A&A, 695, A31 (2025)
https://doi.org/10.1051/0004-6361/202452103

Astronomy & Astrophysics

Recovering the properties of the interstellar medium through integrated spectroscopy: Application to the $z \sim 0$ ECO volume-limited star-forming galaxy sample

V. Lebouteiller[1,⋆], C. T. Richardson[2], M. S. Polimera[3,4], D. S. Carr[4], Z. L. Hutchens[4,5], S. J. Kannappan[4], L. Ramambason[6], A. J. Moffett[7], M. Varese[1], and S. C. Madden[1]

[1] Université Paris-Saclay, Université Paris Cité, CEA, CNRS, AIM, 91191 Gif-sur-Yvette, France
[2] Elon University, 100 Campus Drive, Elon, NC 27278, USA
[3] Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA
[4] Department of Physics & Astronomy, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
[5] Department of Physics & Astronomy, University of North Carolina Asheville, Asheville, NC 28804, USA
[6] Institut fur Theoretische Astrophysik, Zentrum für Astronomie, Universität Heidelberg, Albert-Ueberle-Str. 2, 69120 Heidelberg, Germany
[7] Department of Physics and Astronomy, University of North Georgia, 3820 Mundy Mill Road, Oakwood, GA 30566, USA

Received 3 September 2024 / Accepted 24 January 2025

ABSTRACT

Context. Deriving physical parameters from integrated galaxy spectra is paramount to interpret the cosmic evolution of the star formation, chemical enrichment, and energetic processes at play. Previous studies have highlighted the power of interstellar medium tracers but also the associated complexities that can be captured only through sophisticated modeling approaches.
Aims. We developed modeling techniques to characterize the ionized gas properties in the subset of 2052 star-forming galaxies from the volume-limited, dwarf-dominated, $z \sim 0$ ECO catalog (stellar mass range $M_* \sim 10^{8-11}$ $M_\odot$). Our study sheds light on the internal distribution and average values of parameters such as the metallicity, ionization parameter, and electron density within galaxies.
Methods. We used the MULTIGRIS statistical framework to evaluate the performance of various models using strong lines as constraints. The reference model involves physical parameters distributed as power laws with free parameter boundaries. Specifically, we used combinations of 1D photoionization models (i.e., considering the propagation of radiation toward a single cloud) to match optical H II region lines, in order to provide probability density functions of the inferred parameters.
Results. The inference predicts nonuniform physical conditions within galaxies. The integrated spectra of most galaxies are dominated by relatively low-excitation gas with a metallicity around 0.3 $Z_\odot$. Using the average metallicity in galaxies, we provide a new fit to the mass-metallicity relationship which is in line with direct abundance method determinations from the low-metallicity calibrated range up to high-metallicity stacks. The average metallicity shows a weakly bimodal distribution which may be due to external (e.g., refueling of non-cluster early-type galaxies) or internal processes (higher star-formation efficiency in metal-rich regions). The specific line set used for inference affects the results and we identify potential issues with the use of the [S II] line doublet.
Conclusions. Complex modeling approaches may capture diverse physical conditions within galaxies but require robust statistical frameworks. Such approaches are limited by the inherent 1D model database as well as caveats regarding the gas geometry. Our results highlight, however, the possibility to extract useful and significant information from integrated spectra.

Key words. methods: statistical – ISM: general – ISM: structure – galaxies: fundamental parameters – galaxies: ISM

1. Introduction

Spectroscopic diagnostics of the interstellar medium (ISM) in galaxies hold tremendous diagnostic power, for example, on the star-formation rate (SFR), gas masses, or the fraction of ionizing radiation due to active galactic nuclei (AGNs). Yet we do not know precisely how to interpret spatially unresolved spectra by accounting for and modeling the complex mechanisms that produce the observed, integrated, emission. As we accumulate more high-z observations as well as wide-field and all-sky spectroscopy lacking spatial resolution, it becomes urgent to design a modeling framework to derive robust and reliable physical parameter distributions highlighting galaxy evolution mechanisms. Furthermore, integrated observations may encompass enough information to actually recover such distributions, thereby enabling some of the power of integral-field spectroscopy (IFS) without actually performing it.

Historically, long-slit spectroscopy or integrated spectroscopy have been used to probe some kinds of average physical conditions in galaxies or to identify dominant excitation sources. Integrated line ratios are, for instance, commonly interpreted using 1D photoionization grids (i.e., assuming a single cloud illuminated by a unique radiation source – or possibly several co-spatial radiation sources – with a plane-parallel or spherical geometry) in order to trace the gas electron density and pressure, the ionization parameter (U), the metallicity (Z), SFR, or excitation mechanism diagnostics (see, e.g., Kewley et al. 2019). Such a hypothesis may reflect either a single "representative" H II region or, equivalently, an ensemble of H II regions with similar properties (see illustration in Fig. 1).

⋆ Corresponding author; vianney.lebouteiller@cnrs.fr

It is generally difficult to evaluate the reliability and actual meaning of the quantities derived thusly and it is therefore crucial to understand potential biases and selection effects due to the hypothesis of a single emission component versus the combined emission of various components and physical mechanisms (see, e.g., Sorba & Sawicki 2015, 2018 for stellar mass determinations). Star-forming galaxies are indeed known to include the following:

- Gas that follows a density distribution related to turbulence, self-gravitation, and rotational support (e.g., Khullar et al. 2021);
- Metallicity variations within galaxies in the form of gradients but also higher order variations (e.g., Poetrodjojo et al. 2018; Williams et al. 2022; Nakajima et al. 2024);
- Molecular clouds, some associated with recent star formation (e.g., Tacconi et al. 2020; Saintonge & Catinella 2022);
- A collection of H II regions following some luminosity function (e.g., Santoro et al. 2022), with some of them leaking ionizing photons (e.g., Della Bruna et al. 2021);
- Stellar age gradients (e.g., Riggs et al. 2024); and
- Additional excitation sources such as Wolf-Rayet stars, high-mass X-ray binaries, or AGNs, with the stellar clusters and AGN that may or not be co-spatial, resulting in coincident or non-coincident geometries (Richardson et al. 2022).

In some rare cases (e.g., guided by imagery or spatially resolved spectroscopy), it is possible to describe a specific star-forming region or even a specific galaxy as a single dominant stellar cluster surrounded by ISM shells, which enables the use of full 3D or pseudo 3D models (with photoionization and chemistry) with arbitrary geometries (e.g., M^3; Jin et al. 2022, PyCloudy; Morisset 2013; Fitzgerald et al. 2020). Three-dimensional Monte-Carlo radiative transfer codes are also useful in that they can handle potentially complex geometries and density structures (e.g., Baes et al. 2003, 2011; De Looze et al. 2014) with promising avenues toward a fully self-consistent 3D model (Romero et al. 2023). Despite these possible improvements, the geometry is never a free parameter in the models that involve chemistry and photoionization. It must also be added that radiation magneto-hydrodynamical simulations are now able to solve large chemical networks (e.g., Katz et al. 2022; Katz 2022), while simulation databases are increasingly available (e.g., Katz et al. 2023) but the comparison to observations remains difficult due to the generally restricted parameter space (e.g., regarding cosmic ray ionization rates or the dust-to-gas mass ratio).

Due to the difficulty in designing 3D models, single 1D models are frequently used, often in conjunction with modern statistical frameworks (e.g., `BEAGLE`; Chevallard & Charlot 2016, `CIGALE`; Burgarella et al. 2005; Boquien et al. 2019; Yang et al. 2022). It should be noted that 1D models can be somewhat optimized by considering different geometries (e.g., spherical geometry with thin shells or filled spheres; e.g., Stasińska et al. 2015), by including incomplete shells (matter- and density-bounded regions; e.g., Péquignot 2008; Cormier et al. 2015; Ramambason et al. 2020, 2022), or by accounting for the flux-averaged integrated emission of evolving H II regions (e.g., Dopita et al. 2006a; Groves et al. 2008; Pellegrini et al. 2019).

Another avenue consists in combining 1D models, either representing a few "sectors" around stellar clusters (e.g., Péquignot 2008; Cormier et al. 2015; Lebouteiller et al. 2017; Cormier et al. 2019; Madden et al. 2020; Ramambason et al. 2022) or statistical distributions of many emitting components within a galaxy (e.g., Richardson et al. 2014, 2016, 2019; Lebouteiller & Ramambason 2022; Ramambason et al. 2024; Marconi et al. 2024; Varese et al. 2025; Morisset et al. 2025). Such combinations provide useful alternatives to 3D models as long as projection effects are not an important issue.

Classically, the approach to model an integrated galaxy is often driven by the availability of tracers. In general, one prefers the simplest possible model (i.e., the smallest possible set of free parameters) that matches available observations, while more complex configurations (such as combinations of 1D models) are usually introduced out of necessity. This raises an important phenomenological question as to whether models should consider the following:

- An optimal model "architecture" (i.e., choice of physical and geometrical parameters) adapted to the data and preventing too much overfitting. The drawback is that different physical descriptions of the galaxy are solely considered based on what tracers are available, resulting in "representative" models that are often difficult to interpret.
- A model architecture driven by a – possibly complex – physical description of the galaxy, with the most likely parameter values inferred from available observations. This may result in weakly constrained model parameters when few tracers are available but the model itself remains identical with additional tracers.

In the present study, we start from the principle that the model architecture represents a physical object and is expected to be as robust as possible against the set of available tracers. For this we relied on combinations of 1D models ("topological models" as pioneered in Péquignot et al. 2002; Péquignot 2008), as they enable a high enough level of complexity that may approach the actual distribution of source and ISM clouds – despite several biases and caveats – while also being easily parameterized. In other words, we consider that the improvements enabled by such combinations compared to single 1D models largely compensate for potential biases. Apart from the physically motivated necessity to include distributions of components and parameters in order to extract specific parameters of interest related to galaxy evolution, we are also interested in actually recovering the intrinsic variation of physical conditions (metallicity, density, etc.), keeping in mind that IFS samples of dwarf galaxies are particularly small and that such indirect methods may provide promising alternatives.

Assuming such a modeling approach, it is essential to construct a reliable framework to compare models and observations. Probabilistic methods are most adapted as they remain useful when the set of tracers changes and/or when parameters are correlated. Full precomputed grids (including all potential parameter combinations) enable brute-force methods with a Bayesian likelihood calculated for every model and they are relatively quick to process large observation sets (e.g., Blanc et al. 2015; Thomas et al. 2016). However, for this work, we relied on on-the-fly Bayesian likelihood calculations within the statistical framework MULTIGRIS[1] (Lebouteiller & Ramambason 2022). MULTIGRIS enables one to account for nuisance variables and is better adapted to a large number of parameters. For completeness, it must be mentioned that neural networks are also increasingly used to match model predictions and observations, especially in the case of a large number of parameters (e.g., Kang et al. 2022; Morisset et al. 2025). Furthermore, new deep learning methods enable model outcomes to be emulated in even faster ways than regular interpolation methods (e.g., Palud et al. 2023).

We focus on a star-forming galaxy sample extracted from the volume-limited Environmental COntext (ECO) survey

[1] `https://gitlab.com/multigris/mgris`

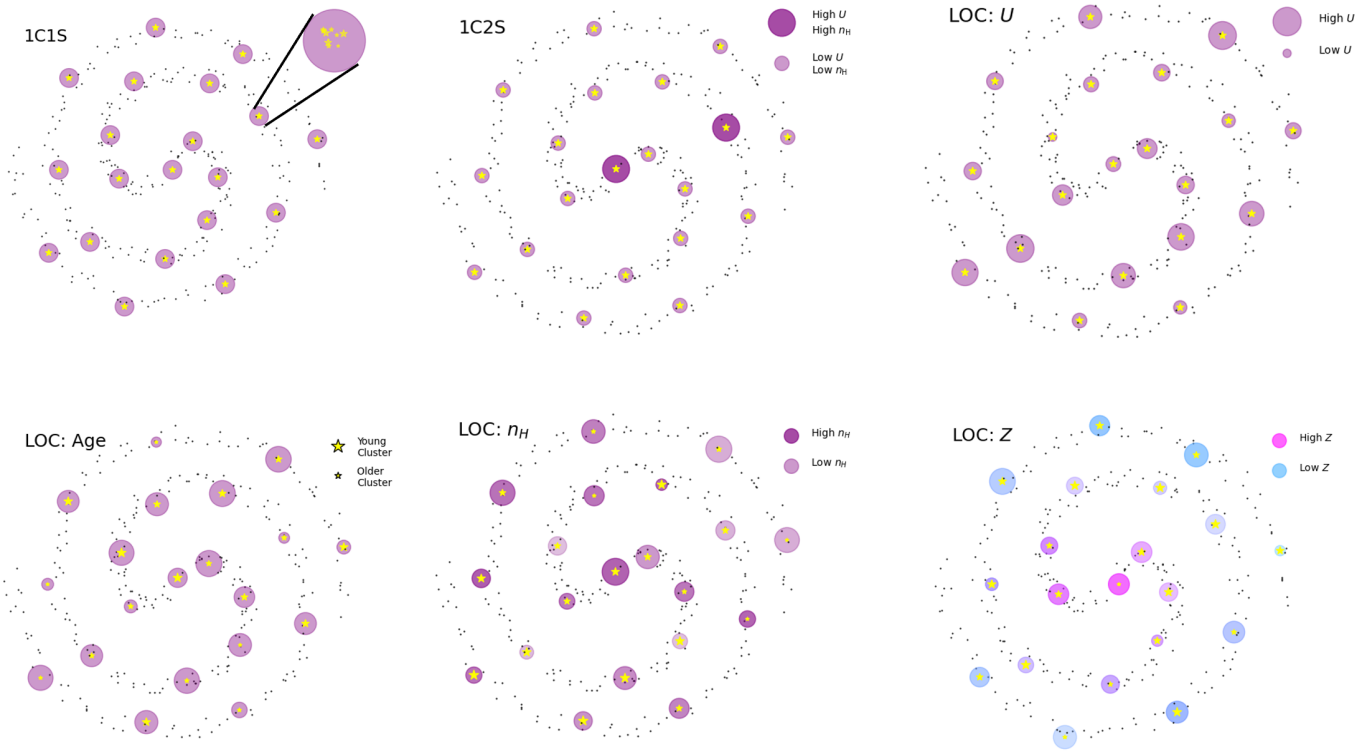

Fig. 1. Illustration of topologies using multicomponent models (1C1S and 1C2S) and integrated distributions (LOC). 1C1S assumes a stellar population described with a single age associated with an ISM component with uniform conditions. 1C2S uses the same stellar population hypothesis as 1C1S but enables two distinct sets of uniform ISM conditions. LOC models gradually consider parameters as power-law distributions.

(Moffett et al. 2015; Polimera et al. 2022; Hutchens et al. 2023), which is complete into the dwarf galaxy regime and for which there should be minimal AGN contamination. Most of the galaxies are dwarf galaxies, but we can study statistically meaningful correlations between parameters and internal variations over a large mass range. Assuming that past versions of massive galaxies resemble low-mass galaxies in the present sample at $z \sim 0$, we may interpret our results as a potential probe of evolutionary pathways. Specifically, we wish to examine the following: 1) the range of physical conditions present in the ECO sample under the most realistic hypothesis of multiple emitting components and how these conditions evolve as a function of statistically averaged quantities (e.g., as a function of metallicity); and 2) the link between physical parameters such as U and Z, and their connection to galaxy evolution parameters such as SFR.

In Sect. 2 we present the framework to assess the various model architectures used to model galaxies. We then present the application of this framework to the ECO star-forming galaxy sample with constraints from optical line spectroscopy (Sect. 3). Results are described in Sect. 4 where we examine the following: 1) the influence of the line set used, and find potential issues with the [S II] lines; 2) the differences between various model architectures, and find that architectures using statistical distributions outperform single 1D models; and 3) the main model fit parameters. Section 5 explores the following: 1) correlations between physical parameters, with strong correlations of all parameters with metallicity Z; and 2) the recovery of internal distributions within galaxies with an emphasis on the metallicity dispersion. We also discuss the potential existence and implication of a metallicity bimodality as well as the mass-metallicity relationship. For the latter, we do find evidence of a significant increase in metallicity between stellar masses $10^{9.5}$–10^{10} $M_\odot$.

2. Modeling framework

2.1. Definition of model architectures (topological models)

Photoionization models are used to describe parameters related to the sources (spectral energy distribution, luminosity), the ISM (density n, chemical composition, distance from source, etc.), and the gas excitation conditions (ionization parameter U). The simplest approach to model a full galaxy considers a single "virtual" stellar cluster representing all clusters in the galaxy with spherically symmetric ISM conditions. This single 1D model may be either interpreted as representing 1) the full galaxy with all excitation sources being co-spatial, resulting in "effective" (or "representative") galaxy-wide parameters, or 2) a collection of strictly identical 1D components (clusters surrounded with ISM) whose total luminosity amounts to that of the galaxy. Both interpretations are equivalent as long as the radiation transfer is controlled by the absorption of UV ionizing photons by the gas (classical H II regions hypothesis described by the Strömgrem sphere assumption; Osterbrock & Ferland 2006). We show in Figure 1 (top left) an illustration of such models to mimic a galaxy's emitting components.

We may then consider a linear combination of two or few "components", with each component representing a single 1D photoionization model. The combination may describe either 1) several ISM components surrounding a single stellar cluster (i.e., well adapted to single H II regions with relatively dense and diffuse "sectors" around the young massive cluster; see, e.g., Cormier et al. 2019; Ramambason et al. 2022), 2) several stellar clusters each surrounded with identical ISM conditions (i.e., well adapted to the case of young SF regions and old stellar populations), or 3) any combination of the above. A critical caveat is that the linear combination assumes that the components are independent. Consequently, radiation escaping from

one component does not affect other components. We will generally refer to these relatively simple architectures of one or a few 1D models as multicomponent models or "xCyS", for x clusters associated with y ISM shell components around each cluster.

As the number of components increases, it becomes necessary to tie them through a statistical distribution described by specific hyperparameters. This is both to keep a manageable number of free parameters as well as to consider a physically meaningful distribution. This is the motivation behind the locally optimally emitted cloud (LOC; Ferguson et al. 1997; Richardson et al. 2014) hypothesis that the observed emission is the result of strong selection effects due to the fact that some emission lines are brighter under certain conditions. Therefore, we may consider a large number of clouds whose model parameters (e.g., density, stellar cluster age) are distributed as power or normal laws defined and integrated between either fixed or free boundaries. We will generally refer to these more complex model architectures as "integrated" distributions or simply "LOC". Progressively more complex LOC models are illustrated in Figure 1.

The LOC equation for an extensive observable L (e.g., a line flux) for a series of parameter sets $\mathbf{p} = (p_0 \dots p_n)$ gives

$$L_{\rm tot} = \sum_{\mathbf{p_{min}}}^{\mathbf{p_{max}}} \prod_{i=0}^{n} \Phi(p_i) I(\mathbf{p}) \Delta(p_i), \tag{1}$$

with $I(\mathbf{p})$ the observable for a given parameter set, $\Delta(p_i)$ the grid parameter interval, and $\Phi(p_i)$ the weight associated with the parameter p_i according to some statistical distribution (e.g., power law). For instance, the weight for a power-law distribution of some parameter p would be

$$\Phi(p) = \left\{ \begin{array}{ll} 10^{\alpha_p} & \text{if } p \in [p_{\rm min}, p_{\rm max}] \\ 0 & \end{array} \right\}. \tag{2}$$

The weight for a parameter that does not follow any particular distribution and that is described instead by a single value $p_{\rm val}$ would be defined as

$$\Phi(p) = \delta_p = \left\{ \begin{array}{ll} 1 & \text{if } p = p_{\rm val} \\ 0 & \end{array} \right\}. \tag{3}$$

The combined weight for a parameter set $\mathbf{p}$ is then

$$\Phi(\mathbf{p}) = \prod_{i=0}^{n} \Phi(p_i). \tag{4}$$

The free parameters are either the model parameters themselves ($p_{\rm val}$ for a single valued distribution) or the hyperparameters (e.g., α_p, $p_{\rm min}$, $p_{\rm max}$ for a power-law distribution).

For any parameter distribution considered, the average parameter value is a useful quantity to calculate, e.g., to compare to single 1D model approaches. The average parameter value is defined as

$$p_{\rm avg} = \frac{\sum_{\mathbf{p_{min}}}^{\mathbf{p_{max}}} \Phi(p) p}{\sum_{\mathbf{p_{min}}}^{\mathbf{p_{max}}} \Phi(p)}, \tag{5}$$

where p is in log scale[2].

[2] Since the integration is performed in the logarithmic space for all parameters, it is natural to calculate the LOC average as the average of the logarithmic quantities. For instance, the combination of two models with all parameters being the same except for different densities 10^3 and 10^0 cm^{-3} compared to two models of densities 10^3 and 10^{-1} cm^{-3} should lead to significantly different results which would be better reflected by the average of the log densities rather than the linear ones.

We emphasize that the distribution of components that we may recover using this method does not correspond to the distribution of what actually composes a galaxy but it is instead the luminosity-weighted distribution of the components that contribute to the optical emission line fluxes. For benchmark purposes, we compared the LOC hyperparameters inferred on-the-fly to results obtained with precomputed LOC grids, (i.e., with tabulated hyperparameters). The inference of hyperparameters quickly becomes more efficient compared to pre-tabulated grids when the number of parameters increases and also provides a flexible framework with priors and potential nuisance variables.

2.2. Statistical framework with `MULTIGRIS`

While LOC distributions may be adapted to particular types of galaxies or regions within galaxies, there is often no prior knowledge as to what distributions should be considered. In the following, we propose one potential method to compare various architectures. We use the statistical framework provided by `MULTIGRIS` (Lebouteiller & Ramambason 2022) which performs on-the-fly inference of combination of 1D models through Markov Chain Monte Carlo (MCMC) sampling. In the following, we refer to the model $\mathcal{M}$ as a model "architecture" defined by a certain distribution of parameters. The posterior probability distribution for a given model $\mathcal{M}$ is defined as

$$p(\boldsymbol{\theta}|\boldsymbol{O}, \mathcal{M}) = \frac{p(\boldsymbol{O}|\boldsymbol{\theta}, \mathcal{M}) p(\boldsymbol{\theta}|\mathcal{M})}{p(\boldsymbol{O}|\mathcal{M})}, \tag{6}$$

with $\boldsymbol{O}$ the data, $\boldsymbol{\theta}$ the parameters, $p(\boldsymbol{O}|\boldsymbol{\theta}, \mathcal{M})$ the likelihood, $p(\boldsymbol{\theta}|\mathcal{M})$ the prior probability, and $p(\boldsymbol{O}|\mathcal{M})$ the marginal likelihood.

Several LOC distributions have been implemented in `MULTIGRIS` (power laws, broken power laws, and normal distributions), with the ability to provide priors on the hyperparameters (slope, mean, standard deviation, etc.). Since most ISM models are usually too long to run for individual MCMC draws, the inference relies on the 1D model grid sampling, together with an interpolation method which can be either nearest neighbors or multidimensional linear interpolation. In practice, hyperparameters for LOC models are either continuous (slope α_p for the power law, mean, and standard deviation for a normal distribution) or using linear or nearest neighbor interpolation (boundaries $p_{\rm min,max}$). The first application of LOC distributions with `MULTIGRIS` was presented in Ramambason et al. (2024) where it was required to explain the emission of the CO(1–0) emission in metal-poor galaxies, with the cloud depth in particular described by a broken power-law distribution.

2.3. Comparative and performance metrics

Several metrics are important to consider when evaluating a model. In particular, we are interested in how well the model captures the data ("goodness" of fit), which is well described by the posterior predictive p-value (PPP). The PPP performs statistical tests of many simulated datasets from the model, using the parameters inferred from the observed data. The PPP is then the proportion of these simulated test statistics that are more extreme than the test statistic calculated from the real data, and is defined as

$$p(\boldsymbol{O}_{\rm rep}|\boldsymbol{O}) = \int_{\boldsymbol{\theta}} p(\boldsymbol{O}_{\rm rep}|\boldsymbol{\theta}) p(\boldsymbol{\theta}|\boldsymbol{O}) d\boldsymbol{\theta}, \tag{7}$$

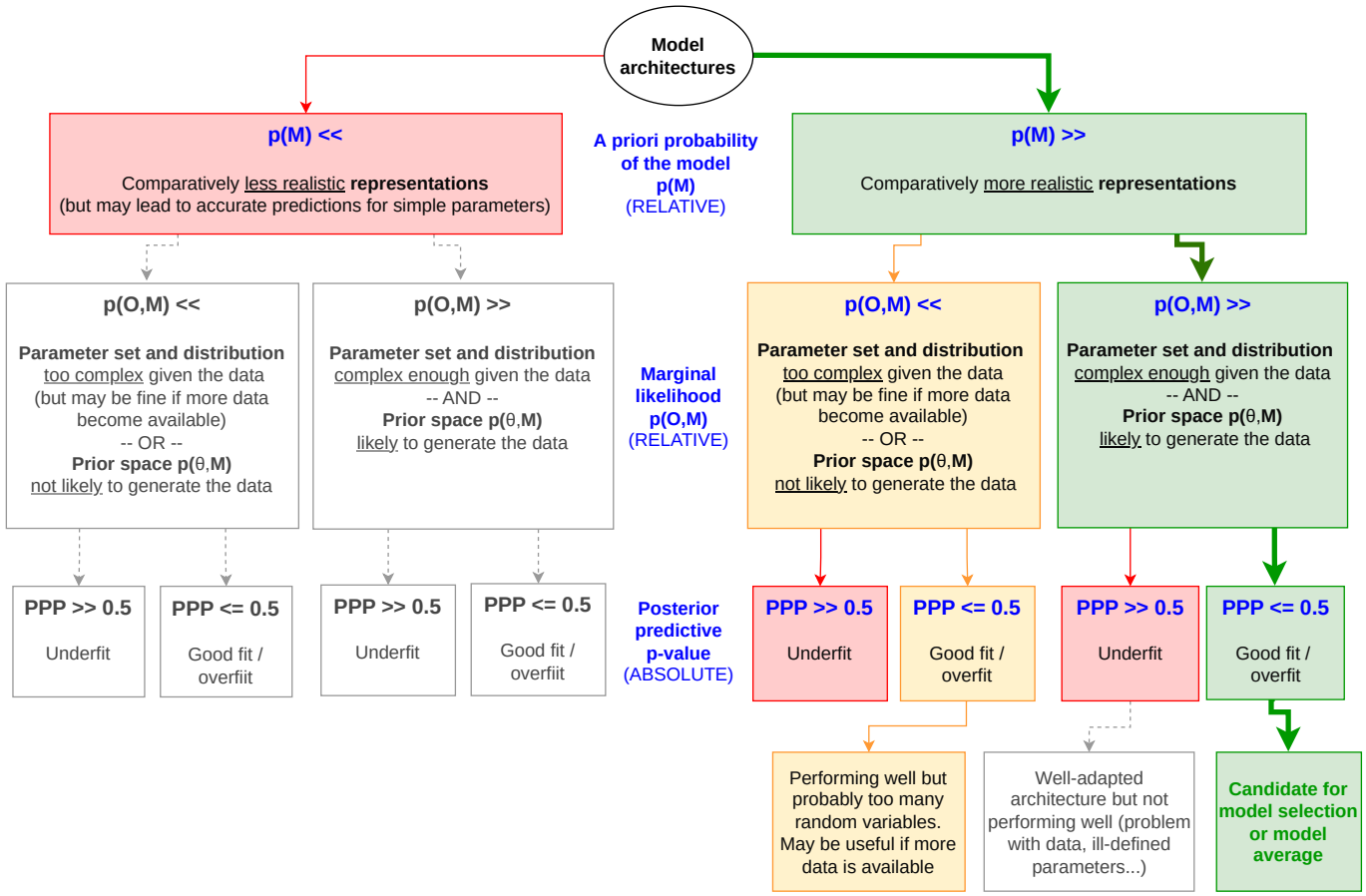

Fig. 2. Decision tree for model architectures. The successive metrics are indicated in the middle. Green arrows and boxes indicate the path of maximum likelihood for model consideration.

with $\boldsymbol{O}_{\rm rep}$ as generated sets of replicated observables (see, e.g., Galliano et al. 2021). Ideally the PPP should be around 0.5, while values near 1 imply a poor fit (underfit) and values near 0 imply a an overfit.

Another important quantity is the marginal likelihood which integrates all parameter combinations from the prior space and therefore enables hypothesis testing (i.e., what is the simplest model adapted to the data), defined as:

$$p(\boldsymbol{O}|\mathcal{M}) = \int_{\theta} p(\boldsymbol{O}|\boldsymbol{\theta}, \mathcal{M}) p(\boldsymbol{\theta}|\mathcal{M}) d\boldsymbol{\theta}. \quad (8)$$

When comparing two models to each other, one may use the Bayes factor which is the ratio of the marginal likelihoods. However, it is necessary to consider the a priori probability of the models themselves, $p(\mathcal{M})$, describing how well the set of parameters is adapted to the object we wish to model (independently on the exact set of observations). The Bayes Factor for two models $\mathcal{M}_1$ and $\mathcal{M}_2$ then becomes

$$BF = \frac{p(\boldsymbol{O}|\mathcal{M}_1)}{p(\boldsymbol{O}|\mathcal{M}_2)} \frac{p(\mathcal{M}_1)}{p(\mathcal{M}_2)}. \quad (9)$$

While the PPP are easily calculated and interpreted in absolute ways, the marginal likelihood and the a priori probability of the models are much more complicated. The marginal likelihood is often too difficult to evaluate for simple random walkers as it is necessary to sample well enough the entire prior space. `MULTIGRIS` uses the Sequential Monte-Carlo method which runs a large number of small Markov Chains across the prior space in a series of steps until convergence to the posterior distribution (Del Moral et al. 2012). This makes it possible to estimate the marginal likelihood for each model[3]. The a priori model probability $p(\mathcal{M})$ is difficult, if not impossible, to evaluate. By default one may simply consider that a set of models are equally plausible to represent a galaxy and therefore ignore these probabilities, or else that one model is far more plausible than another one.

Armed with the above quantities, we propose the following decision tree when several model architectures are to be compared, with all steps illustrated in Figure 2.

- First, a qualitative assessment of $p(\mathcal{M})$ is necessary to decide arbitrarily which model architectures are plausible to start with. Plausible model architectures are considered realistic representations of galaxies that depend the least possible on the exact set of tracers used for constraints. The probability $p(\mathcal{M})$ typically involves parameters that cannot be varied, including the overall model architecture itself (e.g., few sectors versus LOC). It is worth noting that implausible models may actually lead to accurate predictions for some simple galaxy parameters (e.g., SFR, ionized gas mass).
- Second, it is important to evaluate the likelihood of the prior space to generate the data, which is encompassed by the marginal likelihood, $p(\boldsymbol{O}|\mathcal{M})$. Large values imply that the model architecture is complex enough given the data and that

[3] While the Sequential Monte-Carlo method is particularly adapted to multi-model posterior distributions, we keep in mind that minor modes may be dropped during the resampling phase, effectively cropping the posterior distribution. Our inference runs use the largest possible number of particles to alleviate this issue.

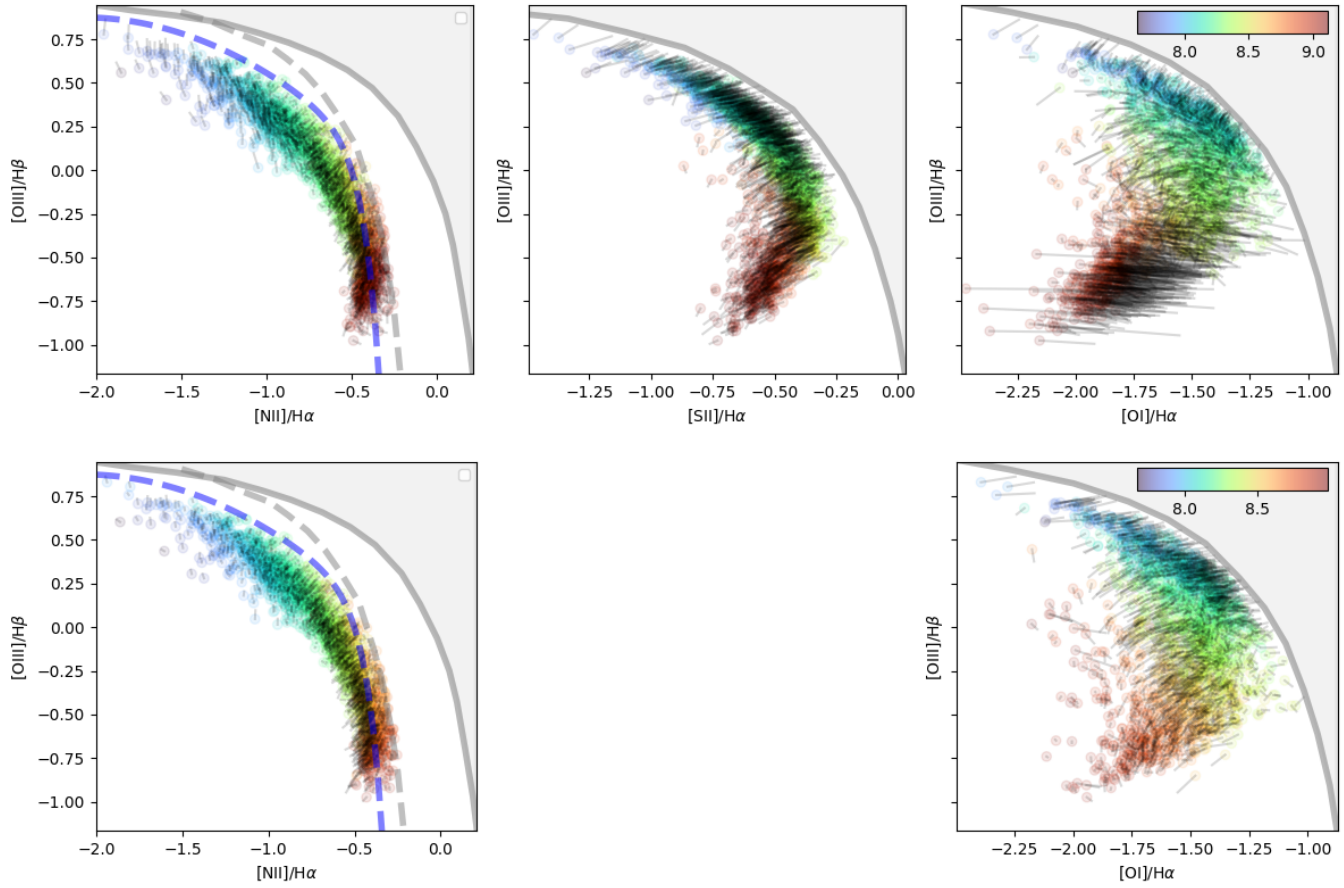

Fig. 3. Excitation diagrams for the inference run with (top) and without (bottom) [S II] lines. The color points show the modeled values (see Sect. 3.3.2), with the color scaling with the metallicity. The solid gray curves show the extreme starburst delimiting line from Kewley et al. (2001), while the dashed gray curve is from Kauffmann et al. (2003), and the dashed blue curve is from Stasińska et al. (2006). The thin black lines shows the distance between the observed and modeled ratios. For [N II]/Hα we use the specific prescription from P22 using both [N II] lines with a scaling factor.

the prior space is likely to generate the data. Inversely, low values imply either that the model architecture is too complex given the data – but may be fine if more data become available – and/or that the prior space is simply not likely to generate the data. The combination of the first and second step corresponds to the Bayes Factor described in Equation (9).

– Finally, plausible, complex enough, model architectures whose prior space is likely to generate the data can be selected based on their predictive diagnostics. The *PPP* can be used to distinguish between overfitting, ideal fitting, or underfitting. Model architectures selected so far but resulting in underfits correspond to well-adapted architecture, with some issues regarding for instance the data or ill-defined parameters. On the other hand, overfitting is not a problem per se since it merely reflects relatively weak constraints (or overestimated observational uncertainties) and our approach is committed to the most realistic models possible. With such overfits, the resulting posterior distribution may be particularly wide, or even identical to the prior distribution, but the predictions should remain reliable enough to interpret. In other words, overfitting may describe a realistic model that simply needs more observables to constrain better.

We emphasize that *PPP* is the only metrics in the sequence that can be interpreted in an absolute way. The final models passing these steps may then be directly considered or even averaged. We also consider another metric for convenience in some plots of this study, which is the fraction of posterior draws matching the observed values within 3σ.

3. Application to ECO star-forming galaxy sample

3.1. Observations and sample

We use data from the Environmental COntext (ECO) catalog DR3 (Moffett et al. 2015; Hutchens et al. 2023), which is a volume-limited data set in the northern spring sky spanning a recession velocity range of $3000 < cz$ [km s^{-1}] < 7000 (where cz is corrected for Local Group motion and is based on group-averaged cz values to minimize peculiar velocities). Being volume-limited, the sample mostly comprises low-mass dwarf galaxies. ECO has been crossmatched with SDSS spectroscopic observations by (Polimera et al. 2022, hereafter P22) and Polimera et al. (in prep.). We use the MPA-JHU catalog for the line flux measurements (Tremonti et al. 2004), with the internal extinction corrections based on the Balmer decrement method calculated in P22. The catalog was filtered in order to use reliable detections with a signal-to-noise ratio (S/N) greater than 5 for the strong lines Hα, Hβ, [O I] λ6300, [O III] λ4959, [O III] λ5007, [N II] λ6548, [N II] λ6584, [S II] λ6717, and [S II] λ6731 (see P22).

We select only the subset of star-forming galaxies, relying on excitation diagrams using [O III]/Hβ and ([N II], [S II], [O I])/Hα (see P22 and Fig. 3). Specifically we use the "definite-SF"

category in P22. This category will be revised in Polimera et al. (in prep.) based on the demarcation line of Stasińska et al. (2006), and we have verified that the trends presented in this study remain unchanged depending on which demarcation line is used (see Appendix A). We note that the S/N requirement for all lines, and especially on [O I] λ6300, is meant to be able to help distinguish between star-formation and AGN activity but produces a somewhat biased toward brighter star-forming galaxies.

The final sample is therefore almost volume-limited and comprises 2052 galaxies. Ancillary data is available for ECO, of which we use in particular the SFR (derived from GALEX and WISE including machine-learning UV magnitude predictions for half of the sample withough deep near-UV data; Carr et al. 2024 and Polimera et al., in prep.) and the stellar mass (Hutchens et al. 2023).

3.2. Database of 1D models

Our modeling approach considers statitical distributions of 1D models to describe the integrated emission of galaxies (Sect. 2). We adopt the methodology from Richardson et al. (2022) for the 1D photoionization models in this work. The stellar SEDs originate from the BPASS v2.0 (Eldridge et al. 2017) models using an instantaneous star-formation burst (i.e., simple stellar population). The post-starburst ages span 1−25 Myr, which ensures that multiple Wolf-Rayet stages are captured (see D'Agostino et al. 2019). In actuality, the post-starburst age essentially accounts for the hardness of the radiation field since any given star-forming region realistically has a mixture of stellar ages. The stellar metallicities range from 0.05 $Z_\odot$ to 2.0 $Z_\odot$. Our model grid extends to 3.0 $Z_\odot$, but BPASS models are unavailable at these metallicities, so we substitute 2.0 $Z_\odot$ models in the cases where $Z > 2.0\,Z_\odot$.

The abundance scalings in the cloud are taken from the Galactic Concordance Abundances described in Nicholls et al. (2017) where the solar standard is defined as 12 + log(O/H) = 8.76. We use the depletion patterns described in the appendix of Richardson et al. (2022) for a fixed depletion strength of $F_* = 0.45$. The model grid uses the parameter $Z/Z_\odot$ for metallicity, which refers to the abundances in the cloud prior to grain depletion. Therefore, one needs to deplete the oxygen abundance by −0.11 dex to obtain "gas-phase abundances" in terms of 12 + log(O/H). We assume a grain composition like the Orion Nebula (Baldwin et al. 1991), in addition to including polycyclic aromatic hydrocarbons (Abel et al. 2008), and use a D/G ratio scaled with metallicity according to the empirical relation from Rémy-Ruyer et al. (2014). The hydrogen density at the face of the gas cloud, log $n_{\rm H}$ varies from 0.5 to 4.0 in 0.5 dex increments, while the ionization parameter U, also defined at the ionized face, varies from −4.5 to −0.5 in 0.25 dex increments. The models are run until an electron fraction of $n_e/n_{\rm H} = 0.01$.

The integrated distributions (LOC) models are drawn from this grid of 1D models. We briefly describe tests using other photoionization codes in Sect. 5.3. Our grid includes a potential radiative component powered by an AGN, but for the present star-forming sample we select only a subgrid with an "AGN fraction" (i.e., fraction of ionizing radiation due to an AGN) of $f_{\rm AGN} = 0$. We defer the study of galaxies with significant or dominant AGN fraction but tests have been performed to verify that the AGN fraction for the present sample, if let free, never reaches above 8% (see Polimera et al. 2022) with most galaxies showing $f_{\rm AGN} < 4\%$.

Before inferring the metallicity and other parameters, we first compare the metallicity from the grid of single 1D models against empirical line ratio diagnostics from Garg et al. (2024)[4]. The metallicity from the grid controls the elemental abundances that, in turn, are used to compute the radiative transfer within Cloudy. For this comparison, we restrict the parameter ranges (average for LOC) to 12+log(O/H) = [7.4, 9.0], $U = [-3.5, -2]$, and $n = [1, 2]$ to match the ranges used for calibration. For the LOC distributions, we consider boundaries for U and n that are typically found in the present study ([−4, −2]) and [1, 3] respectively) and ensure that the average values remain within the calibration ranges. We find good agreements overall across the considered metallicity range, especially with the N2O2, N2S2, N2, S2, R3N2, and O3N2 diagnostics (Fig. 4). This implies that the metallicity parameter in the grid corresponds, to first order, to the metallicity obtained from the empirical diagnostics.

Differences likely stem from the assumed N/O abundance ratio prescription in the various calibrations (see specific discussion in Garg et al. 2022). Any deviation will result in important biases if we wish to compare the metallicity from the grid (e.g., the value inferred through modeling observations) to the metallicity from empirical calibrations and may highlight specific model assumptions regarding, for instance, the depletion pattern of some elements as a function of metallicity.

The best agreement between our models and empirical diagnostics is found for N2O2, but the [O II] doublet is unfortunately difficult to measure and available only for a small number of ECO sources and with large uncertainties. All other relatively reliable empirical diagnostics show some kind of deviations compared to our models. Therefore, we keep in mind in the following that the metallicity inferred from our models may deviate somewhat from that obtained with empirical diagnostics.

3.3. Model architectures for ECO star-forming galaxies

3.3.1. Relevant architectures

Based on the a priori model probability $p(\mathcal{M})$ criterion (Sect. 2.3), LOC models (using either power-law or normal distributions) are preferred to multicomponent "xCyS" models due to the evidence of heterogeneity of ISM and energetic source properties (see introduction). Building upon the recent modeling effort from Ramambason et al. (2024), we present here the first application of LOC models with the age and Z following a statistical distribution, in addition to n and U. For the integration boundaries of LOC models ($p_{\rm min,max}$), MULTIGRIS considers by default the minimum and maximum values in the grid for each parameter, but free boundaries are considered in the present study.

Overall, many different architectures have been investigated with the number of random variables ranging from 5 to $\gtrsim$20 (Table 1) and we will only focus on a few model architectures afterward. For simplicity, we present here the results assuming power-law distributions only. The reasoning is that, on first order, narrow normal distributions may be approximated by single parameter values while broad normal distributions may be approximated by flat power laws. Although it would be interesting to thoroughly test various distributions, we keep in mind that we may not be able to afford fine-tuning in the model architecture (e.g., testing a power law versus normal distributions)

[4] Contrary to what is indicated in Garg et al. (2024), we do use the [O II] doublet sum for N2O2 and we do consider a sum of logarithms for N2S2.

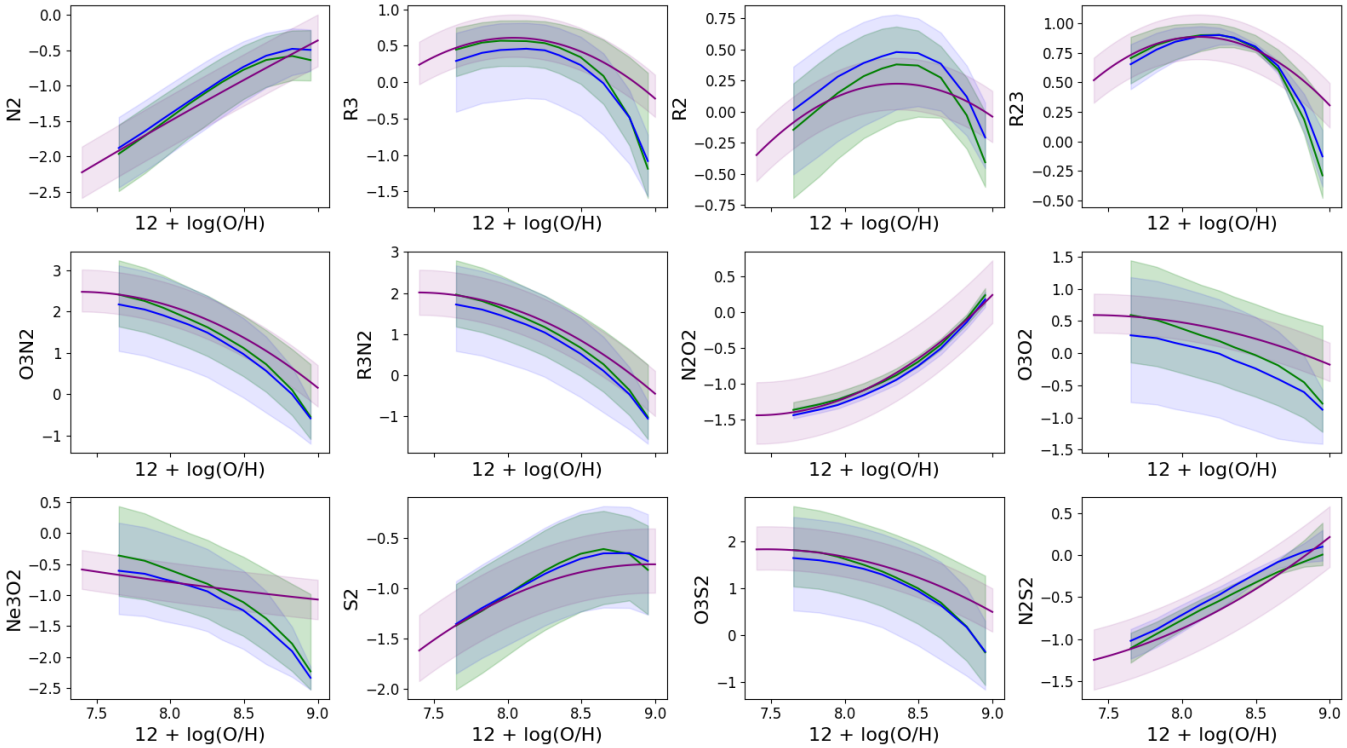

Fig. 4. Empirical line ratio diagnostics as a function of metallicity for the model grid (the single 1D model is in blue and the LOC distribution is in green), compared to the calibration from Garg et al. (2024) in purple. The shaded areas correspond to the range of physical conditions used in our photoionization grid (see text).

Table 1. Subset of tested model architectures.

	1C1S	2C1S	LOC U/n	LOC U/n/age/Z
	$\delta_{\rm age}$	$\delta_{\rm age1}$, $\delta_{\rm age2}$	$\delta_{\rm age}$	$\alpha_{\rm age}$, age$_{\rm min,max}$
	δ_U	δ_{U1}, δ_{U2}	α_U, $U_{\rm min,max}$	α_U, $U_{\rm min,max}$
	δ_n	δ_{n1}, δ_{n2}	α_n, $n_{\rm min,max}$	α_n, $n_{\rm min,max}$
	δ_Z	δ_{Z1}, δ_{Z2}	δ_Z	α_Z, $Z_{\rm min,max}$
RVs	4	8	8	12

Notes. α, δ, as well as minimum and maximum values refer to the distribution hyperparameters (see Eqs. (2) and (3)). RVs indicates the raw number of random variables (ignoring potential correlations).

considering the various other caveats (e.g., projection effects, 1D model grid hypotheses).

We consider "xCyS" models strictly for comparison and continuity with previous works, despite the fact that these models are considered comparatively less realistic because they assume that the excitation sources or the ISM are fully described by one or two components typically, i.e., by a single parameter value (e.g., single density) or by a combination of two parameter values (e.g., two components with a single density for each).

3.3.2. Model selection

Based on the marginal likelihood $p(\boldsymbol{O}|\mathcal{M})$, the single component model (1C1S) is not favored compared to the two component models (1C2S or 2C1S) because the prior space is too simple and therefore not likely to generate the data. LOC models, on the other hand, may quickly become too complex given the data. As mentioned earlier, this is not a problem per se, and we aim to compare and select the best model architectures that are deemed realistic enough.

Based on the *PPP*, we find that 1C2S or 2C1S models perform quite well, and the only reason we do not fully consider them for interpretation is that they were not selected initially as realistic models based on the a priori probability of the model $p(\mathcal{M})$ (Sect. 3.3.1). The reason why these models perform so well is due to the parameter flexibility: the parameters for the few (or single) considered components are independent, a single value per component is fine-tuned to match the observations, and the value is not necessarily expected to correspond to a physically meaningful region of the galaxy.

Among LOC models, the best models are those with free parameter boundaries ($p_{\rm min,max}$) for the integration. Even though models with free boundaries involve a larger number of random variables (Table 1), the LOC models with free boundaries do not indicate significant overfitting (*PPP* < 0.5) and show only slightly lower marginal likelihoods compared to the models with fixed (minimum and maximum in the grid) boundaries due to the expanded parameter prior space. Our final model architectures are LOC models with power-law distributions and free parameter boundaries.

It should be emphasized that the match between the LOC models and observations is driven simultaneously by the topology assumptions (parameters describing such combinations) and by the inherent 1D model database (emission line predictions from a set of parameters such as metallicity, ionization parameter, etc.). We choose to design the most adequate photoionization grid possible and focus on topology improvements, but we keep in mind that our results are strongly dependent on the reference grid and that other conclusions could be reached with different prescriptions, e.g., for the radiation sources, or the ISM

composition (see, e.g., Lecroq et al. 2024). We also remind that the distributions we implement do not correspond to the entire ISM of galaxies but are biased toward the emitting components and, as such, are driven by H II region properties in this star-forming galaxy sample.

4. Results

An illustration of the inference results obtained for individual galaxies is presented in Appendix B. By default, we consider in the following probability density functions (PDFs) for the entire sample. For this, we do not simply select the mean of each parameter for each object but concatenate the draws for all the objects in order to keep the information contained within the confidence intervals.

4.1. Influence of the line set

Figure 5 shows that the predicted line fluxes agree within $\approx 2\sigma$ for all galaxies in the sample with the LOC architecture using all available lines. Looking at the specific posterior predictive p-value (PPP) for each line for the 1C1S and LOC architectures using all available lines (Fig. 6), we find that underfitting is largest for [S II], [O I], and [N II], all being slightly overpredicted by the model. Motivated by the underfitting of [S II] lines as well as potential deviations between empirical metallicity diagnostics involving [S II] and the metallicity from the grid (Sect. 3.2), we ran the inference without the [S II] lines as constraints. We also considered runs with the [O II] doublet sum, despite with poor signal-to-noise ratio, instead of the [S II] lines.

Figure 7 shows that LOC models ignoring the [S II] lines perform much better. This remains true even using [O II] instead of [S II]. Furthermore, Figure 6 shows that [N II] and [O I] are dramatically better matched ($PPP \leq 0.5$) in runs ignoring [S II] (see also Fig. 3). In the following, we will consider runs that include or do not include [S II] or [O II] to study the impact of the line set on our results. The overprediction of [S II] in the models may be due to the 1D model grid assumptions (e.g., need to account for sulfur depletion, need for more refined stellar atmospheres for the relevant energy range; Sect. 3.2) and/or to systematic effects in the line measurement available in the SDSS catalogs (see discussion in Polimera et al. 2022) that may be due to the difficulty in removing nearby telluric features. We also have verified that ignoring the [O I] line for the inference of the entire sample does not significantly modify the metallicity determination.

4.2. Performance of single 1D models (1C1S) versus LOC

Here we wish to compare the – often used – single 1D model approach (1C1S) to the LOC one. For this comparison we are therefore interested in biases specifically due to the model architecture. For this test the LOC boundaries ($p_{\rm min,max}$) are not linearly interpolated but the slope (α_p) is continuously sampled. Hence we consider 1C1S with all parameters linearly interpolated instead of nearest neighbor for a fairer comparison, keeping in mind that LOC models would perform even better through boundary interpolation between models.

Figure 7 shows that, even considering the best possible 1C1S models (with linear interpolation for all parameter), LOC models always perform better in all metrics. While we show for reference the results for the inference ignoring [S II] in Figure 7, the same results hold for all inference runs, and also hold for tests

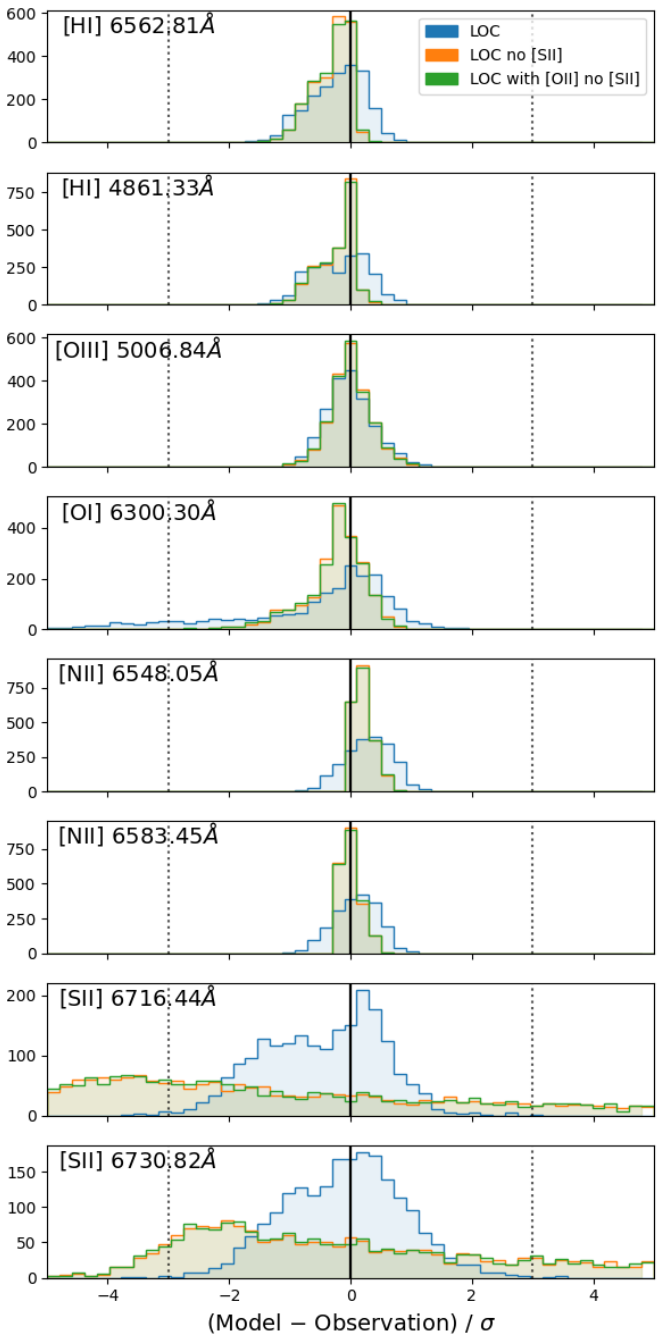

Fig. 5. Histogram comparison between predicted and observed fluxes scaled by the observed errorbar for the entire sample. The vertical lines delimit the agreement within 3σ.

with only one or two parameters following statistical distributions instead of single values.

The fact that LOC models globally outperform 1C1S models strengthens the hypothesis that physical conditions are not uniform within the galaxies of the sample, and consequently that the integrated tracers we observe do contain useful information about the distribution of matter and radiation sources. Since our tests show that PPP does not improve significantly beyond (any) two parameters being distributed as power laws, we conclude that this is most likely the minimum amount of complexity dictated by the set of tracers we use. However, we keep in mind that physical motivations exist for all parameters to follow some

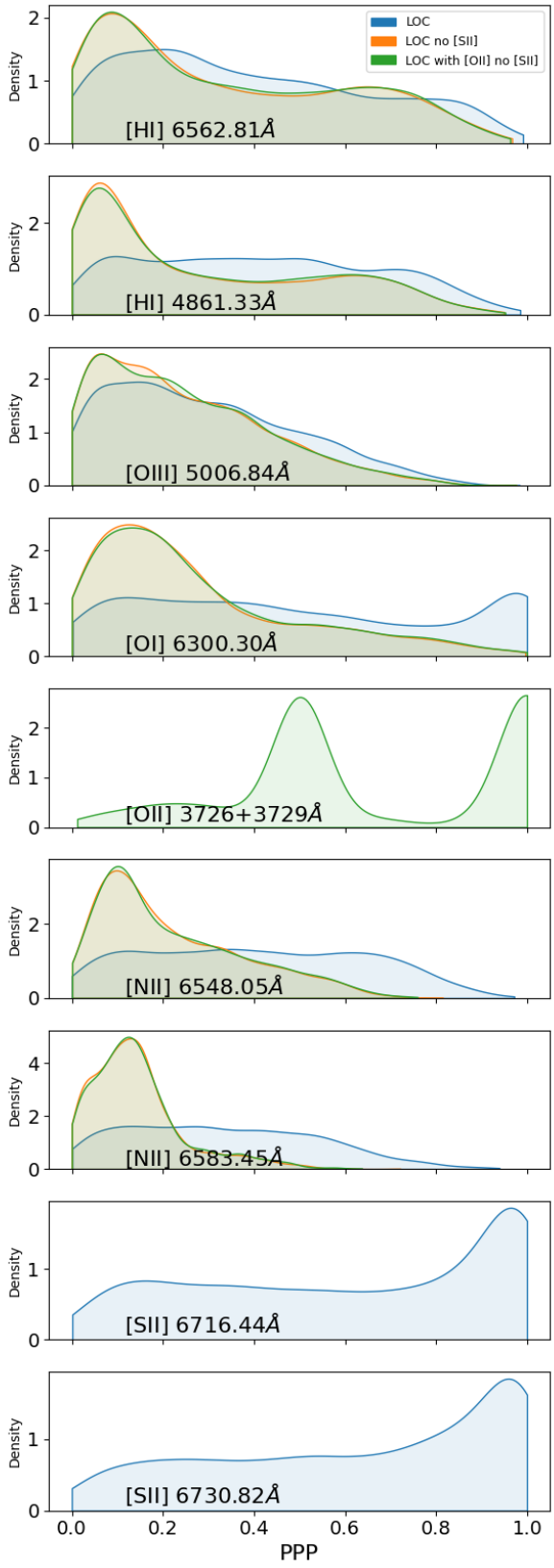

Fig. 6. Posterior predictive p-value (PPP) for each line.

kind of statistical distributions and that LOC models are considered more likely a priori (independently on observed tracers, i.e., with a larger $p(\mathcal{M})$ value) than multicomponent "xCyS" models.

4.3. Physical meaningfulness of LOC average and single 1D model values

While we show in Sect. 4.2 that the LOC approach outperforms the single 1D model approach, potential biases on physical parameter determinations due to the chosen approach need to be addressed. Output physical parameters using single 1D models are often interpreted as some kind of average quantities and we put this assumption to the test by comparing the single parameter values of 1C1S models to the average quantities from LOC models (see Eq. (5)).

Globally the 1C1S values are always compatible with the LOC averages within a factor of a few (Fig. 8). We keep in mind that for all parameters, 1C1S or LOC averages never reach the grid minimum or maximum values and that there is no edge effect. We may either consider the 1C1S approach as some kind of reference because it is often used and it is therefore reassuring that the LOC average value matches the 1C1S value, or we may also consider the LOC approach to be more realistic and viable and it is reassuring that 1C1S models provide values that do not deviate significantly. However, it should be restated that the 1C1S models globally perform less well than LOC and that the biases we identify are interpreted as biases due to the single 1D model hypothesis (in other words, we are not trying to validate the average LOC quantities).

Looking more closely, we find some small deviations. For the metallicity, there is little bias for very low- and very high-Z but there is a "kink" around solar metallicities with the 1C1S Z somewhat lower than the LOC average Z, especially for the run that includes the [S II] lines. We find that this bias is not due to one single specific parameter in the LOC models, but to all of the parameters in aggregate. For instance, the relatively wider range of Z (boundaries) around solar metallicities could explain in part this bias but also the consideration of ranges (LOC) for U or age that results in a range of Z because of the dependency between both U and age with Z (see discussion in Sect. 4.4.2).

The age (i.e., spectral hardness of the input radiation field) parameter shows significant biases, with the 1C1S age being overestimated in the most metal-poor galaxies (blue points in Fig. 8) and being underestimated in slightly subsolar to solar metallicity galaxies (orange points). There is almost no bias for the highest-Z galaxies (red points).

For the ionization parameter U, there is no clear bias. For the density n, there is also no clear bias apart from globally slightly lower values with 1C1S compared to LOC averages. This is likely due to the relatively poorer performance of 1C1S (all parameters linearly interpolated) including and especially with [S II] lines.

Globally, we notice that there seems to be a special regime around solar metallicity corresponding to a turnover in U and age that may lead to significant biases using single 1D model results. As a function of metallicity, there is a clear trend for the age (and therefore hardness) to decrease between 20 Myr and 5 Myr until solar metallicity and increase again. Similarly, as a function of metallicity, there is a clear trend for the ionization parameter to decrease (until slightly subsolar metallicities) and increase again. Thus, we definitely see a relationship between age, Z, and U (see also Sect. 4.4.2). We note that these turnovers are not driven by the grid minimal or maximal values for each parameter in the grid.

In summary, biases may exist using single 1D models that may affect the interpretation of the inferred parameters or other, related, parameters. The physical parameters tackled in this study are related to H II regions, and their average

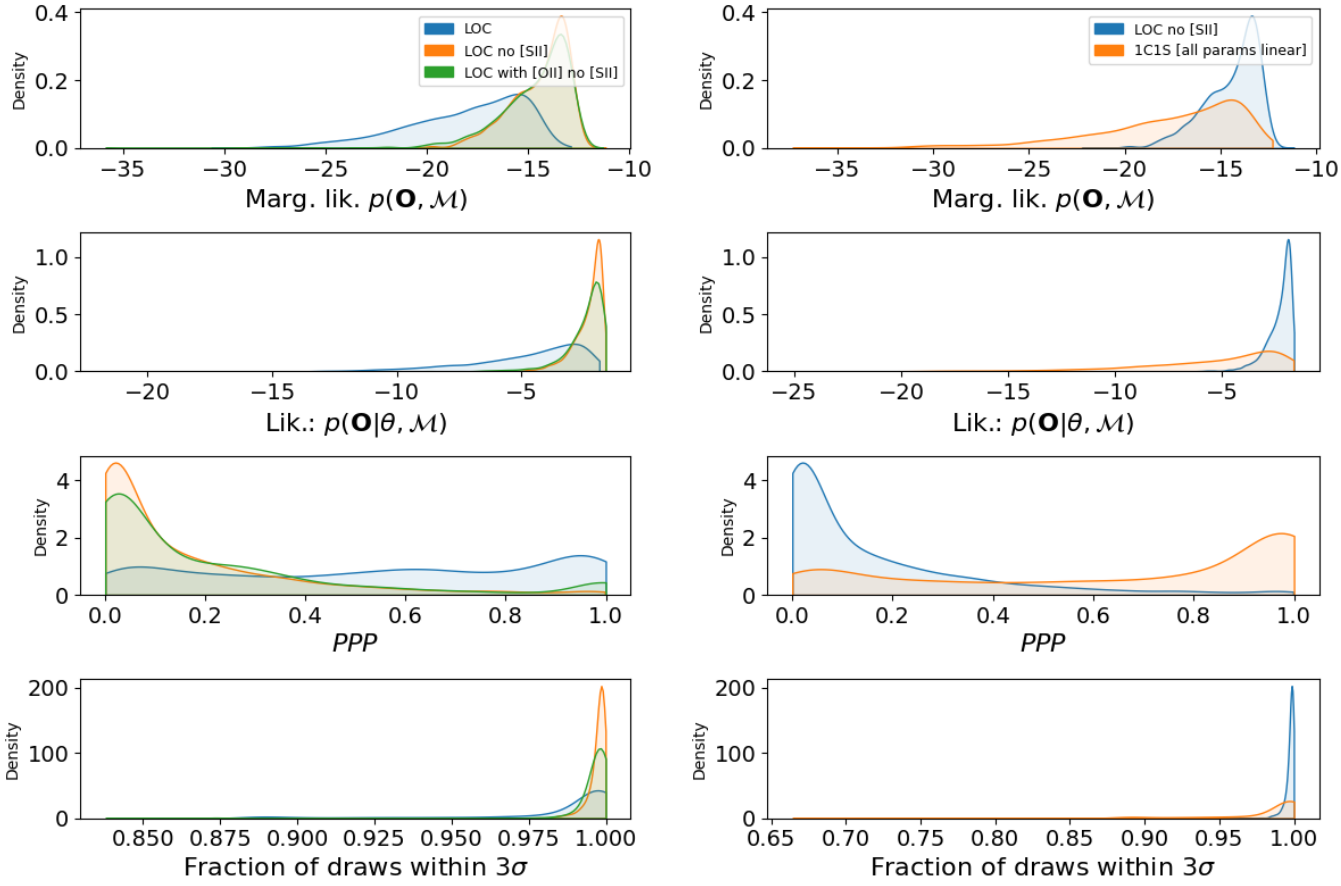

Fig. 7. Performance metrics for the entire sample. From top to bottom: the marginal likelihood and the likelihood (evidence), the absolute posterior predictive p value (PPP), and the fraction of posterior draws matching the observations within 3σ.

values do not depend significantly on the ISM topology. However, stronger biases may exist for other specific galaxy parameters (e.g., H_2 masses, escape fraction of ionizing photons, etc.; Ramambason et al. 2024).

4.4. Inferring internal parameter distributions within galaxies

In the following we examine the PDF of the average parameter values within galaxies $Z_{\rm avg}$, $U_{\rm avg}$, $n_{\rm avg}$, age$_{\rm avg}$ (not to be confused with the average value across the sample). For some parameters (e.g., $Z_{\rm avg}$ and $U_{\rm avg}$), the PDFs for individual objects are much narrower than the sample PDF, implying that the latter describes well the different properties of the studied galaxies (Figs. 9 and B.3). For the other parameters (in particular $n_{\rm avg}$), the PDFs for individual objects are identical on first order and the overall sample PDF thus reflects a common PDF, valid for all galaxies.

The PDFs of the power-law slopes (α) are similar on first order for all galaxies for any given physical parameter, which may indicate a universal origin of the distribution but could also reflect the difficulty in constraining α from the observed tracers used for inference. Considering the PDFs from Figure B.3, we propose that small variations of α for a given galaxy will lead to significant variations of the boundaries, and that the observed tracers mostly constrain the average physical parameter value. Small variations do exist, however, for all hyperparameters, from galaxy to galaxy and from the prior distribution, and we investigate them in the following.

4.4.1. Hyperparameters

The hyperparameter α (slope of the power-law distribution) reflects the relative proportions of emitting components with given physical properties. The slopes for age, density, and metallicity are close to 0 (Fig. 9), hinting that for most galaxies the emission is not significantly dominated by a given dense versus diffuse, old versus young, or high versus low metallicity. For most galaxies, however, the emission is dominated by relatively low excitation components ($\alpha_U < 0$), with log $U_{\rm avg}$ fairly peaked around $\approx$−3.2. The low-excitation components, that contribute most to the total emission of most galaxies, show a narrow range of log $U_{\rm min}$ between −4.5 and −3. Inversely, the higher-excitation components show a wide range of log $U_{\rm max}$ centered around −2.

The distribution of age$_{\rm avg}$ peaks around 5 Myr. The distribution of the lower and upper boundaries peak around age$_{\rm min}$ $\approx$ 2 and age$_{\rm max}$ $\approx$ 10 Myr respectively, but the upper boundary extends to the Wolf-Rayet phase at $\approx$20 Myr, which is the hardest radiation field in the grid (D'Agostino et al. 2019).

The distribution of $n_{\rm avg}$ almost spans the entire parameter space, except if [S II] lines are used as constraints. While it is expected that [S II] lines help constrain the density (e.g., Osterbrock & Ferland 2006), we note that these lines may also cause some biases (Sect. 4.1).

The distribution of $Z_{\rm avg}$ is bimodal, with a stronger bimodality for the inference run using the [S II] lines. Most galaxies lie around $Z_{\rm avg} \sim 0.3\,Z_\odot$ and populate the leftmost peak. The secondary peak lies around $2\,Z_\odot$ if [S II] lines are used, and otherwise around solar metallicity. The distribution of the lower

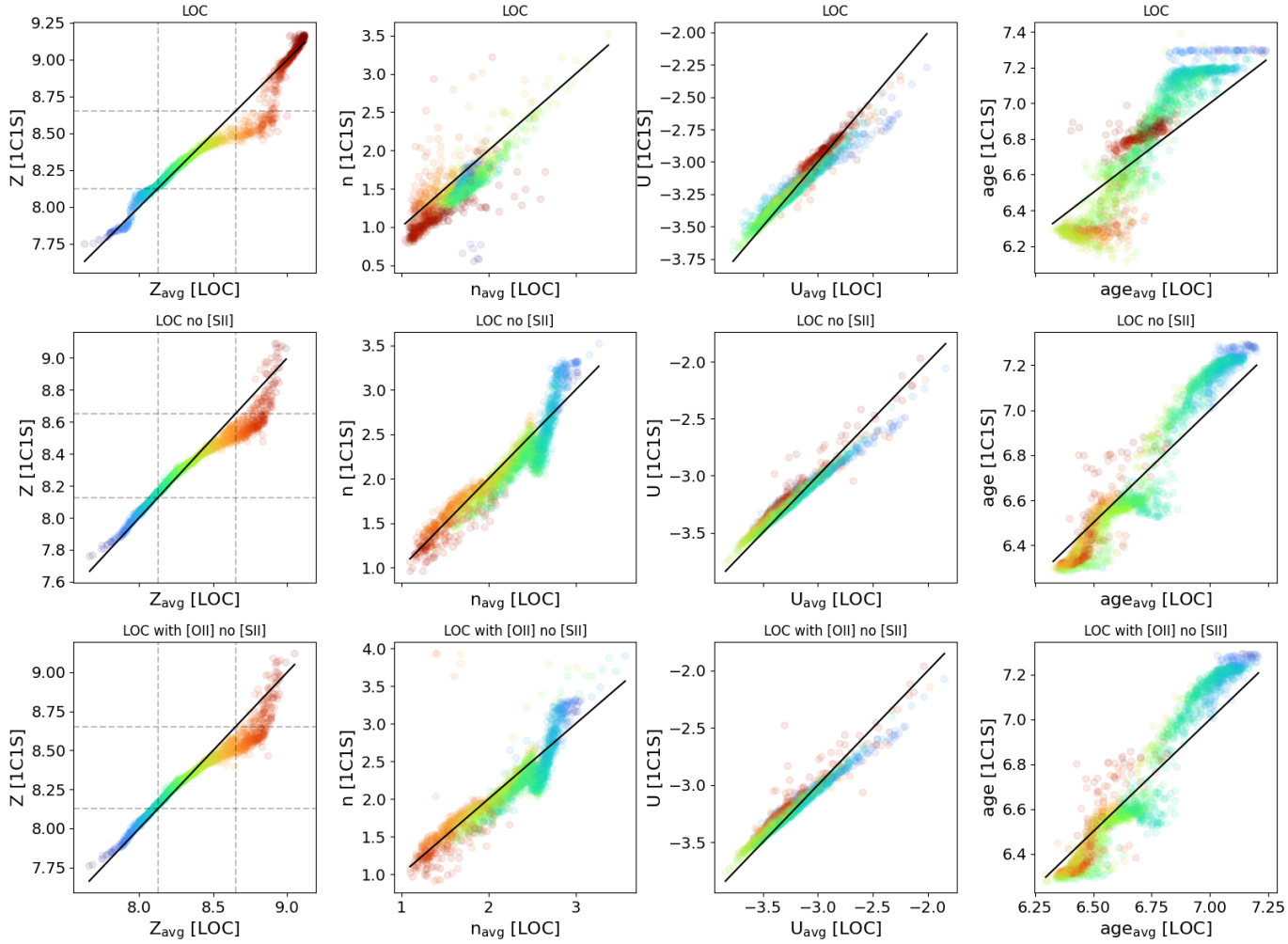

Fig. 8. Comparison of single 1D models (1C1S) versus LOC averages. The color scales with the metallicity parameter. From top to bottom we show the results for the runs ignoring [O II], ignoring [S II] and [O II], and ignoring [S II] (Sect. 4.1). The dotted lines in the leftmost plots indicate the 0.3 $Z_\odot$ and $Z_\odot$ values.

boundary peaks around $Z_{\rm min} \approx 0.15\,Z_\odot$. The distribution of the upper boundary $Z_{\rm max}$ is strongly bimodal (Fig. 9), clearly driving the bimodality of the average metallicity.

In individual galaxies, it must be emphasized that boundaries for all parameters are mostly well separated (Fig. 9), yet there is no prior to force a minimum difference between the maximum and minimum value to be considered for integration. In other words, the inference could have resulted in boundaries being equal or almost equal (i.e., being equivalent to a single 1D model) if this had been a more likely solution (see also Appendix B).

4.4.2. Correlations between physical parameters

We investigate the correlation between physical parameters using their average value in each galaxy. Results are shown in Figure 10 for the inference runs with and without [S II]. We note that there is no degeneracy between the parameters and that the PDFs of individual galaxies clearly prefer one solution (see example in Fig. B.1).

There is no clear trend between $\rm age_{avg}$ and $U_{\rm avg}$. However, we find a strong relationship between $\rm age_{avg}$ and $Z_{\rm avg}$. The most metal-poor galaxies are characterized by $\rm age_{avg} \approx 10\,Myr$. Around slightly subsolar metallicities, $\rm age_{avg}$ reaches down to $\approx$3 Myr, i.e., a softer radiation field. For high-metallicity galaxies, the runs including the [S II] lines indicate older ages around $\approx$5 Myr and therefore intermediary hardness, while the runs ignoring [S II] flatten around $\approx$3 Myr. Our sample is selected based on [O I] detection, thereby selecting relatively hard radiation fields, but other high-metallicity galaxies may actually not require such hard radiation field. The tendency for high-Z sources to require a hard radiation field could also be indicative that other high-ionization process may be important (e.g., shocks).

The $U_{\rm avg}$ versus $Z_{\rm avg}$ correlation shows the same trend as $\rm age_{avg}$ versus $Z_{\rm avg}$ but with the turn-off occurring at a lower metallicity. The slight increase of $U_{\rm avg}$ above 0.3 $Z_\odot$ for the runs including [S II] is reminiscent of the result obtained in high-redshift star-forming galaxies (Reddy et al. 2023). The average density $n_{\rm avg}$ also shows a decreasing trend with metallicity, from $\sim$500 $\rm cm^{-3}$ down to $\sim$50 $\rm cm^{-3}$, with a significantly tighter trend for the runs ignoring [S II].

Although not shown, we observe the same trends for the slopes $\alpha_{\rm age}$, α_U, and α_n versus Z as for the average parameters. Since the slopes reflect the weight of regions with given physical parameters toward the integrated galaxy emission, this implies

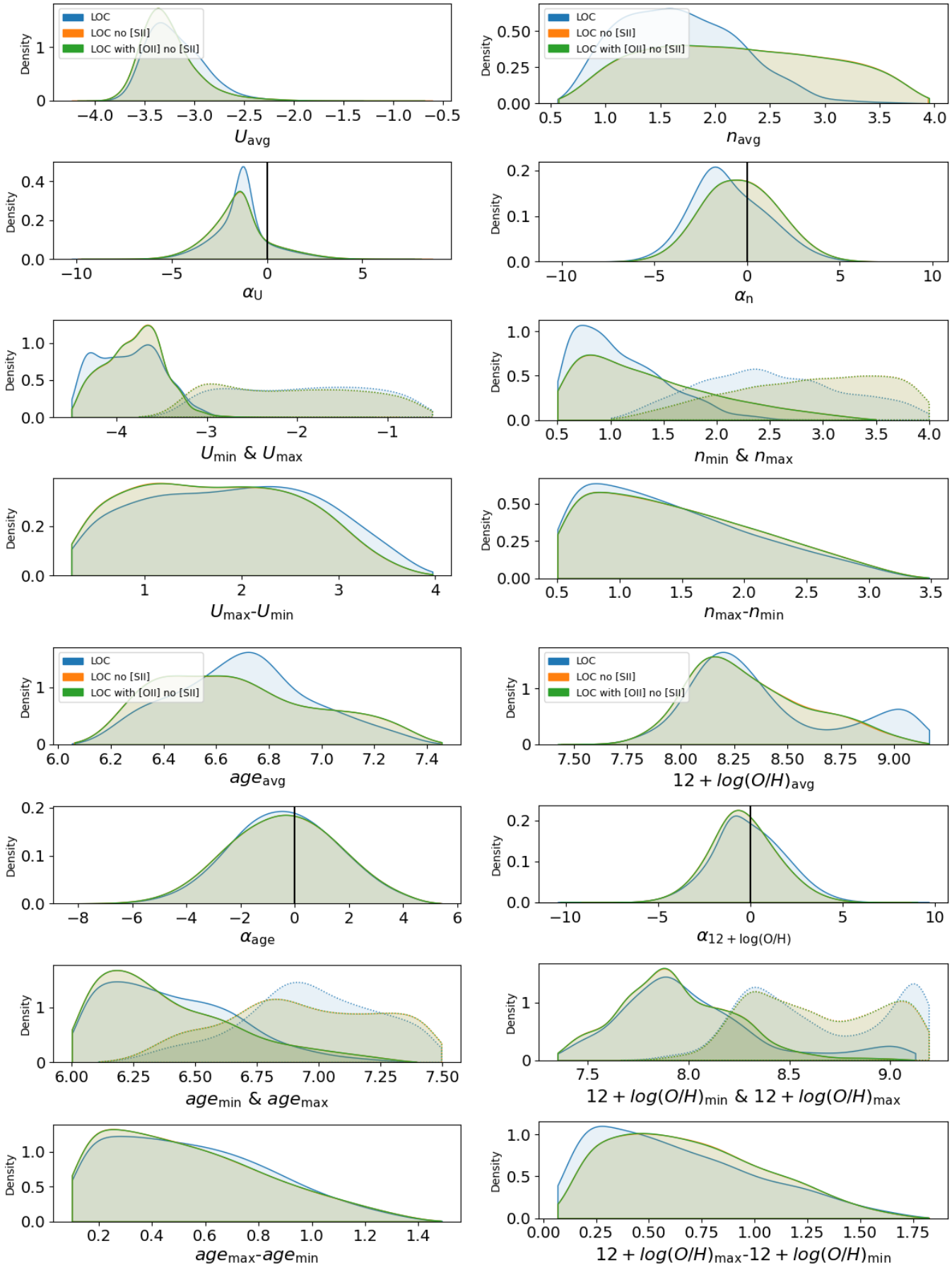

Fig. 9. Hyperparameter and average values.

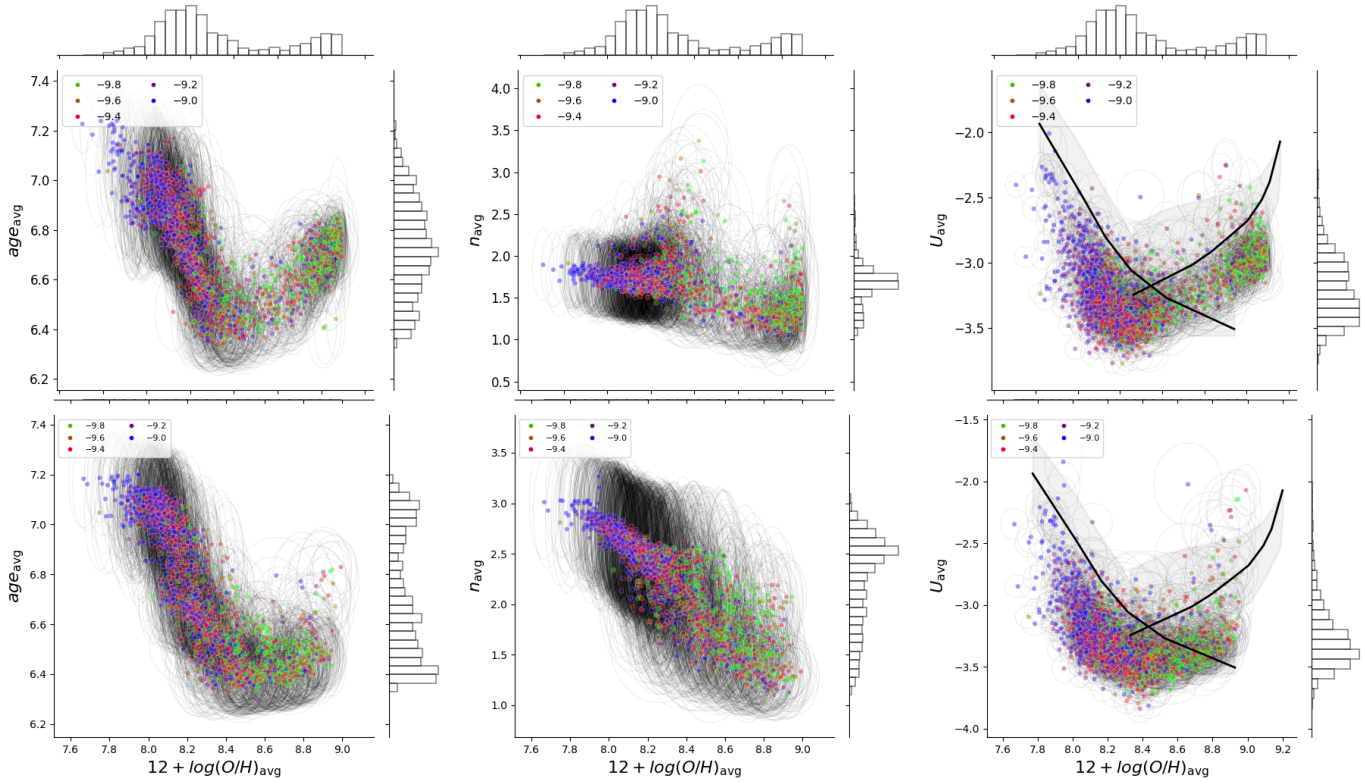

Fig. 10. Correlations between physical parameters for the runs including (top) or excluding (bottom) the [S II] lines. The color scale indicates the sSFR. For U versus Z, the low- and high-Z curves are from Kashino & Inoue (2019) and Ji & Yan (2022) respectively.

that as Z_{avg} decreases, there is an increasing proportion of harder stellar radiation field, high ionization parameter, and high density contributing to integrated galaxy spectrum.

5. Discussion

5.1. Ionization parameter versus metallicity

There is evidence in the literature that the ionization parameter U and the metallicity Z are physically related but the origin of the relationship is still debated (see, e.g., Dopita et al. 2006b; Ji & Yan 2022). First, there is evidence that U anticorrelates with Z in low-metallicity galaxies until about solar metallicity (or stellar mass $\sim 10^{10}$ $M_\odot$) (Kashino & Inoue 2019; Reddy et al. 2023; see black curves in Fig. 10). Independently, there is also evidence that U correlates with Z for metal-rich sources, above solar metallicity (Ji & Yan 2022). Suprisingly, there are few or no samples spanning a wide enough range of metallicities to verify whether these relations are specific to some given metallicity regimes. The ECO sample is ideal for studying this relationship because we have access to a wide range of masses and metallicities.

Our results show a relatively well-behaved relationship between U_{avg} and Z_{avg}, with a steep decline followed by a smooth increase (almost nonexistent if [S II] lines are ignored; see Fig. 10 bottom). Our results are in line with theoretical expectations. The steep decline could be explained by the wind-driven bubble model for H II regions of Dopita et al. (2006a), which would dominate at low-metallicity, together with the lower opacity of low-metallicity stellar atmospheres resulting in greater ionizing flux. This interperation is strengthened by models of multiple H II regions within a single galaxy (Garner et al. 2025). It should be noted that Kashino & Inoue (2019) show that, despite the strong apparent anticorrelation between U and Z in their low-metallicity sample, the U variation depends more heavily on the specific SFR (sSFR). In summary, U may be controlled by a competition between variations of Z and of sSFR, with a moderate U versus Z anticorrelation at low-metallicity steepened by sSFR.

On the other hand, the smooth increase of U_{avg} in the most metal-rich galaxies could be due to an increased SFR. This elevated SFR might be itself related to the quick enrichment of the lower-metallicity regions in metal-rich galaxies (see Sect. 5.2.1).

5.2. Distribution of physical parameters within galaxies

The LOC approach is motivated by the study of potential biases due to a single 1D model approach (Sect. 4.3) but also because it enables additional parameters relevant to galaxy evolution. While many different model architectures could be used to match the observed lines, we chose a plausible architecture with physical parameters distributed as power laws within each galaxy because they likely represent physically meaningful internal distributions (Sect. 3.3.2).

In this section we stand by this hypothesis and investigate what these distributions imply as far as galaxy evolution is concerned. In other words, given the observations, given the grid of 1D photoionization models, and given the assumption of power-law distributions, we wish to find and interpret the most likely internal distributions of physical parameters (Z, U, age, n) within each galaxy of the sample. We show in Figure 11 the variation of the upper and lower boundaries as well as average values as a function of metallicity discussed in the following.

5.2.1. Metallicity internal distribution

We first note that the power-law distribution inferred for any given galaxy should not be confused with the metallicity gradient often observed in disk-dominated galaxies (e.g., Carton et al. 2015; Hu et al. 2018; Bresolin 2019; Simons et al. 2021; including in the Milky Way, Balser et al. 2011) and thought to be the result of star-formation spreading outward through the disk (e.g., Sharda et al. 2021, 2024). A positive or negative slope α_Z in our models does not correspond to the slope of the metallicity gradient (the latter considering radial averages that do not contribute equally to the total emission) but reflects instead a given weight of metal-rich versus metal-poor regions emission within a galaxy toward global emission.

The inferred internal metallicity dispersion ($\Delta Z = Z_{\rm max} - Z_{\rm min}$) reaches up to ~0.7−1 dex around solar metallicity galaxies (Fig. 11), which is compatible with the dispersion often observed in 2D maps (e.g., Poetrodjojo et al. 2018; Nakajima et al. 2024). However, it must be noted that robust 2D metallicity estimates indicate that the abundance gradient should dominate the metallicity variation (e.g., Kreckel et al. 2019; Williams et al. 2022) and that our result may be the consequence of a 3D distribution and the consequence of potentially larger weights from metal-poor regions contributing to the total luminosity.

Our results for the ECO sample also show that $Z_{\rm min}$ and $Z_{\rm max}$ do not evolve the same way as a function of $Z_{\rm avg}$ (Fig. 11). We identify 4 regimes based on the metallicity dispersion ΔZ, most evident in the inference runs using the [S II] lines:

1) Smooth increase of ΔZ until $\approx 1/3\,Z_\odot$ by a factor of ≈2, with little dispersion across galaxies.
2) Sharp increase between $\approx 1/3\,Z_\odot$ and $\approx 1/2\,Z_\odot$ by a factor of ≈4, with a large dispersion across galaxies.
3) Turnoff until super-solar metallicity galaxies.
4) Sharp decrease until $\approx 2.5-3\,Z_\odot$ by a factor of ≈10 (only seen if using [S II] lines as constraints).

We emphasize that the small difference between $Z_{\rm min}$ and $Z_{\rm max}$ in low-metallicity galaxies is a direct result of the inference: other solutions may exist (such as a wide range of metallicity within galaxies together with a very low α_Z), but their likelihood is significantly lower. Furthermore, the trend observed for the metallicity dispersion to be small for either the lowest or highest metallicity galaxies is not due to potential edge effects as we do not observe the same behavior for the boundaries of other parameters as a function of the corresponding average parameter value.

In metal-poor galaxies, the small metallicity dispersion implies that the existence of numerous metal-rich regions in low-Z galaxies is unlikely. The relatively slow evolution of $Z_{\rm min}$ may indicate that metal-poor ($\lesssim 1/3\,Z_\odot$) gas remains present in metal-rich galaxies until at least an average metallicity about solar, but in the form of regions that do not contribute much to the total emission ($\alpha_Z > 0$). Inversely, the small dispersion inferred for the most metal-rich galaxies is only seen for inference runs using [S II] and seems in contradiction with the evidence of relatively metal-poor regions in metal-rich galaxies (e.g., Poetrodjojo et al. 2018). The inference runs ignoring [S II] do predict a relatively large dispersion instead.

Assuming that the sample at $z \sim 0$ may capture the evolution of galaxies versus Z (i.e., assuming a closed-box scenario for which the average metallicity increases monotonously with time and also assuming that metal-poor galaxies are past versions of metal-rich ones), the fact that $Z_{\rm max}$ increases relatively faster (factor of ≈10 between $\approx 1/3$ and $\approx 1/2\,Z_\odot$) than $Z_{\rm min}$ might indicate a faster enrichment of metal-rich regions. One possible interpretation, assuming that the average Z traces an evolutionary pathway, is that

1) Galaxies start forming stars in a gas whose metallicity is relatively uniform and metal-poor (average metallicity below $\lesssim 1/5\,Z_\odot$),
2) Star-formation is slightly more efficient in regions already enriched in heavy elements (e.g., due to increased cooling) leading to an increasing offset between the maximum and minimum metallicity within the galaxy and to the average metallicity of the galaxy being driven by metal-rich regions,
3) The enrichment of the most metal-rich regions eventually plateaus around solar metallicity, which could be due to the fact that the added metal mass released through a typical star-formation episode becomes small compared to the existing metal content.

A symmetric behavior is observed for $Z_{\rm min}$ and $Z_{\rm max}$, with two reference metallicity thresholds: $1/3\,Z_\odot$ corresponding to a sharp increase in the metal enrichment, and $\sim Z_\odot$ corresponding to a saturation in enrichment.

5.2.2. Internal distribution of other parameters

The ionization parameter dispersion within galaxies ($U_{\rm max} - U_{\rm min}$) is the largest within the most metal-poor galaxies and decreases sharply until $\sim 1/2\,Z_\odot$ (Fig. 11). Since the density boundaries do not evolve much versus Z, the wide range of U in metal-poor galaxies could be due to a wide range of the ionizing photon flux and/or the distance between the stars and the illuminated gas shells. We remark that the lower boundary drives the average ionization parameter (due to the negative slope α_U (Sect. 4.4.1).

There is no clear evolution of the density boundaries versus Z apart from a slight decrease of both boundaries. As a consequence, there is also little evolution of the density dispersion ($n_{\rm max} - n_{\rm min}$). The age boundaries tightly follow each other versus Z and, as a consequence, the difference depends relatively little on Z.

5.3. On a potential metallicity bimodality

We find a bimodal distribution of the average metallicity $Z_{\rm avg}$ in the galaxies of the ECO sample when using all available lines (Fig. 9). The low-metallicity probability peak, where most galaxies lie, is centered around 12 + log(O/H) ≈ 8.25 (≈$0.3\,Z_\odot$) and the high-metallicity peak reaches up to 12 + log(O/H) ≈ 9.0. However, if [S II] lines are ignored for the inference (for the LOC approach or single 1D model alike), the bimodality is much weaker and the secondary peak lies around the solar value. The bimodality is driven by the upper boundary $Z_{\rm max}$ (Sect. 4.4.1), while the lower boundary $Z_{\rm min}$ hardly seems to reach the metallicity threshold for rapid enrichment (Sect. 5.2.1).

The N2S2 empirical diagnostic (Dopita et al. 2016) provides a significantly smoother PDF (Fig. A.1) and provides metallicities as low as $\approx 1/30\,Z_\odot$ while the metallicity we infer does not reach below $1/10\,Z_\odot$. We emphasize that the lines involved in the N2S2 diagnostic are not particularly well reproduced by the various models we consider and that the difference seen for the PDFs may be partly due to the inability of the models to reproduce better [S II] and/or to systematic effects in the line measurement available in the SDSS catalogs (Sect. 4.1). It also shows that the bimodality in Z, if real, may be difficult to identify solely based on empirical diagnostics.

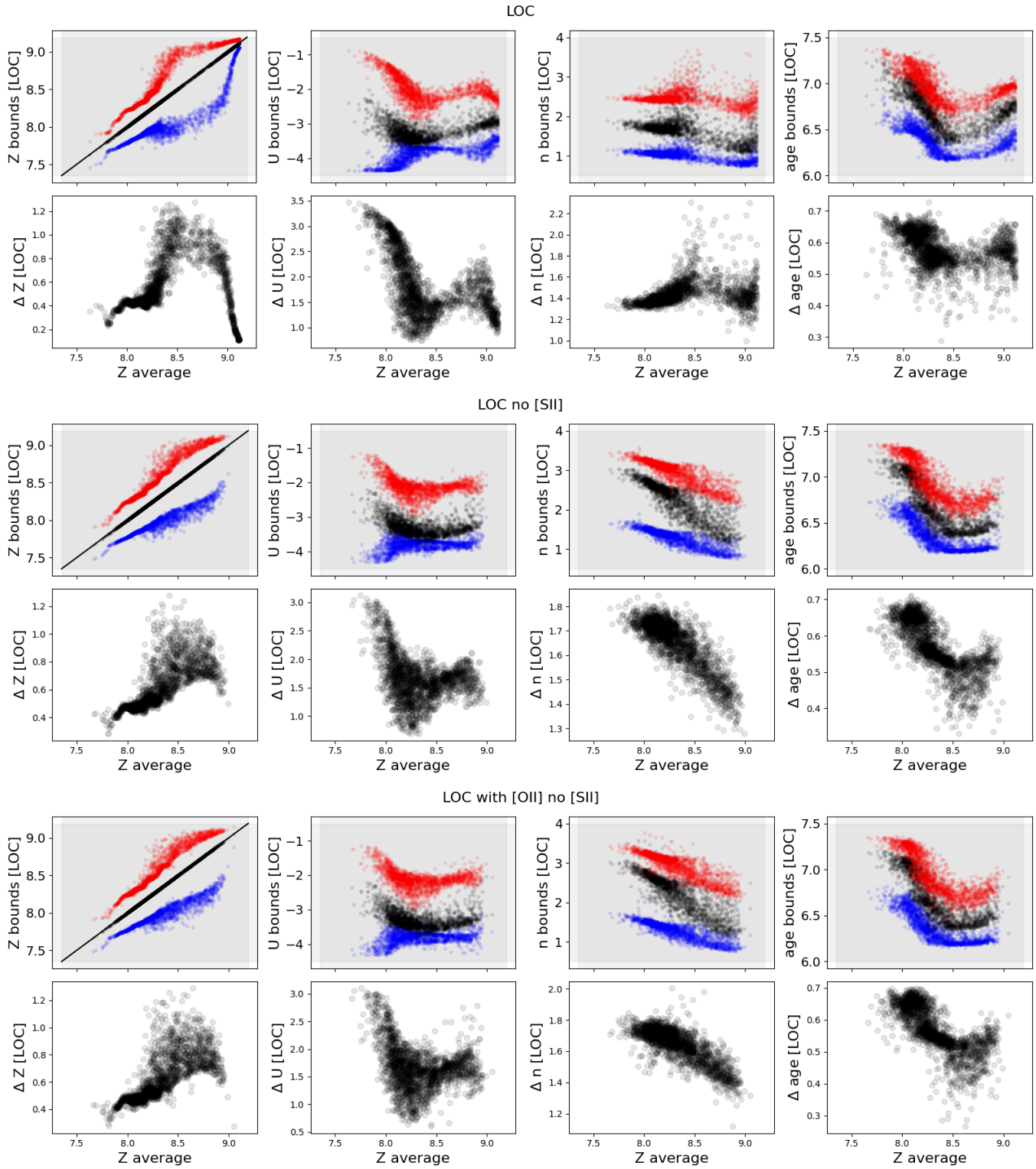

Fig. 11. Evolution of the parameter boundaries $p_{\rm min,max}$ versus the average metallicity $Z_{\rm avg}$ for the three inference runs (including [S II] on top, ignoring [S II] in the middle, and replacing [S II] by [O II] in the bottom). For each run, the upper row shows $p_{\rm min,max}$ in blue and red respectively (with the gray rectangles showing the full parameter range in the grid), while the bottom row shows the difference between $p_{\rm max} - p_{\rm min}$.

We tested inference runs with different underlying photoionization grids and including the [S II] lines: BOND (Vale Asari et al. 2016) and SFGX (Ramambason et al. 2022). Although not shown, the PDFs from SFGX and from the present grid are similar and both show a bimodality. The bimodality is more pronounced with the present grid because the maximum metallicity in SFGX is only 0.1 log solar. BOND does reach higher values but does not show any bimodality, and in fact provides similar results to the N2S2 calibration. We conclude that the Z bimodality is mostly driven by the grid presently used and the underlying abundance patterns that are assumed (similar prescription in the current grid and SFGX, mostly drawn from

Nicholls et al. 2017). Ignoring the [S II] lines in the inference somewhat mitigates these issues, implying that the bimodality, if real, is likely not a strong one.

For completeness, although the present sample was selected to be compatible with star-forming criteria and negligible AGN contamination (Sect. 3.1), we cannot exclude that high-Z galaxies may correspond to sources with a contribution from ionization mechanisms other than UV photoionization from massive stars (e.g., shocks and/or AGN). In fact, the high-Z sample is partly populated with "ambiguous" galaxies lying between the SF-AGN demarcation lines of Kauffmann et al. (2003) and Stasińska et al. (2006) and may therefore imply weak AGN contamination (see Fig. 3 and Appendix A). The interpretation of a potential metallicity bimodality in star-forming galaxies therefore depends heavily on the maximum AGN contamination allowed, especially if one considers that nuclear activity might never be null. Nevertheless, we do note that the "ambiguous" galaxies remain far off the AGN domain in the [O I] diagnostic plot (Fig. 3).

5.4. Mass-metallicity relationship

The ECO star-forming galaxy sample provides an opportunity to study the mass-metallicity relationship (MZR; see review in Maiolino & Mannucci 2019) in a consistent way for a wide range of galaxy mass. We use the metallicity inferred in the present study and the stellar mass is taken from Hutchens et al. (2023).

Figure 12 shows that the MZR is smooth and narrow until stellar masses 10^9 $M_\odot$, but there is a large spread of metallicities in the range $10^{9.5-10}$ $M_\odot$, which eventually leads to a high-metallicity plateau for the most massive galaxies. The Z and M_* PDFs are remarkably different, the former being somewhat bimodal (Sect. 5.3) and the latter showing a single-peak broad distribution.

The main result shows that the inferred MZR in the low-mass regime ($\lesssim 10^{9.5}$ $M_\odot$, i.e., the bulk of the ECO sample) is compatible with the low-mass galaxy fit in Indahl et al. (2021), which uses the robust direct method (T_e-method with calibrated ionization correction factors; see also Berg et al. 2012; Kirby et al. 2013). This agreement with Indahl et al. (2021) as well as with stellar abundances in Maiolino & Mannucci (2019) suggests that the inferred metallicity (i.e., using photoionization models and strong lines) are reliable. Figure 13 (left panel) shows that the low-mass fit remains unchanged whether [S II] lines are used as constraints or not. Figure 13 (right panel) shows that the MZR in the low-mass regime using the empirical N2S2 diagnostic agrees less well with Indahl et al. (2021), suggesting that the [S II] line measurement may lead to systematics (Sect. 4.1).

Concerning high-mass galaxies ($\gtrsim 10^{9.5}$ $M_\odot$), we find significant lower metallicities than both Tremonti et al. (2004) and Mingozzi et al. (2020), the latter studies using strong lines and the theoretical method (stellar population + photoionization grids). We argue that this may be a consequence of ignoring the [S II] lines for inference in the present study, as including these lines results in significantly higher metallicities (Fig. 13 left panel), at the expense of a strong metallicity bimodality. Our fit for high-mass galaxies ignoring [S II] lines is in line with the study of Andrews & Martini (2013) which uses the direct method on stacks, strengthening the reliability of our inferred metallicity across the full mass range.

In summary, the MZR we infer without [S II] lines is in line with studies using the direct method from the calibrated range up to methods using stacks, hinting that our models are able

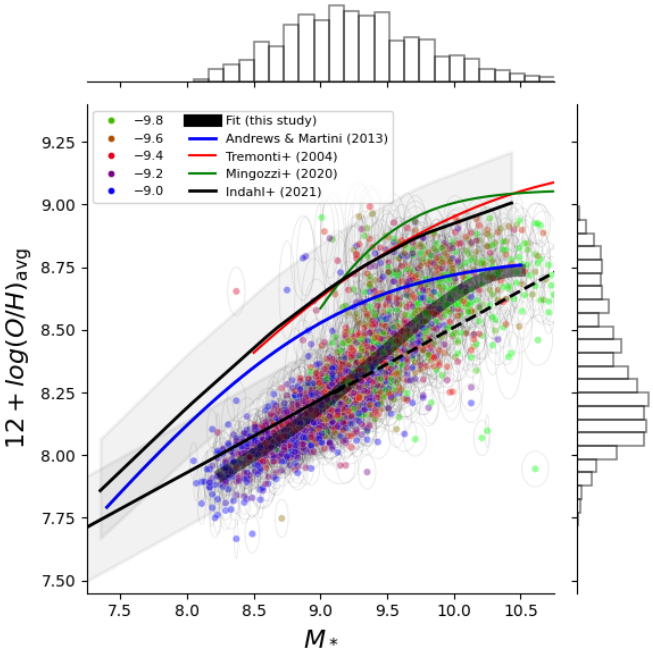

Fig. 12. Metallicity-mass relationship using the metallicity inferred ignoring the [S II] lines. The thick gray curve shows the 4th order polynomial fit (see text). The blue, red, and green curves show the correlations from Andrews & Martini (2013), Tremonti et al. (2004), and Mingozzi et al. (2020) respectively, while the bottom and top stripes show the low-mass galaxy fit and the SDSS star-forming galaxy fit from Indahl et al. (2021) respectively. The dashed line shows the extrapolation from Indahl et al. (2021).

to capture the physical conditions of the gas. Our sample was indeed selected to ensure sufficient S/N in the strongest lines but we did not consider T_e-sensitive auroral lines (Sect. 3.1). Among these, the [O III] λ4363 line is detected in only ≈17% and ≈7% of the sample above 2σ and 3σ respectively. Nevertheless, we have verified that the model predictions for this line (i.e., not using it for inference) agree within 2σ for the galaxies with detections. We also verified that using it for the inference does not modify our results across the mass and metallicity ranges.

The fourth order polynomial fit, valid in the range $\log M* = [8.25, 10.5]$, provides

$$12 + \log(\mathrm{O/H}) \approx -0.035756 M_*^4 + 1.25737 M_*^3 - 16.4913 M_*^2 + 95.9941 M_* - 201.99. \quad (10)$$

The MZR is often reported in the literature to show a transition between a positive correlation to an almost constant metallicity, thought to be the consequence of galactic outflow efficiency versus galaxy mass (e.g., Tremonti et al. 2004; De Vis et al. 2019), i.e., with no gap or sharp transition between two metallicities. A bimodality in galaxy parameters is, however, known to exist between blue star-forming disks and red spheroids dominated by old stellar populations, with a mass transition ≈$10^{10.5}$ $M_\odot$ and has been attributed to cold flows and shock heated streams (e.g., Dekel & Birnboim 2008). Various studies have proposed "inverse" morphological transformations (from early- to late-type) through a disk regrowth process possibly enabled by gas accretion, which may explain the existence of non-cluster "blue-sequence" E/S0 galaxy population as well as extended UV emission around some early-type galaxies (see, e.g., Stark et al. 2013; Moffett et al. 2015). This led Kannappan et al. (2013) to hypothesize that the transition is due

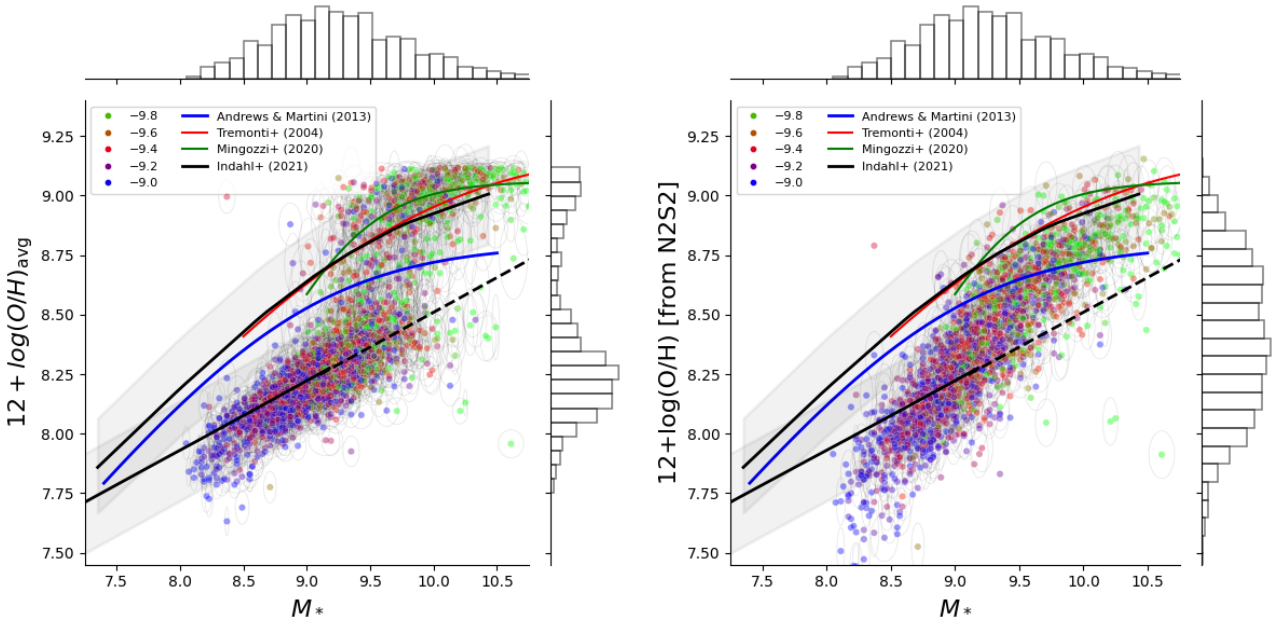

Fig. 13. Metallicity-mass relationship using the metallicity inferred with MULTIGRIS including [S II] lines (left) and using the metallicity calculated from N2S2 empirical calibration (right).

to a different refueling regime with high levels of external gas accretion and stellar mass growth. The "blue-sequence" E/S0 galaxy population a population exists primarily below a stellar mass of $\sim 10^{9.7-10.5}$ $M_\odot$ and corresponds well to the masses for which the average metallicity we infer increases sharply, potentially suggesting a higher star-formation efficiency. The present results unfortunately do not allow us to distinguish between a higher star-formation efficiency due to external processes (accretion) or internal processes (metallicity threshold; Sect. 5.2.1).

6. Conclusions

We present models of star-forming galaxies from the volume-limited ECO catalog. The main objective is to interpret realistic models of an unbiased sample in order to probe relationships between physical parameters, in particular as a function of metallicity, but also to investigate the metallicity probability density function itself and to recover the internal distribution of physical parameters within galaxies. In summary:

- We designed a framework using probabilistic methods in order to assess various model architectures meant to describe the emitting components of a galaxy. In particular, we considered the combination of many 1D models, that is, the LOC hypothesis. LOC architectures integrate a number of models with different physical properties linked by a given distribution (e.g., a power law) whose parameters are found through inference.
- We applied this framework to the ECO star-forming galaxy sample. We focused on a few model architectures, including a single 1D model approach for comparison. The 1D models, used as single models or within an LOC combination, were computed with Cloudy with specific abundance patterns as a function of metallicity and we made use of the stellar population synthesis code BPASS.
- Guided by potential issues with the line measurements as well as with the model hypotheses, we performed runs ignoring [S II], which globally performed much better and alleviated some issues with [N II] and [O I] predictions.

The main results are as follows:

1. Globally, we find that the LOC models outperform the single 1D models, strengthening the need for relatively complex and realistic architectures. The single 1D models provide values for physical parameters that are close to the average value considering a distribution of components (LOC) – which we consider robust – despite the small biases observed and discussed.
2. For LOC models, the average physical parameter value in a galaxy is always tightly constrained. Other distribution hyperparameters (the slope and boundaries) are much less well constrained but do show small deviations from galaxy to galaxy and with respect to the prior, suggesting that it is possible and meaningful to study these variations for a physical interpretation.
3. We find, in particular, that the integrated emission of galaxies is dominated by relatively low-excitation gas, with an average $U \sim -3.2$. The age distribution peaks around 5 Myr, with the lower and upper boundary around 2 and 10−20 Myr, respectively.
4. The average metallicity shows a weakly bimodal distribution, with most galaxies showing an average metallicity of $\sim 0.3\,Z_\odot$ and a secondary peak around solar metallicity.
5. The lower and upper metallicity boundaries within galaxies do not evolve the same way as a function of the average metallicity. In the most metal-poor galaxies, most emitting components have the same metallicity within a factor of 2−3. As the metallicity increases until about solar values, the most metal-rich regions increase their metallicity sharply while low-metallicity regions remain constant, resulting in a metallicity dispersion up to a factor 5 − 10. For super-solar metallicity galaxies, the most metal-poor regions finally get enriched. We propose that this reflects an evolutionary sequence involving a combination of metallicity thresholds for efficient star formation ($\approx 1/3\,Z_\odot$) and saturation ($\approx Z_\odot$).
6. The average metallicity bimodality is driven by the upper boundary $Z_{\rm max}$ and the secondary peak could be a consequence of efficient or rapid enrichment of the most metal-rich regions.

7. We find correlations between all parameters (age, ionization parameter, and – though to a lesser extent – density) versus Z, with the lowest metallicity galaxies having a younger age, higher density, and higher ionization parameter. We find, however, a flattening of age and U for galaxies above $\sim 0.5\,Z_\odot$.
8. Finally, we examined the MZR and find results in line with direct abundance method determinations, from the calibrated range at a low metallicity to methods using stacks at a high metallicity. This suggests that the models are able to capture physical conditions of the gas and that the inferred metallicity is reliable. We identified two regimes, the low-mass regime below $\sim 10^{9.5}\,M_\odot$, reproducing the low-mass galaxy fit from Indahl et al. (2021), and a sharp metallicity increase for more massive galaxies. This transition may be related to a specific refueling of non-cluster early-type galaxies but we cannot exclude purely internal processes such as a metallicity threshold for efficient star formation.

Acknowledgements. This work is supported by the FACE Foundation Transatlantic Research Partnership Fund (award TJF21_053). CR gratefully acknowledges the support of the Elon University Japheth E. Rawls Professorship, Elon University's FR&D committee, and the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS), which is supported by the National Science Foundation. This work used the ACCESS resource Expanse at the San Diego Supercomputing Center through allocation AST140040. LR gratefully acknowledges funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through an Emmy Noether Research Group (grant number CH2137/1-1).

References

Abel, N. P., van Hoof, P. A. M., Shaw, G., Ferland, G. J., & Elwert, T. 2008, ApJ, 686, 1125
Andrews, B. H., & Martini, P. 2013, ApJ, 765, 140
Baes, M., Davies, J. I., Dejonghe, H., et al. 2003, MNRAS, 343, 1081
Baes, M., Verstappen, J., De Looze, I., et al. 2011, ApJS, 196, 22
Baldwin, J. A., Ferland, G. J., Martin, P. G., et al. 1991, ApJ, 374, 580
Balser, D. S., Rood, R. T., Bania, T. M., & Anderson, L. D. 2011, ApJ, 738, 27
Berg, D. A., Skillman, E. D., Marble, A. R., et al. 2012, ApJ, 754, 98
Blanc, G. A., Kewley, L., Vogt, F. P. A., & Dopita, M. A. 2015, ApJ, 798, 99
Boquien, M., Burgarella, D., Roehlly, Y., et al. 2019, A&A, 622, A103
Bresolin, F. 2019, MNRAS, 488, 3826
Burgarella, D., Buat, V., & Iglesias-Páramo, J. 2005, MNRAS, 360, 1413
Carr, D. S., Kannappan, S. J., Norris, M. A., et al. 2024, ApJ, 968, 18
Carton, D., Brinchmann, J., Wang, J., et al. 2015, MNRAS, 451, 210
Chevallard, J., & Charlot, S. 2016, MNRAS, 462, 1415
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
D'Agostino, J. J., Kewley, L. J., Groves, B., et al. 2019, ApJ, 878, 2
De Looze, I., Cormier, D., Lebouteiller, V., et al. 2014, A&A, 568, A62
De Vis, P., Jones, A., Viaene, S., et al. 2019, A&A, 623, A5
Dekel, A., & Birnboim, Y. 2008, MNRAS, 383, 119
Del Moral, P., Doucet, A., & Jasra, A. 2012, Stat. Comput., 22, 1009
Della Bruna, L., Adamo, A., Lee, J. C., et al. 2021, A&A, 650, A103
Dopita, M. A., Fischera, J., Sutherland, R. S., et al. 2006a, ApJS, 167, 177
Dopita, M. A., Fischera, J., Sutherland, R. S., et al. 2006b, ApJ, 647, 244
Dopita, M. A., Kewley, L. J., Sutherland, R. S., & Nicholls, D. C. 2016, Ap&SS, 361, 61
Eldridge, J. J., Stanway, E. R., Xiao, L., et al. 2017, PASA, 34
Ferguson, J. W., Korista, K. T., Baldwin, J. A., & Ferland, G. J. 1997, ApJ, 487, 122
Fitzgerald, K., Harvey, E. J., Keaveney, N., & Redman, M. P. 2020, Astron. Comput., 32, 100382
Galliano, F., Nersesian, A., Bianchi, S., et al. 2021, A&A, 649, A18
Garg, P., Narayanan, D., Byler, N., et al. 2022, ApJ, 926, 80
Garg, P., Narayanan, D., Sanders, R. L., et al. 2024, ApJ, 972, 113
Garner, R., Kennicutt, R., Rousseau-Nepton, L., et al. 2025, ApJ, 978, 70
Groves, B., Nefs, B., & Brandl, B. 2008, MNRAS, 391, L113
Hu, N., Wang, E., Lin, Z., et al. 2018, ApJ, 854, 68
Hutchens, Z. L., Kannappan, S. J., Berlind, A. A., et al. 2023, ApJ, 956, 51
Indahl, B., Zeimann, G., Hill, G. J., et al. 2021, ApJ, 916, 11
Ji, X., & Yan, R. 2022, A&A, 659, A112
Jin, Y., Kewley, L. J., & Sutherland, R. 2022, ApJ, 927, 37
Kang, D. E., Pellegrini, E. W., Ardizzone, L., et al. 2022, MNRAS, 512, 617
Kannappan, S. J., Stark, D. V., Eckert, K. D., et al. 2013, ApJ, 777, 42
Kashino, D., & Inoue, A. K. 2019, MNRAS, 486, 1053
Katz, H. 2022, MNRAS, 512, 348
Katz, H., Liu, S., Kimm, T., et al. 2022, MNRAS, submitted [arXiv:2211.04626]
Katz, H., Rosdahl, J., Kimm, T., et al. 2023, Open J. Astrophys., 6, 44
Kauffmann, G., Heckman, T. M., Tremonti, C., et al. 2003, MNRAS, 346, 1055
Kewley, L. J., Dopita, M. A., Sutherland, R. S., Heisler, C. A., & Trevena, J. 2001, ApJ, 556, 121
Kewley, L. J., Nicholls, D. C., & Sutherland, R. S. 2019, ARA&A, 57, 511
Khullar, S., Federrath, C., Krumholz, M. R., & Matzner, C. D. 2021, MNRAS, 507, 4335
Kirby, E. N., Cohen, J. G., Guhathakurta, P., et al. 2013, ApJ, 779, 102
Kreckel, K., Ho, I. T., Blanc, G. A., et al. 2019, ApJ, 887, 80
Lebouteiller, V., & Ramambason, L. 2022, A&A, 667, A34
Lebouteiller, V., Péquignot, D., Cormier, D., et al. 2017, A&A, 602, A45
Lecroq, M., Charlot, S., Bressan, A., et al. 2024, MNRAS, 527, 9480
Madden, S. C., Cormier, D., Hony, S., et al. 2020, A&A, 643, A141
Maiolino, R., & Mannucci, F. 2019, A&ARv, 27, 3
Marconi, A., Amiri, A., Feltre, A., et al. 2024, A&A, 689, A78
Mingozzi, M., Belfiore, F., Cresci, G., et al. 2020, A&A, 636, A42
Moffett, A. J., Kannappan, S. J., Berlind, A. A., et al. 2015, ApJ, 812, 89
Morisset, C. 2013, pyCloudy: Tools to manage astronomical Cloudy photoionization code, Astrophysics Source Code Library, [record ascl:1304.020]
Morisset, C., Charlot, S., Sánchez, S. F., et al. 2025, MNRAS, accepted [arXiv:2501.05424]
Nakajima, K., Ouchi, M., Isobe, Y., et al. 2024, ApJ, submitted [arXiv:2412.04541]
Nicholls, D. C., Sutherland, R. S., Dopita, M. A., Kewley, L. J., & Groves, B. A. 2017, MNRAS, 466, 4403
Osterbrock, D. E., & Ferland, G. J. 2006, Astrophysics of Gaseous Nebulae and Active Galactic nuclei (Sausalito, CA: University Science Books)
Palud, P., Einig, L., Le Petit, F., et al. 2023, A&A, 678, A198
Pellegrini, E. W., Drory, N., Guillermo, A., et al. 2019, arXiv e-prints [arXiv:1905.00311]
Péquignot, D. 2008, A&A, 478, 371
Péquignot, D., Amara, M., Liu, X. W., et al. 2002, Rev. Mex. Astron. Astrofis. Conf. Ser., 12, 142
Poetrodjojo, H., Groves, B., Kewley, L. J., et al. 2018, MNRAS, 479, 5235
Polimera, M. S., Kannappan, S. J., Richardson, C. T., et al. 2022, ApJ, 931, 44
Ramambason, L., Schaerer, D., Stasińska, G., et al. 2020, A&A, 644, A21
Ramambason, L., Lebouteiller, V., Bik, A., et al. 2022, A&A, 667, A35
Ramambason, L., Lebouteiller, V., Madden, S. C., et al. 2024, A&A, 681, A14
Reddy, N. A., Topping, M. W., Sanders, R. L., Shapley, A. E., & Brammer, G. 2023, ApJ, 952, 167
Rémy-Ruyer, A., Madden, S. C., Galliano, F., et al. 2014, A&A, 563, A31
Richardson, C. T., Allen, J. T., Baldwin, J. A., Hewett, P. C., & Ferland, G. J. 2014, MNRAS, 437, 2376
Richardson, C. T., Allen, J. T., Baldwin, J. A., et al. 2016, MNRAS, 458, 988
Richardson, C. T., Polimera, M. S., Kannappan, S. J., Moffett, A. J., & Bittner, A. S. 2019, MNRAS, 486, 3541
Richardson, C. T., Simpson, C., Polimera, M. S., et al. 2022, ApJ, 927, 165
Riggs, C. L., Brooks, A. M., Munshi, F., et al. 2024, ApJ, 977, 20
Romero, M., Corcho-Caballero, P., Millán-Irigoyen, I., Mollá, M., & Ascasibar, Y. 2023, MNRAS, 521, 1727
Saintonge, A., & Catinella, B. 2022, ARA&A, 60, 319
Salvatier, J., Wiecki, T. V., & Fonnesbeck, C. 2016, PeerJ Comput. Sci., 2, e55
Santoro, F., Kreckel, K., Belfiore, F., et al. 2022, A&A, 658, A188
Sharda, P., Krumholz, M. R., Wisnioski, E., et al. 2021, MNRAS, 502, 5935
Sharda, P., Ginzburg, O., Krumholz, M. R., et al. 2024, MNRAS, 528, 2232
Simons, R. C., Papovich, C., Momcheva, I., et al. 2021, ApJ, 923, 203
Sorba, R., & Sawicki, M. 2015, MNRAS, 452, 235
Sorba, R., & Sawicki, M. 2018, MNRAS, 476, 1532
Stark, D. V., Kannappan, S. J., Wei, L. H., et al. 2013, ApJ, 769, 82
Stasińska, G., Cid Fernandes, R., Mateus, A., Sodré, L., & Asari, N. V. 2006, MNRAS, 371, 972
Stasińska, G., Izotov, Y., Morisset, C., & Guseva, N. 2015, A&A, 576, A83
Tacconi, L. J., Genzel, R., & Sternberg, A. 2020, ARA&A, 58, 157
Thomas, A. D., Groves, B. A., Sutherland, R. S., et al. 2016, ApJ, 833, 266
Tremonti, C. A., Heckman, T. M., Kauffmann, G., et al. 2004, ApJ, 613, 898
Vale Asari, N., Stasińska, G., Morisset, C., & Cid Fernandes, R. 2016, MNRAS, 460, 1739
Varese, M., Lebouteiller, V., Ramambason, L., et al. 2025, A&A, 693, A147
Williams, T. G., Kreckel, K., Belfiore, F., et al. 2022, MNRAS, 509, 1303
Yang, G., Boquien, M., Brandt, W. N., et al. 2022, ApJ, 927, 192

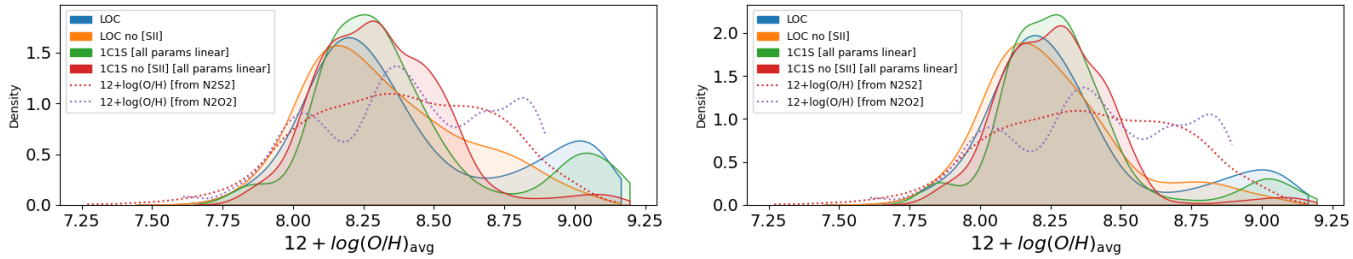

Fig. A.1. Left: average metallicity PDF including the N2S2 and N2O2 calibrations from Garg et al. (2024). Right: same as left plot but using the star-formation / AGN demarcation line of Stasińska et al. (2006) instead of Kauffmann et al. (2003).

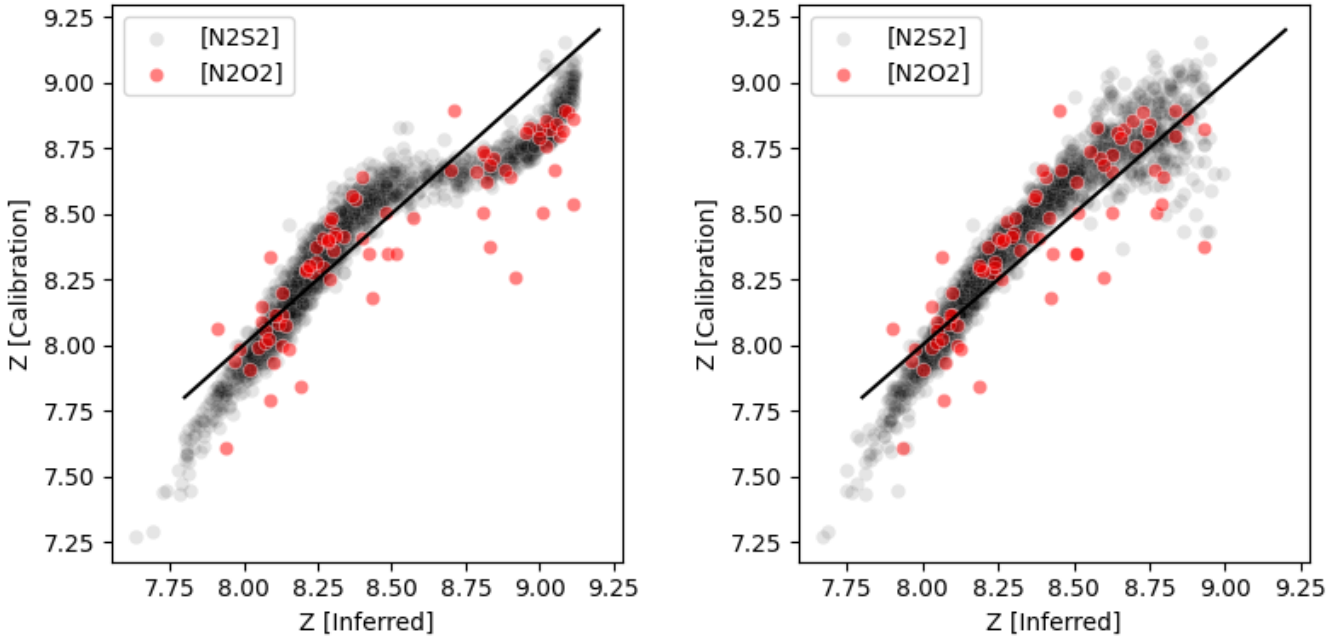

Fig. A.2. N2S2 (black) and N2O2 (red) empirical diagnostics versus the inferred (average) metallicity for the LOC architecture using [S II] lines (left) or ignoring them (right).

Appendix A: Empirical metallicity diagnostics and star-formation / AGN demarcation line

Figure A.1 shows the metallicity distribution inferred through various model architectures compared to the metallicity calculated from empirical diagnostics (N2S2 and N2O2). The N2S2 diagnostic is discussed in Sect. 5.3. The N2O2 diagnostic suffers from the weak statistics due to the low S/N in the [O II] line for most of the sample. Figure A.2 illustrates how the metallicity we infer differs from the N2S2 and N2O2 empirical diagnostics.

The sample used in this study was drawn from the ECO catalog, selecting only star-forming galaxies (Sect. 3.1). As explained in the main text, the exact choice of the demarcation line between gas excitation dominated by star-formation or not has some impact on the results. Stasińska et al. (2006) provided an updated demarcation compared to Kauffmann et al. (2003) to account for galaxies with weak AGN contribution (typically $\lesssim 3\%$). We have verified that the results presented in this study remain unchanged whatever the choice of the demarcation (in other words including or not galaxies with potentially weak AGN contributions). Most galaxies falling between the two demarcation lines are high-metallicity galaxies (Fig. 3) and the main impact of using the demarcation from Stasińska et al. (2006) instead of Kauffmann et al. (2003) is to reduce the statistics of the high-Z galaxies in the various plots, with no change to the actual trends. Unsurprisingly, the only diagnostic plot which changes significantly is the metallicity PDF itself. Results in Section 5.3 use the demarcation from Kauffmann et al. (2003) and indicate a potential secondary high-Z "peak" (highly populated if using [S II] lines for inference, weakly populated if ignoring [S II]). When using Stasińska et al. (2006) instead, we see that the secondary peak is much smaller (Fig. A.1).

Appendix B: Results for individual galaxies

Figure B.1 shows that the inferred average parameter values for each individual galaxy show no significant correlations or degeneracy within the confidence intervals. The inference method in MULTIGRIS makes use of the Sequential Monte-Carlo method from python package PyMC (Salvatier et al. 2016), which is well adapted to complex, potentially multimodal posterior distributions (see Lebouteiller & Ramambason 2022). We have used the largest possible number of particles to alleviate issues with minor modes being cropped.

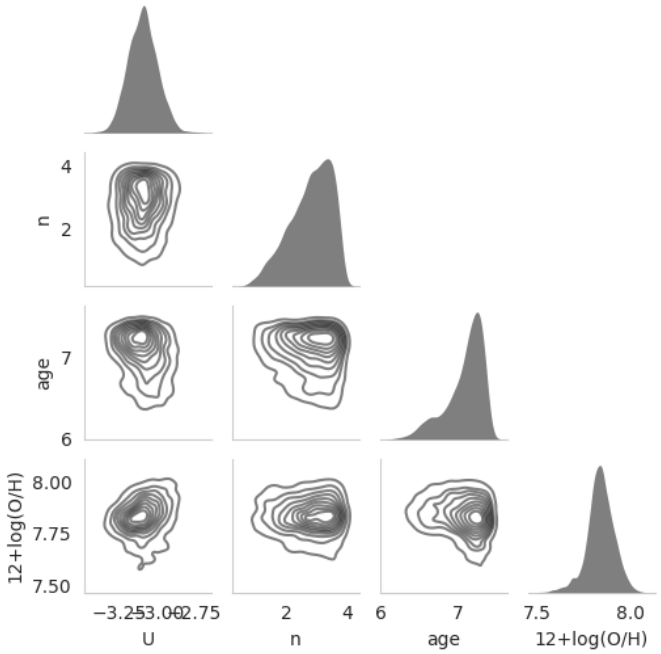

Fig. B.1. Illustration of the pairwise relationships between the inferred average parameter values for a low-metallicity source, illustrating that the posterior distributions do not show strong correlations.

Examples of inference results for individual galaxies are shown in Figure B.2. The average U and Z values within any galaxy (defined in Eqn. 5) are generally well constrained, and some boundaries (e.g., upper boundary $U_{\rm max}$) are relatively uncertain. However, it is worth noting that there is often minimal overlap between the lower and upper boundaries for each parameter, reflecting the fact that, given the choice between a single valued parameter or an LOC distribution, the latter is always preferred and thus likely to represent a more realistic model architecture.

Figure B.3 shows the parameter distribution for each galaxy (not to be confused with the parameter distribution of the average values, shown in Fig. 9). This figure shows that the PDF of some parameters is driven by the actual variation between galaxies (e.g., average metallicity) while the PDF of some other parameters is somewhat driven by a similar PDF inferred for all galaxies (e.g., age dispersion).

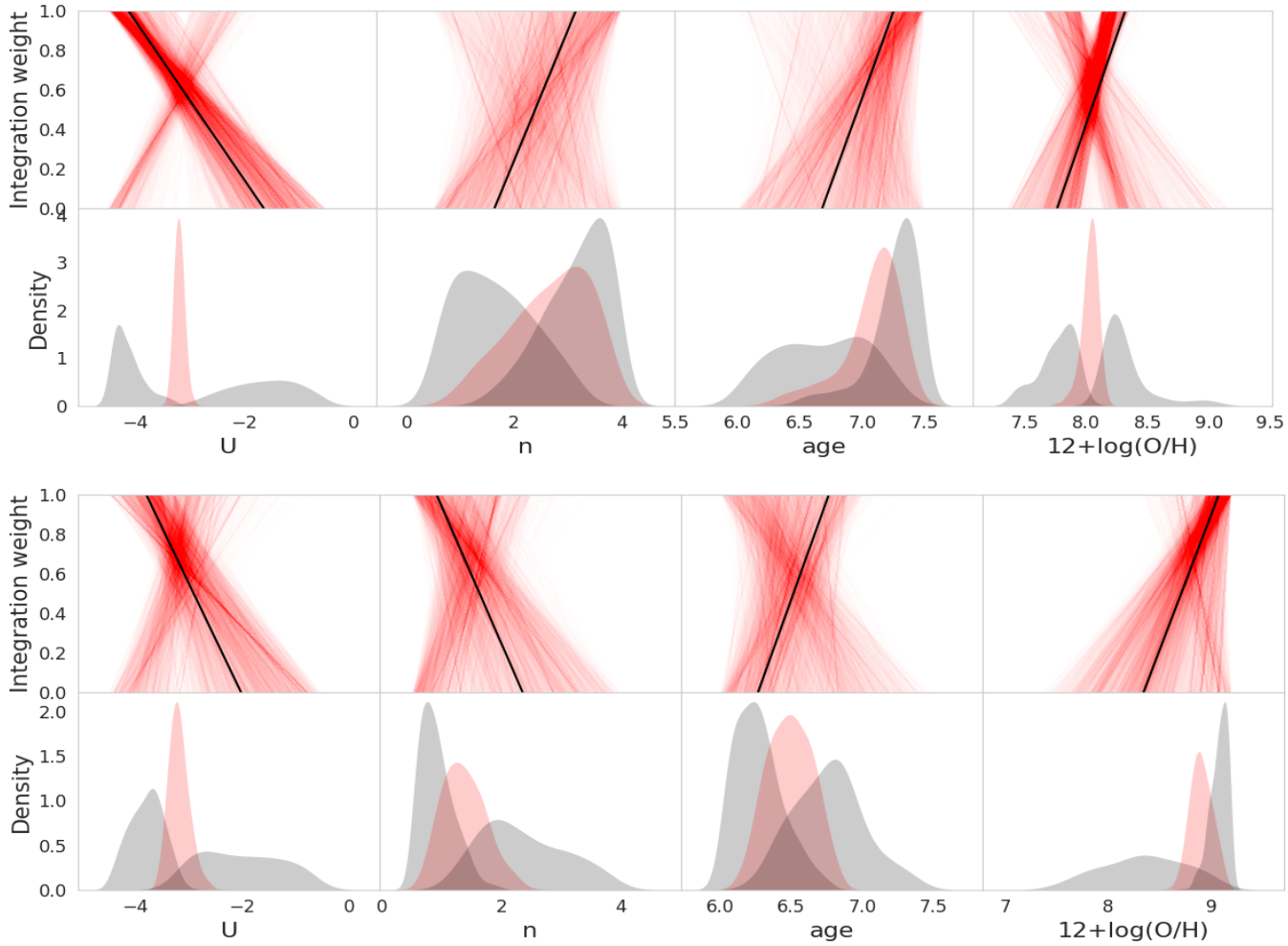

Fig. B.2. Illustration of inferred power-law distributions for typical subsolar metallicity (on top) and super-solar metallicity (on bottom) galaxies. For each metallicity case we show the individual draws for the integration weight ($\Phi(p)$; Eqn. 2) on top and the PDF for the lower and upper boundaries for integration ($p_{\rm min,max}$; gray) and the average parameter value ($p_{\rm avg}$; red) the bottom).

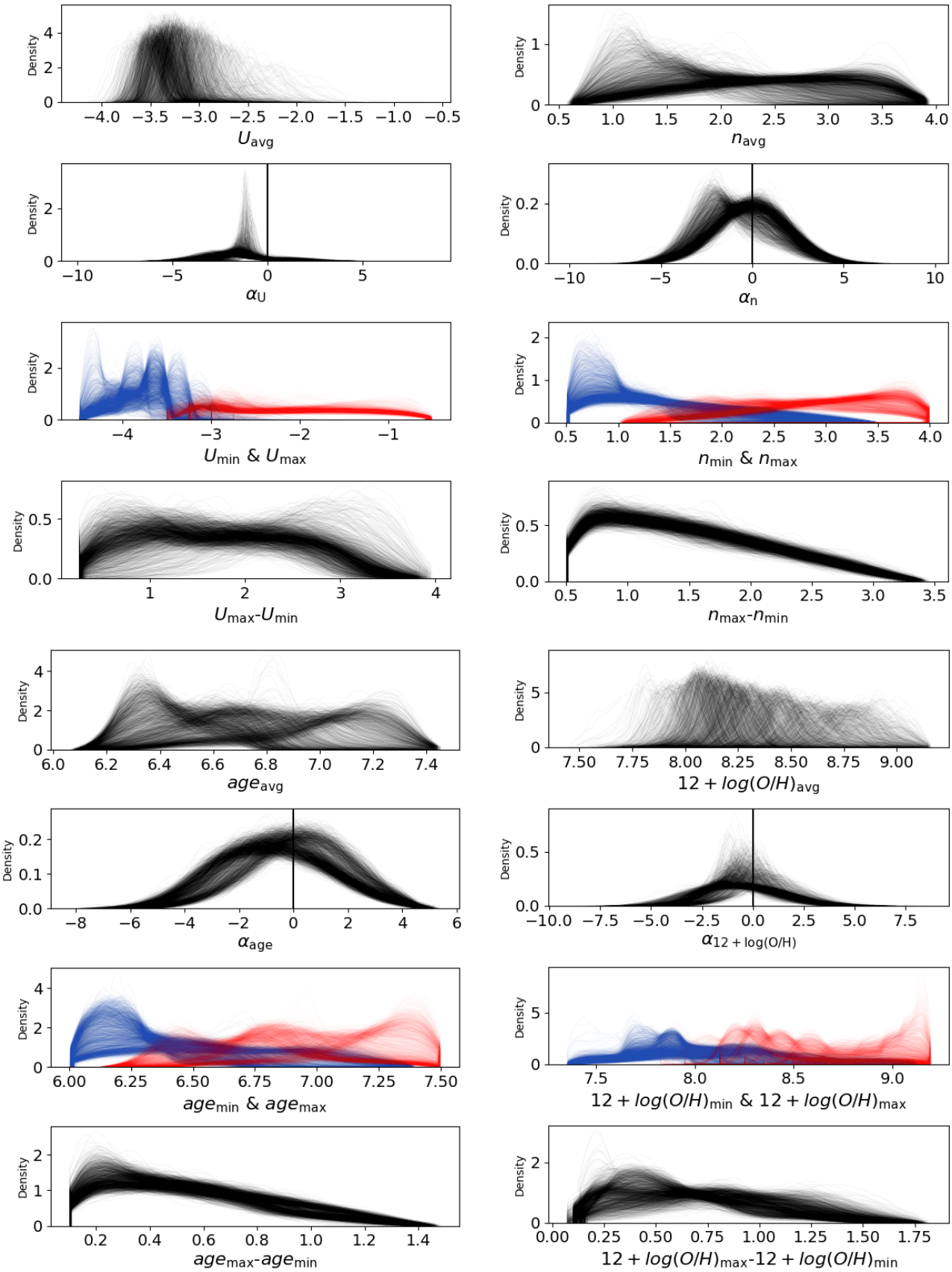

Fig. B.3. Illustration of the parameter distribution for each individual galaxy (here LOC approach ignoring [S II] lines; see Sect. 4.1). Each individual curve represents the PDF of a single galaxy. We plot the lower and upper boundaries in different colors for clarity.