works which have used photoionization and photodissociation models to constrain the key physical parameters in H II regions and PDRs (see examples in Section 3.1.2). More specifically, it builds on previous works in our team which have introduced and iteratively improved strategies to combine components as "topological" models, as described in Section 3.2.1. Because our new code heavily relies on pre-existing open source codes, such as the radiative transfer code `Cloudy` (Ferland et al. 2017) and the Bayesian statistics python package PyMC3 (Salvatier et al. 2016), this work has mostly focused on combining all those tools in a robust and consistent way. Hence, this has allowed us to reach an unprecedented level of complexity in terms of modeling based on photoionization and photodissociation models, and in terms of statistical formalism. In this conclusion, we briefly summarize why those improvements are important in the context of ISM studies and what steps forward were taken with respect to previous works.

While the topological models combining several components as sectors were already introduced in previous studies (from Section 3.2.1), the novelty of our approach has consisted in developing a *flexible* way to combine components. Instead of fixing arbitrarily the number of sectors/components needed to reproduce the observations, we could, for the first time, consider the number of sectors as a free parameter in our models and compare different configurations, under various assumptions. Additionally, the code enables considering various topologies (single or multisector, power-law and normal distributions) and new ones can be implemented relatively easily. The introduction of a metric (see, e.g., the marginal likelihood in Section 7.5) which allows us to compare quantitatively how different models perform at reproducing data, assuming some prior knowledge on the configuration, is an important step to understand the biases introduced by our modeling assumptions. While the choice of this metric is not easy and brings up questions regarding, for example, the minimum number of component needed to model a galaxy or the minimum number of lines needed to constrain a given parameter (see discussions in Lebouteiller & Ramambason 2022b and in Section 8.1.3), such metrics hold much more information than a simple χ^2 value.

Another interesting feature which comes with the use of Bayesian statistics is the possibility to derive posterior PDFs rather than a single most likely value. The interpretation of PDFs is more complex than interpreting a mean/median value with the associated error bars, and it has required us to invest some time to implement the tools that allow us to plot and examine PDFs (see Section 7.4.4). Nevertheless, by looking at the posterior PDFs rather than at the mean values, it becomes possible to examine the degeneracies, to study the correlations between multiple parameters, or to include informative priors when they are available. The study of posterior PDFs also allows us to determine, based on a simple visual examination, which parameters are not well constrained by the models. Hence, the Bayesian framework brings an important piece of information about the possibility to constrain a given parameter, based on a given set of data, and assuming a given configuration. Knowing which parameters are *not* well constrained in our models is important and may motivate further investigations with other approaches.

In the context of this thesis we have shown, for example, that the PDFs of the H II region escape fractions are well constrained by suites of IR emission lines arising from H II regions and PDRs (see Chapter 9). This finding motivates future works on other samples (see Section 11.2.1), as it shows that the complex $f_{\rm esc,HII}$ can be constrained based only on the information coming from spectra. Our work provides, for the first time, predictions for $f_{\rm esc,HII}$ in unresolved galaxies, based only on their emission

lines. Although this parameter is sensitive to modeling assumptions on the number of sectors, the trends we derive with physical parameters are robust and hold for various configurations. To which extent this analysis can be carried out in less ideal cases than the DGS, with fewer emission lines available, remains to be discussed (see Section 11.2.1). We are also limited by some missing physics in our models, such as the absence of a diffuse gas component, which needs to be accounted for to estimate global escape fractions from galaxies.

The study of the molecular gas (see Chapter 10) has revealed more limitations of our multicomponent models, which were unable to match the CO emission observed in the DGS galaxies. It has motivated the introduction of other topologies based on distributions of parameters (namely, power-laws and broken power-laws). While our results regarding the CO-dark molecular gas are in line with previous results from Madden et al. (2020), assuming much simpler, single-sector topologies, we find, however, that the CO-bright gas content in our model is mostly driven by our modeling assumptions and cannot be properly constrained. This may not be an issue to constrain the total molecular gas masses, since we find that the molecular gas mass is mostly dominated by CO-dark gas, and that our predictions experience little variations with changes of topologies. However, we find that quantifying the CO-dark gas fractions or the CO-to-H_2 conversion factor is not straightforward with our approach, as the variations we observe are mostly driven by our modeling assumptions on the dense gas distribution.

These conclusions highlight the impossibility to constrain all the parameters based on a single modeling approach, even when accounting for the complex multiphase topology of the ISM. The combination of results obtained with different approaches (e.g., empirical studies, 3D radiative transfers, or simulations at various spatial scales, as described in the overview of Section 3.1) is an essential step to enlarge our global understanding of the physics of the ISM. In that view, the Bayesian framework offers the interesting possibility to update our beliefs based on prior knowledge obtained with other methods. Reciprocally, the posterior PDFs obtained as outputs can be used to narrow down the parameters space explored with, e.g., refined numerical simulations. This complementarity between various methods may allow further progress as well as gains in terms of efficiency and computing power.

Part VI

Annexes

Appendix

Cloudy labels

Al 2 2660.35A	Ca 2 8498.02A	Fe 2 2739.55A	Fe 4 4906.56A
Al 2 2669.15A	Ca 2 8542.09A	Fe 2 2755.74A	Fe 4 6739.82A
Al 9 2.04444m	Ca 2 8662.14A	Fe 2 3318.38A	Fe 5 25.9131m
Ar 2 6.98337m	Ca 5 4.15739m	Fe 2 35.3394m	Fe 5 4071.24A
Ar 3 21.8253m	Ca 8 2.32117m	Fe 2 4.11394m	Fe 5 4227.19A
Ar 3 3109.18A	Cl 2 14.3639m	Fe 2 4177.20A	Fe 6 19.5527m
Ar 3 5191.82A	Cl 2 33.2721m	Fe 2 4233.17A	Fe 6 5145.76A
Ar 3 6.36676m	Cl 2 6161.84A	Fe 2 4244.81A	Fe 7 4988.55A
Ar 3 7135.79A	Cl 2 8578.70A	Fe 2 4287.39A	Fe 7 5720.71A
Ar 3 7751.11A	Cl 2 9123.60A	Fe 2 4352.78A	Fe 7 6086.97A
Ar 3 8.98898m	Cl 3 3342.80A	Fe 2 4359.33A	Fe10 6374.54A
Ar 4 2853.66A	Cl 3 3353.17A	Fe 2 4474.90A	Fe11 7891.87A
Ar 4 2868.22A	Cl 3 5517.71A	Fe 2 4492.63A	H 1 1.28180m
Ar 4 4711.26A	Cl 3 5537.87A	Fe 2 4515.19A	H 1 1.87510m
Ar 4 4740.12A	Cl 3 8433.66A	Fe 2 4584.00A	H 1 1.94454m
Ar 4 7170.70A	Cl 3 8480.86A	Fe 2 4889.62A	H 1 12.3684m
Ar 4 7237.77A	Cl 3 8500.01A	Fe 2 4905.34A	H 1 1215.67A
Ar 4 7263.33A	Cl 3 8547.96A	Fe 2 5.33881m	H 1 19.0565m
Ar 5 13.0985m	Cl 4 11.7629m	Fe 2 5158.78A	H 1 2.16551m
Ar 5 6435.12A	Cl 4 20.3197m	Fe 2 5220.06A	H 1 2.62513m
Ar 5 7.89971m	Cl 4 7530.54A	Fe 2 5261.62A	H 1 276.210m
Ar 5 7005.83A	Cl 4 8045.62A	Fe 2 5273.35A	H 1 3.29607m
Ar 6 4.52800m	Cl 5 6.70733m	Fe 2 5333.65A	H 1 3.73951m
C 1 1.06950m	Cr 2 8125.30A	Fe 2 5376.45A	H 1 3889.05A
C 1 370.269m	F 2 4789.45A	Fe 2 7155.16A	H 1 4.05113m
C 1 609.590m	F 4 3996.92A	Fe 2 7172.00A	H 1 4.65247m
C 1 8727.13A	F 4 4059.90A	Fe 2 8616.95A	H 1 4101.73A
C 1 9088.00A	F100 100.000m	Fe 3 22.9190m	H 1 4340.46A
C 1 9850.26A	F12 12.0000m	Fe 3 33.0270m	H 1 4861.33A
C 2 157.636m	F25 25.0000m	Fe 3 4607.11A	H 1 5.12722m
C 2 2837.00A	F60 60.0000m	Fe 3 4658.01A	H 1 5.90655m
C 2 4267.00A	Fe 1 2526.51A	Fe 3 4666.94A	H 1 518.618m
C 2 6578.05A	Fe 1 2729.74A	Fe 3 4701.62A	H 1 590.935m
C 2 6582.88A	Fe 2 1.25668m	Fe 3 4733.84A	H 1 6562.81A
C 2 9903.00A	Fe 2 1.32055m	Fe 3 4754.64A	H 1 7.45777m
C 3 2512.20A	Fe 2 1.37182m	Fe 3 4777.61A	H 1 7.50038m
C 3 4069.00A	Fe 2 1.64355m	Fe 3 4881.12A	H2 1.62938m
C 3 4649.00A	Fe 2 1.80939m	Fe 3 4930.44A	H2 1.63092m
C 4 4659.00A	Fe 2 16.7235m	Fe 3 4987.33A	H2 1.63619m
CO 1300.05m	Fe 2 17.9314m	Fe 3 5270.40A	H2 1.63877m
CO 2600.05m	Fe 2 22.8960m	Fe 4 2.80554m	H2 1.64487m
CO 325.137m	Fe 2 25.9811m	Fe 4 2.83329m	H2 1.64998m
CO 371.549m	Fe 2 2585.88A	Fe 4 2.83562m	H2 1.65803m
CO 433.438m	Fe 2 2599.40A	Fe 4 2.86447m	H2 1.66604m
CO 520.089m	Fe 2 26.5501m	Fe 4 2567.61A	H2 1.67457m
CO 650.074m	Fe 2 2611.87A	Fe 4 2823.50A	H2 1.68726m
CO 866.727m	Fe 2 2613.82A	Fe 4 2829.36A	H2 1.69490m
Ca 2 7291.47A	Fe 2 2625.67A	Fe 4 2835.74A	H2 1.71419m
Ca 2 7323.89A	Fe 2 2631.32A	Fe 4 3094.96A	H2 1.71901m

H2 1.74720m
H2 1.74760m
H2 1.77967m
H2 1.78746m
H2 1.83529m
H2 1.89145m
H2 1.95702m
H2 12.2752m
H2 17.0300m
H2 2.03320m
H2 2.12099m
H2 2.22269m
H2 2.40594m
H2 2.41307m
H2 2.42307m
H2 2.43683m
H2 2.45408m
H2 2.47510m
H2 2.49929m
H2 2.52709m
H2 2.55029m
H2 2.55785m
H2 2.55983m
H2 2.59469m
H2 2.62608m
H2 2.63426m
H2 2.67820m
H2 2.70951m
H2 2.71785m
H2 2.71922m
H2 2.72610m
H2 2.73042m
H2 2.77819m
H2 2.80176m
H2 2.83537m
H2 2.89689m
H2 2.96462m
H2 28.2130m
H2 3.11688m
H2 3.36534m
H2 3.36794m
H2 3.37087m
H2 3.38104m
H2 3.38671m
H2 3.40299m
H2 3.40989m
H2 3.43697m
H2 3.48530m
H2 3.54606m

H2 3.62531m
H2 3.72257m
H2 3.84534m
H2 3.99573m
H2 4.17983m
H2 4.40836m
H2 5.05148m
H2 5.50996m
H2 5.80953m
H2 6.10718m
H2 6.90725m
H2 8.02362m
H2 9.66228m
He 1 1.08303m
He 1 2.05813m
He 1 3187.74A
He 1 3613.64A
He 1 3634.08A
He 1 3819.61A
He 1 3888.63A
He 1 3889.00A
He 1 4026.20A
He 1 4471.49A
He 1 4713.03A
He 1 4921.93A
He 1 5047.64A
He 1 5875.64A
He 1 6678.15A
He 1 7065.22A
He 1 7281.35A
He 2 1.01233m
He 2 1640.43A
He 2 4.76203m
He 2 4685.64A
He 2 6.94582m
He 2 9.71043m
IRAC 3.60000m
IRAC 4.50000m
IRAC 5.80000m
IRAC 8.00000m
Inci 1215.00A
Inci 4860.00A
K 3 4.61683m
K 6 5.57324m
K 6 5602.44A
K 6 6228.57A
K 6 8.82061m
LBOL 500.050m
LFIR 350.000m

LTIR 500.500m
MIPS 160.000m
MIPS 24.0000m
MIPS 70.0000m
Mg 1 2852.13A
Mg 1 4562.60A
Mg 1 4571.10A
Mg 2 1.09142m
Mg 2 2790.78A
Mg 2 2795.53A
Mg 2 2798.00A
Mg 2 2802.71A
Mg 2 2928.63A
Mg 2 2936.51A
Mg 2 9218.25A
Mg 2 9244.27A
Mg 4 4.48712m
Mg 5 13.5464m
Mg 5 2782.76A
Mg 5 5.60700m
Mg 7 5.50177m
Mg 8 3.02765m
Mn 1 2797.32A
Mn 2 2588.97A
N 1 1.01089m
N 1 1.03977m
N 1 1.03982m
N 1 1.04072m
N 1 1.04076m
N 1 1.05003m
N 1 1.05070m
N 1 1.05134m
N 1 1.05206m
N 1 1.05338m
N 1 1.05496m
N 1 1.05633m
N 1 1.06232m
N 1 1.06440m
N 1 1.06530m
N 1 1.07135m
N 1 1.07180m
N 1 1.07579m
N 1 1.12271m
N 1 1.12662m
N 1 1.12917m
N 1 1.12943m
N 1 1.13139m
N 1 1.13232m
N 1 1.15661m

N 1 1.21300m
N 1 1.21868m
N 1 1.22039m
N 1 1.22888m
N 1 1.22892m
N 1 1.22985m
N 1 1.23288m
N 1 1.23816m
N 1 1.24613m
N 1 3466.50A
N 1 5197.90A
N 1 5200.26A
N 1 7423.64A
N 1 7442.30A
N 1 7468.31A
N 1 8184.86A
N 1 8188.01A
N 1 8200.36A
N 1 8210.71A
N 1 8216.34A
N 1 8223.13A
N 1 8242.39A
N 1 8680.28A
N 1 8683.40A
N 1 8686.15A
N 1 8703.25A
N 1 8711.70A
N 1 8718.84A
N 1 8728.90A
N 1 9386.80A
N 1 9798.56A
N 1 9810.00A
N 1 9822.75A
N 1 9834.61A
N 1 9872.15A
N 1 9931.47A
N 2 121.767m
N 2 205.244m
N 2 3062.90A
N 2 4041.00A
N 2 5005.00A
N 2 5754.61A
N 2 6548.05A
N 2 6583.45A
N 3 4379.00A
N 3 57.3238m
NH3 523.514m
Na 1 5889.95A
Na 1 5895.92A

Na 3 7.31706m	O 2 3728.81A	S 2 1.02867m
Ne 2 12.8101m	O 2 3867.00A	S 2 1.03205m
Ne 3 15.5509m	O 2 4075.00A	S 2 1.03364m
Ne 3 3342.18A	O 2 4094.00A	S 2 1.03705m
Ne 3 36.0036m	O 2 4111.00A	S 2 4068.60A
Ne 3 3868.76A	O 2 4152.00A	S 2 4076.35A
Ne 3 3967.47A	O 2 4278.00A	S 2 6716.44A
Ne 3 4011.68A	O 2 4294.00A	S 2 6730.82A
Ne 4 4714.23A	O 2 4341.00A	S 3 18.7078m
Ne 4 4715.64A	O 2 4342.00A	S 3 33.4704m
Ne 4 4724.17A	O 2 4607.00A	S 3 3721.63A
Ne 4 4725.58A	O 2 4651.00A	S 3 3797.17A
Ne 5 14.3228m	O 2 7318.92A	S 3 6312.06A
Ne 5 24.2065m	O 2 7319.99A	S 3 8829.39A
Ne 5 2973.20A	O 2 7329.67A	S 3 9068.62A
Ne 5 3345.99A	O 2 7330.73A	S 3 9530.62A
Ne 5 3426.03A	O 3 4363.21A	S 4 10.5076m
Ne 6 7.64318m	O 3 4931.23A	S 9 1.25201m
Ni 2 10.6793m	O 3 4958.91A	SPR1 250.000m
Ni 2 6365.10A	O 3 5006.84A	SPR2 350.000m
Ni 2 6666.80A	O 3 51.8004m	SPR3 500.000m
Ni 2 7377.83A	O 3 88.3323m	Si 1 2528.51A
Ni 2 7411.61A	O 4 25.8832m	Si 2 34.8046m
Ni 4 5363.34A	O 4 4632.00A	Si 6 1.96247m
O 1 1.12863m	P 2 32.8620m	Si 7 2.48071m
O 1 1.12864m	P 2 60.6263m	Si 7 6.51288m
O 1 1.12873m	P 3 17.8797m	Si 9 2.58394m
O 1 1.31639m	PAC1 70.0000m	Si10 1.43008m
O 1 1.31649m	PAC2 100.000m	Si11 1.93446m
O 1 1.31651m	PAC3 160.000m	Ti 2 3237.15A
O 1 145.495m	PAH 11.3000m	Ti 2 3364.74A
O 1 5577.34A	PAH 11.8000m	WISE 12.0820m
O 1 63.1679m	PAH 13.3000m	WISE 22.1940m
O 1 6300.30A	PAH 3.30000m	WISE 3.36800m
O 1 6363.78A	PAH 6.20000m	WISE 4.61800m
O 1 7773.00A	PAH 7.90000m	^{1}3CO 2719.67m
O 1 8446.25A	PAHC 10.9000m	^{13}CO 453.374m
O 1 8446.36A	PAHC 12.6500m	^{13}CO 544.012m
O 1 8446.76A	PAHC 14.1000m	^{13}CO 679.978m
O 2 2470.22A	PAHC 3.23000m	
O 2 2470.34A	PAHC 3.37000m	
O 2 2470.34A	PAHC 5.65000m	
O 2 3726.03A	PAHC 6.90000m	

Table A1: List of the lines (516) saved in the SFGX grid.

Glossary

$\hat{R}$ The rank normalized R-hat diagnostic tests for lack of convergence by comparing the variance between multiple chains to the variance within each chain.

case B Corresponds to an optically thick nebula in which hydrogen recombinations to the ground level are reabsorbed "on-the-spot" and do not contribute to the net recombination.

CO-dark Component of the molecular gas that is unseen in CO emission.

configuration Set of priors for a representative model of a galaxy seen as an arrangement of several sectors around one or several clusters.

corner plot Illustrative representation of different projections of samples in high dimensional spaces..

density-bounded Property of an H II region which is delimited by the lack of matter, with an outer radius smaller than its Strömgren radius..

elpd The elpd is the theoretical expected log pointwise predictive density for a new dataset, which can be estimated using cross-validation. It calculates the predictive density values for each point of observation..

emergent Which includes the effects of absorption by dust and scattering of the line from regions away from that where the lines form. It correspond to the reddened fluxes.

ergodicity Property verified by a process for which the temporal mean is equal to the spatial mean, when its trajectory becomes infinitely long.

escape fraction Fractions of photons that escape from density-bounded regions within a given energy range (usually above 1 Ryd).

evidence See marginal likelihood.

extensive Which scales with the total luminosity and can be expressed as a linear combination using the mixing-weights of each sector/cluster..

importance sampling The importance sampling is a variance reduction technique that can be used in the Monte Carlo method. It assesses the "important" parameters that impact more the estimates and sample them more frequently..

intensive Which does not scale with the total luminosity but can be expressed as a ratio of extensive parameters. It correspond to the dereddened fluxes..

intrinsic Includes all the physics of line formation, including destruction by local background opacities such as grains or photoelectric absorption, but does not include the effects of absorption and scattering of the line from regions away from that where the lines form..

likelihood The likelihood function is the joint probability of the observed data viewed as a function of the parameters of the chosen statistical model..

marginal likelihood Integration of the likelihood over the parameters of a model (see Eqaution 7.2).

p-value P-value is the level of marginal significance within a statistical hypothesis test, representing the probability of the occurrence of a given event.

posterior A posterior probability, in Bayesian statistics, is the distribution of a given parameter of observable after taking into consideration new information from the data.

primary parameter Parameter which is used to uniquely define a model in the grid.

prior A prior probability, in Bayesian statistics, is the distribution of a given parameter of observable before taking into consideration any new information.

radiation-bounded Property of an H II region which is delimited by the photoionisation and recombination equilibrium, with an outer radius equal to its Strömgren radius..

ray-tracing Ray tracing is a numerical method for calculating the path of waves or particles through a system.

secondary parameter Any parameter which can be extracted from the models a posteriori.

Strömgren radius The maximum distance at which ultraviolet from the star will keep hydrogen ionized based om the balance between the photoionization and recombination/.

tempered likelihood The tempered likelihood is used in tempering methods which come from the statistical physics formalism and are designed to find the lowest free energy state of a system of many interacting particles at low temperature.

topology Representative model based on linear combinations of a few or numerous components (see Chapter 4.

Acronyms

AGB Asymptotic Giant Branch.

AGN Active Galactic Nuclei.

BCD Blue Compact Dwarfs.

CNM Cold Neutral Medium.

CR Cosmic rays.

DGS Dwarf Galaxy Survey.

DIG Diffuse ionized gas.

dIrr Dwarf irregular galaxies.

DM Dark matter.

EoR Epoch of Reionization.

ESS Effective Sample Size.

FUV Far Ultra-violet.

GMC Molecular cloud.

GRB Gamma Ray Burst.

HDF5 Hierarchical Data Format (5).

HDI High Density Interval.

HMXB High-mass X-ray binaries.

IF Ionization front.

IGM Intergalactic Medium.

IMF Initial Mass Function.

IR Infrared.

ISM Interstellar Medium.

ISRF Interstellar radiation field.

KDE Kernel Density Estimate.

LMC Large Magellanic Cloud.

LOC Locally Optimally emitting Clouds.

LOO Leave-one-out cross-validation.

LyC Lyman continuum.

MAD Median Absolute Deviation.

MC Molecular cloud.

MCMC Monte Carlo Markov Chain.

MCSE Markov Chain Standard Error.

MIR Mid-infrared.

MZ Mass-metallicity relation.

NIR Near-infrared.

NUTS No-U-Turn sampler.

PAH Polycyclic Aromatic Hydrocarbons.

PDF Probability Distribution Function.

PDR Photodissociation region.

Pop III Population III stars.

RT Radiative transfers.

RV Random variable.

SED Spectral energy distribution.

SFGX Star-forming galaxies with an X-ray source.

SFH Star formation history.

SFR Star formation rate.

SMC Small Magellanic Cloud.

SMC Sequential Monte-Carlo.

SN Supernova.

SQL Structured query language.

sSFR Specific star formation rate.

SUE Skewed Uncertainty Ellipse.

ULX Ultra-Luminous X-ray sources.

WAIC Widely applicable information criterion.

WNM Warm Neutral Medium.

XDR X-ray dominated region.

XMP Extremely Metal Poor galaxies.

Bibliography

Accurso, G., Saintonge, A., Catinella, B., et al. 2017, MNRAS, 470, 4750

Ackermann, M., Ajello, M., Allafort, A., et al. 2012, ApJ, 755, 22

Adamo, A., Östlin, G., Zackrisson, E., et al. 2010, MNRAS, 407, 870

Adams, N. J., Conselice, C. J., Ferreira, L., et al. 2023, MNRAS, 518, 4755

Allen, M. G., Groves, B. A., Dopita, M. A., Sutherland, R. S., & Kewley, L. J. 2008, ApJS, 178, 20

Allen, R. J., Hogg, D. E., & Engelke, P. D. 2015, AJ, 149, 123

Almeida, J. S., Pérez-Montero, E., Morales-Luis, A. B., et al. 2016, ApJ, 819, 110

Aloisi, A., Clampin, M., Diolaiti, E., et al. 2001, AJ, 121, 1425

Aloisi, A., Clementini, G., Tosi, M., et al. 2007, ApJL, 667, L151

Aloisi, A., Savaglio, S., Heckman, T. M., et al. 2003, ApJ, 595, 760

Aloisi, A., van der Marel, R. P., Mack, J., et al. 2005, ApJL, 631, L45

Amorín, R., Muñoz-Tuñón, C., Aguerri, J. A. L., & Planesas, P. 2016, A&A, 588, A23

Anders, P., de Grijs, R., Fritze-v. Alvensleben, U., & Bissantz, N. 2004, MNRAS, 347, 17

Andrae, R., Schulze-Hartung, T., & Melchior, P. 2010, arXiv e-prints, arXiv:1012.3754

Annibali, F., Tosi, M., Pasquali, A., et al. 2015, AJ, 150, 143

Asari, N. V., Stasińska, G., Morisset, C., & Fernandes, R. C. 2016, Mon. Not. R. Astron. Soc., 460, 1739, arXiv: 1605.01057

Asplund, M., Grevesse, N., Sauval, A. J., & Scott, P. 2009, Annual Review of Astronomy and Astrophysics, 47, 481

Atek, H., Shuntov, M., Furtak, L. J., et al. 2022, MNRAS[[arXiv]2207.12338]

Aykutalp, A., Barrow, K. S. S., Wise, J. H., & Johnson, J. L. 2020, ApJL, 898, L53

Baldwin, J. A., Spinrad, H., & Terlevich, R. 1982, MNRAS, 198, 535

Barman, S., Neelamkodan, N., Madden, S. C., et al. 2022, ApJ, 930, 100

Bassett, R., Ryan-Weber, E. V., Cooke, J., et al. 2019, MNRAS, 483, 5223

Basu-Zych, A. R., Lehmer, B., Fragos, T., et al. 2016, ApJ, 818, 140

Bayes, T. 1763, M. F. R. S, Phil. Trans. R. Soc., 53, 370

Beck, A., Lebouteiller, V., Madden, S. C., et al. 2022, A&A, 665, A85

Begley, R., Cullen, F., McLure, R. J., et al. 2022, MNRAS, 513, 3510

Belfiore, F., Santoro, F., Groves, B., et al. 2022, A&A, 659, A26

Bell, E. F. 2003, ApJ, 586, 794

Berg, D. A., Chisholm, J., Erb, D. K., et al. 2021, ApJ, 922, 170

Berg, D. A., Erb, D. K., Henry, R. B. C., Skillman, E. D., & McQuinn, K. B. W. 2019, ApJ, 874, 93

Bergvall, N., Zackrisson, E., Andersson, B. G., et al. 2006, A&A, 448, 513

Berzin, E., Secunda, A., Cen, R., Menegas, A., & Götberg, Y. 2021, ApJ, 918, 5

Béthermin, M., Dessauges-Zavadsky, M., Faisst, A. L., et al. 2020, The Messenger, 180, 31

Bialy, S. & Sternberg, A. 2019, ApJ, 881, 160

Bian, F., Fan, X., McGreer, I., Cai, Z., & Jiang, L. 2017, ApJL, 837, L12

Bianchi, S., De Vis, P., Viaene, S., et al. 2018, A&A, 620, A112

Bisbas, T. G., Bell, T. A., Viti, S., Yates, J., & Barlow, M. J. 2012, MNRAS, 427, 2100

Bisbas, T. G., Haworth, T. J., Barlow, M. J., et al. 2015, MNRAS, 454, 2828

Bisbas, T. G., Schruba, A., & van Dishoeck, E. F. 2019, MNRAS, 485, 3097

Bisbas, T. G., Tan, J. C., & Tanaka, K. E. I. 2021, MNRAS, 502, 2701

Bisbas, T. G., Tanaka, K. E. I., Tan, J. C., Wu, B., & Nakamura, F. 2017, ApJ, 850, 23

Blanc, G. A., Kewley, L., Vogt, F. P. A., & Dopita, M. A. 2014, ApJ, 798, 99

Bolatto, A. D., Leroy, A. K., Rosolowsky, E., Walter, F., & Blitz, L. 2008, ApJ, 686, 948

Bolatto, A. D., Wolfire, M., & Leroy, A. K. 2013, Annual Review of Astronomy and Astrophysics, 51, 207

Boquien, M., Burgarella, D., Roehlly, Y., et al. 2019, A&A, 622, A103

Borthakur, S., Heckman, T. M., Leitherer, C., & Overzier, R. A. 2014, Science, 346, 216

Bradley, T. R., Knapen, J. H., Beckman, J. E., & Folkes, S. L. 2006, A&A, 459, L13

Bromm, V., Yoshida, N., Hernquist, L., & McKee, C. F. 2009, Nature, 459, 49

Bron, E., Agúndez, M., Goicoechea, J. R., & Cernicharo, J. 2018, arXiv:1801.01547 [astro-ph], arXiv: 1801.01547

Bron, E., Le Bourlot, J., & Le Petit, F. 2014, A&A, 569, A100

Bron, E., Roueff, E., Gerin, M., et al. 2021, A&A, 645, A28

Brooks, A. M., Governato, F., Booth, C. M., et al. 2007, ApJL, 655, L17

Brorby, M., Kaaret, P., & Prestwich, A. 2014, MNRAS, 441, 2346

Burkhart, B. 2018, ApJ, 863, 118

Burkhart, B. & Mocz, P. 2019, ApJ, 879, 129

Burton, M. G., Hollenbach, D. J., & Tielens, A. G. G. M. 1990, ApJ, 365, 620

Calzetti, D., Liu, G., & Koda, J. 2012, ApJ, 752, 98

Cardamone, C., Schawinski, K., Sarzi, M., et al. 2009, Monthly Notices of the Royal Astronomical Society, 399, 1191

Carton, D., Brinchmann, J., Contini, T., et al. 2018, MNRAS, 478, 4293

Chastenet, J., Sandstrom, K., Chiang, I.-D., et al. 2019, ApJ, 876, 62

Chevallard, J. & Charlot, S. 2016, MNRAS, 462, 1415

Chevallard, J., Charlot, S., Wandelt, B., & Wild, V. 2013, MNRAS, 432, 2061

Chevance, M. 2016, PhD thesis, Universite de Paris Diderot (Paris VII), France

Chevance, M., Kruijssen, J. M. D., Krumholz, M. R., et al. 2022, MNRAS, 509, 272

Chevance, M., Kruijssen, J. M. D., Vazquez-Semadeni, E., et al. 2020a, , 216, 50

Chevance, M., Madden, S. C., Fischer, C., et al. 2020b, MNRAS, 494, 5279

Chevance, M., Madden, S. C., Lebouteiller, V., et al. 2016, A&A, 590, A36

Ching, J. & Chen, Y. 2007, Journal of Engineering Mechanics, 133

Chisholm, J., Gazagnes, S., Schaerer, D., et al. 2018, 12

Chisholm, J., Prochaska, J. X., Schaerer, D., Gazagnes, S., & Henry, A. 2020, MNRAS, 498, 2554

Chisholm, J., Tremonti, C. A., Leitherer, C., et al. 2015, ApJ, 811, 149

Choi, Y., Dalcanton, J. J., Williams, B. F., et al. 2020, ApJ, 902, 54

Consiglio, S. M., Turner, J. L., Beck, S., & Meier, D. S. 2016, ApJL, 833, L6

Cormier, D., Abel, N. P., Hony, S., et al. 2019, A&A, 626, A23

Cormier, D., Bendo, G. J., Hony, S., et al. 2017, MNRAS, 468, L87

Cormier, D., Lebouteiller, V., Madden, S. C., et al. 2012, A&A, 548, A20

Cormier, D., Madden, S. C., Lebouteiller, V., et al. 2015, A&A, 578, A53

Cormier, D., Madden, S. C., Lebouteiller, V., et al. 2014, A&A, 564, A121

Cowie, L. L. & Songaila, A. 1986, Annual Review of Astronomy and Astrophysics, 24, 499

Cowie, L. L., Songaila, A., Hu, E. M., & Cohen, J. G. 1996, AJ, 112, 839

Cresci, G., Vanzi, L., Sauvage, M., Santangelo, G., & van der Werf, P. 2010, A&A, 520, A82

Cresci, G., Vanzi, L., Telles, E., et al. 2017, A&A, 604, A101

Crocker, A. F., Pellegrini, E., Smith, J. D. T., et al. 2019, ApJ, 887, 105

Crowther, P. A., Schnurr, O., Hirschi, R., et al. 2010, MNRAS, 408, 731

Croxall, K. V., Smith, J. D., Pellegrini, E., et al. 2017, ApJ, 845, 96

Croxall, K. V., van Zee, L., Lee, H., et al. 2009, ApJ, 705, 723

Curti, M., Mannucci, F., Cresci, G., & Maiolino, R. 2020, MNRAS, 491, 944

Dabrowski, I. 1984, Canadian Journal of Physics, 62, 1639

Dale, D. A., Helou, G., Neugebauer, G., et al. 2001, AJ, 122, 1736

Davé, R., Anglés-Alcázar, D., Narayanan, D., et al. 2019, MNRAS, 486, 2827

Davé, R., Rafieferantsoa, M. H., & Thompson, R. J. 2017, MNRAS, 471, 1671

Davé, R., Thompson, R., & Hopkins, P. F. 2016, MNRAS, 462, 3265

de Barros, S., Vanzella, E., Amorín, R., et al. 2016, A&A, 585, A51

De Breuck, C., Weiß, A., Béthermin, M., et al. 2019, A&A, 631, A167

de la Fuente, E., Porras, A., Trinidad, M. A., et al. 2020, MNRAS, 492, 895

De Looze, I., Cormier, D., Lebouteiller, V., et al. 2014, A&A, 568, A62

Dekel, A., Birnboim, Y., Engel, G., et al. 2009, Nature, 457, 451

Dekel, A. & Silk, J. 1986, ApJ, 303, 39

Del Zanna, G., Dere, K. P., Young, P. R., & Landi, E. 2021, ApJ, 909, 38

Della Bruna, L., Adamo, A., Bik, A., et al. 2020, A&A, 635, A134

Dempsey, J., McClure-Griffiths, N. M., Murray, C., et al. 2022, Publications of the Astron. Soc. of Australia, 39, e034

Dere, K. P., Landi, E., Mason, H. E., Monsignori Fossi, B. C., & Young, P. R. 1997, A&AS, 125, 149

Dobbels, W. & Baes, M. 2021, A&A, 655, A34

Dopita, M. A., Rich, J., Vogt, F. P. A., et al. 2014, Astrophysics and Space Science, 350, 741

Dopita, M. A., Sutherland, R. S., Nicholls, D. C., Kewley, L. J., & Vogt, F. P. A. 2013, ApJS, 208, 10

Draine, B. T. 2011, Physics of the Interstellar and Intergalactic Medium

Dunne, L., Maddox, S. J., Papadopoulos, P. P., Ivison, R. J., & Gomez, H. L. 2022, MNRAS, 517, 962

Dunne, L., Maddox, S. J., Vlahakis, C., & Gomez, H. L. 2021, MNRAS, 501, 2573

Eggen, N. R., Scarlata, C., Skillman, E., & Jaskot, A. 2021, ApJ, 912, 12

Eldridge, J. J., Stanway, E. R., Xiao, L., et al. 2017, Publications of the Astron. Soc. of Australia, 34, e058

Ercolano, B., Barlow, M. J., & Storey, P. J. 2005, MNRAS, 362, 1038

Ferland, G. J., Chatzikos, M., Guzmán, F., et al. 2017, Revista Mexicana de Astronomía y Astrofísica, 53, 385

Ferland, G. J., Porter, R. L., van Hoof, P. A. M., et al. 2013, Revista Mexicana de Astronomía y Astrofísica, 49, 137

Fernández, V., Amorín, R., Pérez-Montero, E., et al. 2022, MNRAS, 511, 2515

Fernández, V., Terlevich, E., Díaz, A. I., & Terlevich, R. 2019, MNRAS, 487, 3221

Fernández-Ontiveros, J. A., Pérez-Montero, E., Vílchez, J. M., Amorín, R., & Spinoglio, L. 2021, A&A, 652, A23

Feroz, F. & Hobson, M. P. 2008, MNRAS, 384, 449

Feroz, F., Hobson, M. P., & Bridges, M. 2009, MNRAS, 398, 1601

Field, G. B. & Saslaw, W. C. 1965, ApJ, 142, 568

Finkelstein, S. L., D'Aloisio, A., Paardekooper, J.-P., et al. 2019, ApJ, 879, 36

Fletcher, T. J., Tang, M., Robertson, B. E., et al. 2019, ApJ, 878, 87

Flury, S. R., Jaskot, A. E., Ferguson, H. C., et al. 2022a, ApJS, 260, 1

Flury, S. R., Jaskot, A. E., Ferguson, H. C., et al. 2022b, ApJ, 930, 126

Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, Publications of the ASP, 125, 306

Galliano, F. 2018, MNRAS, 476, 1445

Galliano, F. 2022, Habilitation Thesis, 1

Galliano, F., Nersesian, A., Bianchi, S., et al. 2021, A&A, 649, A18

Gao, Y., Gu, Q., Shi, Y., et al. 2022, A&A, 661, A136

Garg, P., Narayanan, D., Byler, N., et al. 2022, ApJ, 926, 80

Gazagnes, S., Chisholm, J., Schaerer, D., Verhamme, A., & Izotov, Y. 2020, A&A, 639, A85

Gazagnes, S., Chisholm, J., Schaerer, D., et al. 2018, A&A, 616, A29

Gelman, A., H. J. . V. A. 2014, Stat Comput, 24, 997

Genzel, R., Tacconi, L. J., Lutz, D., et al. 2015, ApJ, 800, 20

Girichidis, P., Offner, S. S. R., Kritsuk, A. G., et al. 2020, , 216, 68

Glover, S. C. O. & Clark, P. C. 2012a, MNRAS, 421, 116

Glover, S. C. O. & Clark, P. C. 2012b, MNRAS, 421, 9

Glover, S. C. O. & Clark, P. C. 2012c, MNRAS, 426, 377

Glover, S. C. O., Federrath, C., Mac Low, M. M., & Klessen, R. S. 2010, MNRAS, 404, 2

Glover, S. C. O. & Mac Low, M. M. 2011, MNRAS, 412, 337

Gnat, O. & Ferland, G. J. 2012, ApJS, 199, 20

Gnedin, N. Y. & Madau, P. 2022, Living Reviews in Computational Astrophysics, 8, 3

Gordon, K. D., Engelbracht, C. W., Rieke, G. H., et al. 2008, ApJ, 682, 336

Grebel, E. K. 2001, Astrophysics and Space Science Supplement, 277, 231

Grenier, I. A., Casandjian, J.-M., & Terrier, R. 2005, Science, 307, 1292

Greve, A., Becker, R., Johansson, L. E. B., & McKeith, C. D. 1996, A&A, 312, 391

Gross, A. C., Prestwich, A., & Kaaret, P. 2021, MNRAS, 505, 610

Grudić, M. Y., Hopkins, P. F., Faucher-Giguère, C.-A., et al. 2018, MNRAS, 475, 3511

Guseva, N. G., Izotov, Y. I., Fricke, K. J., & Henkel, C. 2012, A&A, 541, A115

Guseva, N. G., Izotov, Y. I., Fricke, K. J., & Henkel, C. 2015, A&A, 579, A11

Haemmerlé, L., Mayer, L., Klessen, R. S., et al. 2020, , 216, 48

Hao, C.-N., Kennicutt, R. C., Johnson, B. D., et al. 2011, ApJ, 741, 124

Harikane, Y., Ouchi, M., Inoue, A. K., et al. 2020, ApJ, 896, 93

Hashimoto, T., Inoue, A. K., Mawatari, K., et al. 2019, Publications of the ASJ, 71, 71

Haurberg, N. C., Rosenberg, J., & Salzer, J. J. 2013, ApJ, 765, 66

Hayashi, K., Okamoto, R., Yamamoto, H., et al. 2019, ApJ, 878, 131

Heintz, K. E., Björnsson, G., Neeleman, M., et al. 2021, MNRAS, 507, 1434

Heintz, K. E. & Watson, D. 2020, ApJL, 889, L7

Henkel, C., Hunt, L. K., & Izotov, Y. I. 2022, Galaxies, 10, 11

Henry, A., Berg, D. A., Scarlata, C., Verhamme, A., & Erb, D. 2018, ApJ, 855, 96

Henry, A., Scarlata, C., Domínguez, A., et al. 2013, ApJL, 776, L27

Henry, A., Scarlata, C., Martin, C. L., & Erb, D. 2015, ApJ, 809, 19

Hernandez, S., Aloisi, A., James, B. L., et al. 2021, ApJ, 908, 226

Heyer, M. & Dame, T. M. 2015, Annual Review of Astronomy and Astrophysics, 53, 583

Heyer, M., Krawczyk, C., Duval, J., & Jackson, J. M. 2009, ApJ, 699, 1092

Hilditch, R. W., Howarth, I. D., & Harries, T. J. 2005, MNRAS, 357, 304

Hirschmann, M., Charlot, S., Feltre, A., et al. 2017, MNRAS, 472, 2468

Hirschmann, M., Charlot, S., Feltre, A., et al. 2019, MNRAS, 487, 333

Hoffman, M. D. & Gelman, A. 2011, arXiv:1111.4246 [cs, stat], arXiv: 1111.4246

Hogarth, L., Amorín, R., Vílchez, J. M., et al. 2020, MNRAS, 494, 3541

Hosokawa, T. & Inutsuka, S.-i. 2005, ApJ, 623, 917

Hosokawa, T. & Inutsuka, S.-i. 2006, ApJ, 646, 240

Hsyu, T., Cooke, R. J., Prochaska, J. X., & Bolte, M. 2018, ApJ, 863, 134

Hsyu, T., Cooke, R. J., Prochaska, J. X., & Bolte, M. 2020, ApJ, 896, 77

Hu, C.-Y., Sternberg, A., & van Dishoeck, E. F. 2021, ApJ, 920, 44

Hunt, L. K., Testi, L., Casasola, V., et al. 2014, A&A, 561, A49

Hunt, L. K., Thuan, T. X., Sauvage, M., & Izotov, Y. I. 2006, ApJ, 653, 222

Indriolo, N., Geballe, T. R., Oka, T., & McCall, B. J. 2007, ApJ, 671, 1736

Inoue, A. K. 2010, MNRAS, 401, 1325

Israel, F. P. 2005, A&A, 438, 855

Izotov, Y. I., Chaffee, F. H., Foltz, C. B., et al. 1999, ApJ, 527, 757

Izotov, Y. I., Chaffee, F. H., & Green, R. F. 2001, ApJ, 562, 727

Izotov, Y. I., Guseva, N. G., Fricke, K. J., et al. 2021a, A&A, 646, A138

Izotov, Y. I., Noeske, K. G., Guseva, N. G., et al. 2004a, A&A, 415, L27

Izotov, Y. I., Orlitová, I., Schaerer, D., et al. 2016a, Nature, 529, 178

Izotov, Y. I., Papaderos, P., Guseva, N. G., Fricke, K. J., & Thuan, T. X. 2004b, A&A, 421, 539

Izotov, Y. I., Schaerer, D., Thuan, T. X., et al. 2016b, MNRAS, 461, 3683

Izotov, Y. I., Schaerer, D., Worseck, G., et al. 2018a, MNRAS, 474, 4514

Izotov, Y. I., Schaerer, D., Worseck, G., et al. 2020, MNRAS, 491, 468

Izotov, Y. I., Stasińska, G., Meynet, G., Guseva, N. G., & Thuan, T. X. 2006, A&A, 448, 955

Izotov, Y. I. & Thuan, T. X. 1999, ApJ, 511, 639

Izotov, Y. I., Thuan, T. X., & Guseva, N. G. 2017, MNRAS, 471, 548

Izotov, Y. I., Thuan, T. X., & Guseva, N. G. 2019, MNRAS, 483, 5491

Izotov, Y. I., Thuan, T. X., & Guseva, N. G. 2021b, MNRAS, 508, 2556

Izotov, Y. I., Thuan, T. X., & Privon, G. 2012, MNRAS, 427, 1229

Izotov, Y. I., Worseck, G., Schaerer, D., et al. 2021c, MNRAS, 503, 1734

Izotov, Y. I., Worseck, G., Schaerer, D., et al. 2018b, MNRAS, 478, 4851

Jaskot, A. E., Dowd, T., Oey, M. S., Scarlata, C., & McKinney, J. 2019, ApJ, 885, 96

Jaskot, A. E. & Oey, M. S. 2013, ApJ, 766, 91

Jenkins, E. B. & Wallerstein, G. 2017, ApJ, 838, 85

Ji, X. & Yan, R. 2022, A&A, 659, A112

Jiao, Q., Zhao, Y., Lu, N., et al. 2019, ApJ, 880, 133

Jin, Y., Kewley, L. J., & Sutherland, R. 2022, ApJ, 927, 37

Joblin, C., Bron, E., Pinto, C., et al. 2018, A&A, 615, A129

Johnson, K. E., Indebetouw, R., Watson, C., & Kobulnicky, H. A. 2004, AJ, 128, 610

Jones, A. P., Köhler, M., Ysard, N., Bocchio, M., & Verstraete, L. 2017, A&A, 602, A46

Juvela, M. 2019, A&A, 622, A79

Juvela, M. 2020, A&A, 644, A151

Kaaret, P. & Feng, H. 2013, in AAS/High Energy Astrophysics Division, Vol. 13, AAS/High Energy Astrophysics Division #13, 402.06

Kaaret, P., Schmitt, J., & Gorski, M. 2011, ApJ, 741, 10

Kang, D. E., Pellegrini, E. W., Ardizzone, L., et al. 2022, MNRAS, 512, 617

Kannan, R., Garaldi, E., Smith, A., et al. 2022, MNRAS, 511, 4005

Kannappan, S. J. & Wei, L. H. 2008, in American Institute of Physics Conference Series, Vol. 1035, The Evolution of Galaxies Through the Neutral Hydrogen Window, ed. R. Minchin & E. Momjian, 163–168

Katz, H., Garel, T., Rosdahl, J., et al. 2022, MNRAS, stac1437

Katz, H., Kimm, T., Ellis, R. S., Devriendt, J., & Slyz, A. 2022a, arXiv e-prints, arXiv:2207.04751

Katz, H., Rosdahl, J., Kimm, T., et al. 2022b, MNRAS, 510, 5603

Kehrig, C., Vílchez, J. M., Guerrero, M. A., et al. 2018, MNRAS, 480, 1081

Kehrig, C., Vílchez, J. M., Pérez-Montero, E., et al. 2015, ApJL, 801, L28

Kehrig, C., Vílchez, J. M., Pérez-Montero, E., et al. 2016, MNRAS, 459, 2992

Kehrig, C., Vílchez, J. M., Telles, E., Cuisinier, F., & Pérez-Montero, E. 2006, A&A, 457, 477

Kennicutt, Robert C., J. 1998, Annual Review of Astronomy and Astrophysics, 36, 189

Kennicutt, Robert C., J., Bresolin, F., Bomans, D. J., Bothun, G. D., & Thompson, I. B. 1995, AJ, 109, 594

Kennicutt, Robert C., J., Hao, C.-N., Calzetti, D., et al. 2009, ApJ, 703, 1672

Kewley, L. J., Nicholls, D. C., & Sutherland, R. S. 2019, Annual Review of Astronomy and Astrophysics, 57, 511

Kinman, T. D. & Davidson, K. 1981, ApJ, 243, 127

Kobulnicky, H. A., Dickey, J. M., Sargent, A. I., Hogg, D. E., & Conti, P. S. 1995, AJ, 110, 116

Kobulnicky, H. A. & Skillman, E. D. 1997, ApJ, 489, 636

Köppen, J., Weidner, C., & Kroupa, P. 2007, MNRAS, 375, 673

Kramer, C., Cubick, M., Röllig, M., et al. 2008, A&A, 477, 547

Krumholz, M. R. 2012, ApJ, 759, 9

Krumholz, M. R. & Dekel, A. 2012, ApJ, 753, 16

Kumari, N., Amorín, R., Pérez-Montero, E., Vílchez, J., & Maiolino, R. 2021, MNRAS, 508, 1084

Kunth, D., Leitherer, C., Mas-Hesse, J. M., Östlin, G., & Petrosian, A. 2003, ApJ, 597, 263

Kunth, D., Lequeux, J., Sargent, W. L. W., & Viallefond, F. 1994, A&A, 282, 709

Kunth, D. & Östlin, G. 2000, Astronomy and Astrophysics Reviews, 10, 1

Kunth, D. & Sargent, W. L. W. 1986, ApJ, 300, 496

Lacerda, E. A. D., Cid Fernandes, R., Couto, G. S., et al. 2018, MNRAS, 474, 3727

Lagos, P., Demarco, R., Papaderos, P., et al. 2016, MNRAS, 456, 1549

Lagos, P. & Papaderos, P. 2013, Advances in Astronomy, 2013, 631943

Lambert-Huyghe, A., Madden, S. C., Lebouteiller, V., et al. 2022, A&A, 666, A112

Larson, R. B. 1981, MNRAS, 194, 809

Laseter, I. H., Barger, A. J., Cowie, L. L., & Taylor, A. J. 2022, ApJ, 935, 150

Le Petit, F., Nehmé, C., Le Bourlot, J., & Roueff, E. 2006, ApJS, 164, 506

Lebouteiller, V., Cormier, D., Madden, S. C., et al. 2019, A&A, 632, A106, arXiv: 1911.03280

Lebouteiller, V., Heap, S., Hubeny, I., & Kunth, D. 2013, A&A, 553, A16

Lebouteiller, V., Péquignot, D., Cormier, D., et al. 2017, A&A, 602, A45

Lebouteiller, V. & Ramambason, L. 2022a, MULTIGRIS: Multicomponent probabilistic grid search, Astrophysics Source Code Library, record ascl:2207.001

Lebouteiller, V. & Ramambason, L. 2022b, A&A, 667, A34

Lee, H., Skillman, E. D., Cannon, J. M., et al. 2006, ApJ, 647, 970

Lee, M. Y., Madden, S. C., Le Petit, F., et al. 2019, A&A, 628, A113

Lee, M. Y., Madden, S. C., Lebouteiller, V., et al. 2016, A&A, 596, A85

Lehmer, B. D., Ferrell, A. P., Doore, K., et al. 2020, ApJS, 248, 31

Leitet, E., Bergvall, N., Hayes, M., Linné, S., & Zackrisson, E. 2013, A&A, 553, A106

Leitherer, C., Hernandez, S., Lee, J. C., & Oey, M. S. 2016, ApJ, 823, 64

Leitherer, C., Schaerer, D., Goldader, J. D., et al. 1999, ApJS, 123, 3

Lequeux, J. 2005, The Interstellar Medium

Lequeux, J., Peimbert, M., Rayo, J. F., Serrano, A., & Torres-Peimbert, S. 1979, A&A, 80, 155

Leroy, A., Bolatto, A. D., Simon, J. D., & Blitz, L. 2005, ApJ, 625, 763

Leroy, A., Cannon, J., Walter, F., Bolatto, A., & Weiss, A. 2007, ApJ, 663, 990

Lewis, J. S. W., Ocvirk, P., Aubert, D., et al. 2020, MNRAS, 496, 4342

Li, Y., Gu, M. F., Yajima, H., Zhu, Q., & Maji, M. 2020, MNRAS, 494, 1919

Lian, J., Thomas, D., & Maraston, C. 2018, MNRAS, 481, 4000

Liszt, H. & Lucas, R. 1996, A&A, 314, 917

Liu, D., Daddi, E., Schinnerer, E., et al. 2021, ApJ, 909, 56

Lucas, R. & Liszt, H. 1996, A&A, 307, 237

Mac Low, M.-M. & Klessen, R. S. 2004, Reviews of Modern Physics, 76, 125

Madden, S. C., Cormier, D., Hony, S., et al. 2020, A&A, 643, A141

Madden, S. C., Poglitsch, A., Geis, N., Stacey, G. J., & Townes, C. H. 1997, ApJ, 483, 200

Madden, S. C., Rémy-Ruyer, A., Galametz, M., et al. 2013, Publications of the ASP, 125, 600

Magnelli, B., Saintonge, A., Lutz, D., et al. 2012, A&A, 548, A22

Maillard, V., Bron, E., & Le Petit, F. 2021, A&A, 656, A65

Maiolino, R. & Mannucci, F. 2019, Astronomy and Astrophysics Reviews, 27, 3

Maji, M., Verhamme, A., Rosdahl, J., et al. 2022, A&A, 663, A66

Mannucci, F., Belfiore, F., Curti, M., et al. 2021, MNRAS, 508, 1582

Mannucci, F., Cresci, G., Maiolino, R., Marconi, A., & Gnerucci, A. 2010, MNRAS, 408, 2115

Marsh, K. A., Kirk, J. M., André, P., et al. 2016, MNRAS, 459, 342

Matteucci, F. & Recchi, S. 2001, in Cosmic evolution, ed. E. Vangioni-Flam, R. Ferlet, & M. Lemoine, 267

Mauerhofer, V., Verhamme, A., Blaizot, J., et al. 2021, A&A, 646, A80

McElroy, D., Walsh, C., Markwick, A. J., et al. 2013, A&A, 550, A36

McKee, C. F. & Ostriker, J. P. 1977, ApJ, 218, 148

McQuinn, K. B. W., Skillman, E. D., Cannon, J. M., et al. 2010, ApJ, 721, 297

Mengersen, K. L. & Tweedie, R. L. 1996, The Annals of Statistics, 24, 101

Meyer, D. M., Jura, M., & Cardelli, J. A. 1998, ApJ, 493, 222

Meyer, R. A., Walter, F., Cicone, C., et al. 2022, ApJ, 927, 152

Michel-Dansac, L., Blaizot, J., Garel, T., et al. 2020, A&A, 635, A154

Mingozzi, M., Belfiore, F., Cresci, G., et al. 2020, A&A, 636, A42

Mirabel, I. F., Dijkstra, M., Laurent, P., Loeb, A., & Pritchard, J. R. 2011, A&A, 528, A149

Mison, S., Simons, M., & J.L., B. 2013, Geophysical Journal International, 194, 1701

Mitsuda, K., Inoue, H., Koyama, K., et al. 1984, Publications of the ASJ, 36, 741

Morisset, C. 2011, Cloudy_3D: Quick Pseudo-3D Photoionization Code, Astrophysics Source Code Library, record ascl:1103.015

Morisset, C. 2013, pyCloudy: Tools to manage astronomical Cloudy photoionization code, Astrophysics Source Code Library, record ascl:1304.020

Morisset, C., Delgado-Inglada, G., & Flores-Fajardo, N. 2015, Revista Mexicana de Astronomía y Astrofísica, 51, 103

Morisset, C., Delgado-Inglada, G., Sánchez, S. F., et al. 2016, A&A, 594, A37

Mullman, K. L., Lawler, J. E., Zsargó, J., & Federman, S. R. 1998, ApJ, 500, 1064

Naidu, R. P., Forrest, B., Oesch, P. A., Tran, K.-V. H., & Holden, B. P. 2018, MNRAS, 478, 791

Naidu, R. P., Oesch, P. A., Reddy, N., et al. 2017, ApJ, 847, 12

Naidu, R. P., Oesch, P. A., van Dokkum, P., et al. 2022, ApJL, 940, L14

Nakajima, K., Ellis, R. S., Robertson, B. E., Tang, M., & Stark, D. P. 2020, ApJ, 889, 161

Nakajima, K. & Maiolino, R. 2022, MNRAS, 513, 5134

Nakajima, K. & Ouchi, M. 2014, MNRAS, 442, 900

Nanni, A., Burgarella, D., Theulé, P., Côté, B., & Hirashita, H. 2020, A&A, 641, A168

Nguyen, H., Dawson, J. R., Miville-Deschênes, M. A., et al. 2018, ApJ, 862, 49

Nicholls, D. C., Sutherland, R. S., Dopita, M. A., Kewley, L. J., & Groves, B. A. 2017, 20

Niederhofer, F., Hilker, M., Bastian, N., & Ercolano, B. 2016, A&A, 592, A47

Oey, M. S., Herrera, C. N., Silich, S., et al. 2017, ApJL, 849, L1

Offner, S. S. R., Clark, P. C., Hennebelle, P., et al. 2014, in Protostars and Planets VI, ed. H. Beuther, R. S. Klessen, C. P. Dullemond, & T. Henning, 53

Oliveira, J. M., van Loon, J. T., Sewiło, M., et al. 2019, MNRAS, 490, 3909

Olivier, G. M., Berg, D. A., Chisholm, J., et al. 2022, ApJ, 938, 16

Olsen, K., Greve, T. R., Narayanan, D., et al. 2017, ApJ, 846, 105

Olsen, K. P., Burkhart, B., Mac Low, M.-M., et al. 2021, ApJ, 922, 88

Osterbrock, D. & Ferland, G. 2006, Astrophysics of Gaseous Nebulae and Active Galactic Nuclei, 2nd edn. (Sausalito, California: University Science Books)

Östlin, G., Amram, P., Bergvall, N., et al. 2001, A&A, 374, 800

Östlin, G., Rivera-Thorsen, T. E., Menacho, V., et al. 2021, ApJ, 912, 155

Otí-Floranes, H., Mas-Hesse, J. M., Jiménez-Bailón, E., et al. 2012, A&A, 546, A65

Ott, J., Walter, F., & Brinks, E. 2005, MNRAS, 358, 1453

Pahl, A. J., Shapley, A., Steidel, C. C., Chen, Y., & Reddy, N. A. 2021, MNRAS, 505, 2447

Pahl, A. J., Shapley, A., Steidel, C. C., Reddy, N. A., & Chen, Y. 2022, MNRAS, 516, 2062

Parmar, P. S., Lacy, J. H., & Achtermann, J. M. 1991, ApJL, 372, L25

Pellegrini, E. W., Oey, M. S., Winkler, P. F., et al. 2012, ApJ, 755, 40

Pellegrini, E. W., Rahner, D., Reissl, S., et al. 2020, MNRAS, 496, 339

Péquignot, D. 2008, A&A, 478, 371

Pérez-Montero, E. 2014, MNRAS, 441, 2663

Pérez-Montero, E. & Díaz, Á. I. 2007, MNRAS, 377, 1195

Pérez-Montero, E., García-Benito, R., & Vílchez, J. M. 2019, MNRAS, 483, 3322

Pérez-Montero, E., Kehrig, C., Vílchez, J. M., et al. 2020, A&A, 643, A80

Perez-Montero, E. & Vilchez, J. M. 2009, Monthly Notices of the Royal Astronomical Society, 400, 1721, arXiv: 0908.3069

Pilbratt, G. L., Riedinger, J. R., Passvogel, T., et al. 2010, A&A, 518, L1

Pillepich, A., Springel, V., Nelson, D., et al. 2018, MNRAS, 473, 4077

Pilyugin, L. S. & Thuan, T. X. 2005, ApJ, 631, 231

Pingel, N. M., Dempsey, J., McClure-Griffiths, N. M., et al. 2022, Publications of the Astron. Soc. of Australia, 39, e005

Plat, A., Charlot, S., Bruzual, G., et al. 2019, MNRAS, 490, 978

Poglitsch, A., Krabbe, A., Madden, S. C., et al. 1995, ApJ, 454, 293

Polimera, M. S., Kannappan, S. J., Richardson, C. T., et al. 2022, ApJ, 931, 44

Polles, F. L., Madden, S. C., Lebouteiller, V., et al. 2019, A&A, 622, A119

Ponnada, S., Brorby, M., & Kaaret, P. 2020, MNRAS, 491, 3606

Press, W. H. & Schechter, P. 1974, ApJ, 187, 425

Prestwich, A. H., Jackson, F., Kaaret, P., et al. 2015, ApJ, 812, 166

Prestwich, A. H., Tsantaki, M., Zezas, A., et al. 2013, ApJ, 769, 92

Pérez-Montero, E., García-Benito, R., & Vílchez, J. M. 2019, Monthly Notices of the Royal Astronomical Society, 483, 3322

Rafelski, M., Wolfe, A. M., Prochaska, J. X., Neeleman, M., & Mendez, A. J. 2012, ApJ, 755, 89

Ramambason, L., Lebouteiller, V., Bik, A., et al. 2022, A&A, 667, A35

Ramambason, L., Schaerer, D., Stasińska, G., et al. 2020, A&A, 644, A21

Ramos Padilla, A. F., Wang, L., Ploeckinger, S., van der Tak, F. F. S., & Trager, S. C. 2021, A&A, 645, A133

Ramos Padilla, A. F., Wang, L., van der Tak, F. F. S., & Trager, S. 2022, arXiv e-prints, arXiv:2205.11955

Reach, W. T., Heiles, C., & Bernard, J.-P. 2015, ApJ, 811, 118

Reddy, N. A., Steidel, C. C., Pettini, M., Bogosavljević, M., & Shapley, A. E. 2016, ApJ, 828, 108

Rees, M. J. 1978a, , 17, 371

Rees, M. J. 1978b, Nature, 275, 35

Reines, A. E., Sivakoff, G. R., Johnson, K. E., & Brogan, C. L. 2011, Nature, 470, 66

Relaño, M., De Looze, I., Kennicutt, R. C., et al. 2018, A&A, 613, A43

Remy, Q., Grenier, I. A., Marshall, D. J., & Casandjian, J. M. 2018, A&A, 616, A71

Rémy-Ruyer, A. 2013, PhD thesis

Rey, M. P., Pontzen, A., Agertz, O., et al. 2020, MNRAS, 497, 1508

Richardson, C. T., Allen, J. T., Baldwin, J. A., Hewett, P. C., & Ferland, G. J. 2014, Monthly Notices of the Royal Astronomical Society, 437, 2376

Richardson, C. T., Allen, J. T., Baldwin, J. A., et al. 2016, Mon. Not. R. Astron. Soc., 458, 988

Richardson, C. T., Simpson, C., Polimera, M. S., et al. 2022, ApJ, 927, 165

Rickards Vaught, R. J., Sandstrom, K. M., & Hunt, L. K. 2021, ApJL, 911, L17

Rivera-Thorsen, T. E., Dahle, H., Chisholm, J., et al. 2019, Science, 366, 738

Robert, C. P. 2015, arXiv e-prints, arXiv:1504.01896

Robertson, B. E., Ellis, R. S., Dunlop, J. S., McLure, R. J., & Stark, D. P. 2010, Nature, 468, 49

Robertson, B. E., Ellis, R. S., Furlanetto, S. R., & Dunlop, J. S. 2015, ApJL, 802, L19

Robertson, B. E., Furlanetto, S. R., Schneider, E., et al. 2013, ApJ, 768, 71

Rosdahl, J., Blaizot, J., Katz, H., et al. 2022, MNRAS, 515, 2386

Rubio, M., Elmegreen, B. G., Hunter, D. A., et al. 2015, Nature, 525, 218

Rémy-Ruyer, A., Madden, S. C., Galliano, F., et al. 2014, A&A, 563, A31

Rémy-Ruyer, A., Madden, S. C., Galliano, F., et al. 2015, A&A, 582, A121

Sage, L. J., Salzer, J. J., Loose, H. H., & Henkel, C. 1992, A&A, 265, 19

Saha, K., Tandon, S. N., Simmonds, C., et al. 2020, Nature Astronomy, 4, 1185

Saldana-Lopez, A., Schaerer, D., Chisholm, J., et al. 2022, A&A, 663, A59

Salvatier, J., Wieckiâ, T. V., & Fonnesbeck, C. 2016, PyMC3: Python probabilistic programming framework, Astrophysics Source Code Library, record ascl:1610.016

Sanders, R. L., Jones, T., Shapley, A. E., et al. 2020, ApJL, 888, L11

Sanders, R. L., Shapley, A. E., Jones, T., et al. 2021, ApJ, 914, 19

Sandstrom, K. M., Bolatto, A. D., Draine, B. T., Bot, C., & Stanimirović, S. 2010, ApJ, 715, 701

Sandstrom, K. M., Leroy, A. K., Walter, F., et al. 2013, ApJ, 777, 5

Satyapal, S., Kamal, L., Cann, J. M., Secrest, N. J., & Abel, N. P. 2021, ApJ, 906, 35

Savage, B. D. & Sembach, K. R. 1996, Annual Review of Astronomy and Astrophysics, 34, 279

Saxena, A., Pentericci, L., Schaerer, D., et al. 2020, MNRAS, 496, 3796

Schaerer, D., Fragos, T., & Izotov, Y. I. 2019, A&A, 622, L10

Schaerer, D., Marques-Chaves, R., Barrufet, L., et al. 2022, A&A, 665, L4

Schaye, J., Crain, R. A., Bower, R. G., et al. 2015, MNRAS, 446, 521

Schruba, A., Leroy, A. K., Kruijssen, J. M. D., et al. 2017, ApJ, 835, 278

Schruba, A., Leroy, A. K., Walter, F., et al. 2012, AJ, 143, 138

Scoville, N. Z., Yun, M. S., Clemens, D. P., Sanders, D. B., & Waller, W. H. 1987, ApJS, 63, 821

Shapley, A. E., Steidel, C. C., Strom, A. L., et al. 2016, ApJL, 826, L24

Shi, Y., Wang, J., Zhang, Z.-Y., et al. 2015, ApJL, 804, L11

Shi, Y., Wang, J., Zhang, Z.-Y., et al. 2020, ApJ, 892, 147

Simmonds, C., Schaerer, D., & Verhamme, A. 2021, A&A, 656, A127

Smith, B. D., Regan, J. A., Downes, T. P., et al. 2018, Monthly Notices of the Royal Astronomical Society, 480, 3762

Snow, T. P., Destree, J. D., & Jensen, A. G. 2007, ApJ, 655, 285

Solomon, P. M., Downes, D., Radford, S. J. E., & Barrett, J. W. 1997, ApJ, 478, 144

Solomon, P. M., Rivolo, A. R., Barrett, J., & Yahil, A. 1987, ApJ, 319, 730

Sommovigo, L., Ferrara, A., Pallottini, A., et al. 2020, MNRAS, 497, 956

Stanway, E. R. & Eldridge, J. J. 2019, A&A, 621, A105

Stark, D. V., Masters, K. L., Avila-Reese, V., et al. 2021, MNRAS, 503, 1345

Stasińska, G., Izotov, Y., Morisset, C., & Guseva, N. 2015, A&A, 576, A83

Steidel, C. C., Bogosavljević, M., Shapley, A. E., et al. 2018, ApJ, 869, 123

Strömgren, B. 1939, ApJ, 89, 526

Tacconi, L. J., Genzel, R., Saintonge, A., et al. 2018, ApJ, 853, 179

Tacconi, L. J., Genzel, R., & Sternberg, A. 2020, Annual Review of Astronomy and Astrophysics, 58, 157

Tadaki, K.-i., Tsujita, A., Tamura, Y., et al. 2022, Publications of the ASJ, 74, L9

Tanvir, N. R., Fynbo, J. P. U., de Ugarte Postigo, A., et al. 2019, Monthly Notices of the Royal Astronomical Society, 483, 5380

Taylor, C. L., Kobulnicky, H. A., & Skillman, E. D. 1998, AJ, 116, 2746

Tchernyshyov, K., Meixner, M., Seale, J., et al. 2015, ApJ, 811, 78

Telfer, R. C., Kriss, G. A., Zheng, W., Davidsen, A. F., & Tytler, D. 2002, ApJ, 579, 500

The LUVOIR Team. 2019, arXiv e-prints, arXiv:1912.06219

Thomas, A. D., Dopita, M. A., Kewley, L. J., et al. 2018, ApJ, 856, 89

Thomas, A. D., Kewley, L. J., Dopita, M. A., et al. 2019, ApJ, 874, 100

Thronson, Harley A., J. & Bally, J. 1987, in NASA Conference Publication, Vol. 2466, NASA Conference Publication, ed. C. J. Lonsdale Persson, 267–270

Thuan, T. X., Bauer, F. E., Papaderos, P., & Izotov, Y. I. 2004, ApJ, 606, 213

Tokuda, K., Kondo, H., Ohno, T., et al. 2021, ApJ, 922, 171

Tokuda, K., Minami, T., Fukui, Y., et al. 2022, ApJ, 933, 20

Tremonti, C. A., Heckman, T. M., Kauffmann, G., et al. 2004, ApJ, 613, 898

Turner, J. L., Beck, S. C., Benford, D. J., et al. 2015, Nature, 519, 331

Ucci, G., Ferrara, A., Gallerani, S., & Pallottini, A. 2017, Mon. Not. R. Astron. Soc., 465, 1144

Ucci, G., Ferrara, A., Gallerani, S., et al. 2019, Monthly Notices of the Royal Astronomical Society, 483, 1295

Umeda, H., Ouchi, M., Nakajima, K., et al. 2022, ApJ, 930, 37

Vale Asari, N., Couto, G. S., Cid Fernandes, R., et al. 2019, MNRAS, 489, 4721

Vanzella, E., Caminha, G. B., Calura, F., et al. 2020, MNRAS, 491, 1093

Vanzella, E., de Barros, S., Castellano, M., et al. 2015, A&A, 576, A116

Vanzella, E., de Barros, S., Vasei, K., et al. 2016, ApJ, 825, 41

Vanzella, E., Nonino, M., Cupani, G., et al. 2018, MNRAS, 476, L15

Vehtari, A., Gelman, A., & Gabry, J. 2017, Statistics and Computing, 27, 1413

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P.-C. 2019, arXiv e-prints, arXiv:1903.08008

Verhamme, A., Orlitová, I., Schaerer, D., et al. 2017, A&A, 597, A13

Verhamme, A., Orlitová, I., Schaerer, D., & Hayes, M. 2015, A&A, 578, A7

Vielfaure, J. B., Vergani, S. D., Japelj, J., et al. 2020, A&A, 641, A30

Villa-Vélez, J. A., Buat, V., Theulé, P., Boquien, M., & Burgarella, D. 2021, A&A, 654, A153

Vizgan, D., Greve, T. R., Olsen, K. P., et al. 2022, ApJ, 929, 92

Vrugt, J. A., Ter Braak, C. J. F., Clark, M. P., Hyman, J. M., & Robinson, B. A. 2008, Water Resources Research, 44, W00B09

Wakelam, V., Herbst, E., Loison, J. C., et al. 2012, ApJS, 199, 21

Walker, A. R. 2012, Astrophysics and Space Science, 341, 43

Wang, B., Heckman, T. M., Amorín, R., et al. 2021, ApJ, 916, 3

Wang, B., Heckman, T. M., Leitherer, C., et al. 2019, ApJ, 885, 57

Watanabe, S. & Opper, M. 2012, Journal of machine learning research, 11, 12

Weinberger, R., Springel, V., Hernquist, L., et al. 2017, MNRAS, 465, 3291

Weingartner, J. C. & Draine, B. T. 2001a, ApJ, 548, 296

Weingartner, J. C. & Draine, B. T. 2001b, ApJ, 548, 296

Wofford, A., Vidal-García, A., Feltre, A., et al. 2021, MNRAS, 500, 2908

Wolfire, M. G., Hollenbach, D., & McKee, C. F. 2010, ApJ, 716, 1191

Wolfire, M. G., Vallini, L., & Chevance, M. 2022, Annual Review of Astronomy and Astrophysics, 60, 247

Wong, T., Oudshoorn, L., Sofovich, E., et al. 2022, ApJ, 932, 47

Wu, Y., Charmandaris, V., Hao, L., et al. 2006, ApJ, 639, 157

Xiao, L., Stanway, E. R., & Eldridge, J. J. 2018, MNRAS, 477, 904

Xu, X., Henry, A., Heckman, T., et al. 2022, ApJ, 933, 202

Young, J. S., Xie, S., Tacconi, L., et al. 1995, ApJS, 98, 219

Youngblood, A. J. & Hunter, D. A. 1999, ApJ, 519, 55

Zackrisson, E., Inoue, A. K., & Jensen, H. 2013, ApJ, 777, 39

Zanella, A., Daddi, E., Magdis, G., et al. 2018, MNRAS, 481, 1976

Zastrow, J., Oey, M. S., Veilleux, S., & McDonald, M. 2013, ApJ, 779, 76

Zastrow, J., Oey, M. S., Veilleux, S., McDonald, M., & Martin, C. L. 2011, ApJ, 741, L17

Zavala, J. A., Casey, C. M., Spilker, J., et al. 2022, ApJ, 933, 242

Zhang, K., Yan, R., Bundy, K., et al. 2017, MNRAS, 466, 3217

Zhou, L., Shi, Y., Zhang, Z.-Y., & Wang, J. 2021, A&A, 653, L10