観測統計量と
経験的モデルによる
銀河の化学進化

Chemical Evolution of Galaxies through Observational Statistics and Empirical Modeling

博士(理学)

Doctor of Philosophy

西垣 萌香

NISHIGAKI Moka

Department of Astronomical Science, School of Physical Sciences
The Graduate University for Advanced Studies, SOKENDAI
March 2026

Abstract

The overarching goal of this thesis is to characterize the cosmic evolution of the baryon cycle, the fundamental physical process governing galaxy formation. As a tracer for this process, we focus on gas-phase metallicity, which is readily observable and sensitively reflects the physical state of galaxies, approaching this problem from both observational and theoretical standpoints.

First, we focus on extremely metal-poor galaxies (EMPGs) in the local universe. Using a machine-learning technique, we select EMPG candidates from the extensive imaging data of the Sloan Digital Sky Survey (SDSS) and conduct follow-up spectroscopy with the Seimei, Nayuta, and Magellan telescopes. Consequently, we identify three new EMPGs and construct a statistically significant sample of 80 EMPGs. Despite this extensive search, no galaxies with metallicities below 1% solar are found, placing constraints on the metallicity floor in the local universe. Our statistical analysis reveals that the mass-metallicity-SFR relation of local EMPGs can be explained by the same equilibrium model that describes normal local galaxies, suggesting that the equilibrium state of the baryon cycle extends even in the most chemically primitive galaxies in the local universe. Furthermore, extending this search methodology to the early universe using JWST imaging data, we identify 17 EMPG candidates at $z \sim 4$–5 and conduct spectroscopic follow-up observations with the Subaru/FOCAS to confirm their redshifts.

Next, we systematically investigate line ratios and direct T_e-based metallicities combining data from the DREAMS, JADES, and CEERS surveys. By stacking spectra to detect faint auroral lines, we derive robust metallicities using the direct method. The results reveal that the fundamental metallicity relation (FMR) breaks at redshifts $z \gtrsim 8$, where galaxies exhibit systematically lower metallicities than expected from their stellar mass and star formation rate. This deviation suggests that the baryon cycle in the early universe is in a non-equilibrium state, distinct from the lower redshifts.

Finally, to provide a physical interpretation of these results, we construct a new empirical model, the ChemicalUniverseMachine, which integrates chemical evolution into dark matter halo growth histories. This model quantitatively demonstrates that the observed metallicity deficit at $z \gtrsim 8$ cannot be explained by standard extrapolations. Instead, reproducing the observations requires an evolution in the baryon cycle, specifically either an enhancement in outflow efficiency or an increase in gas inflow relative to star formation. By integrating these results, this thesis presents a unified picture of evolution of the baryon cycle, transitioning from a primordial, non-equilibrium state in the early universe to the equilibrium state established today.

Acknowledgements

The completion of this thesis would not have been possible without the support and encouragement of many people. I would like to express my deepest gratitude to my supervisor, Prof. Masami Ouchi, for his continuous support, invaluable guidance, and patience throughout my Ph.D. studies. His mentorship has been instrumental in shaping this thesis and will guide me throughout my career.

I am also grateful to the members of my thesis committee, Profs. Yusei Koyama, Koichiro Nakanishi, Yuichi Matsuda, Masahiro Nagashima, and Mariko Kubo, for their valuable feedback on this thesis.

I am deeply grateful to Prof. Kimihiko Nakajima, my co-supervisor, for his insightful advice and discussions over the past five years. I am also indebted to Prof. Peter Behroozi for our valuable collaboration on theoretical modeling. I would like to thank Dr. Yoshiaki Ono and Prof. Hidenobu Yajima for their helpful suggestions on the manuscript and theoretical interpretations. I also acknowledge the other 22 collaborators for their contributions to the projects included in this thesis. I appreciate the helpful advice and encouragement provided by Profs. Akio Inoue, Kentaro Nagamine, Hideko Nomura, Masayuki Tanaka, and Kazunori Kohri, as well as my co-supervisors, Prof. Tadafumi Takata and Dr. Nanase Harada.

I am thankful to all the members of the Ouchi Group, including Drs. Daichi Kashino, Yuichi Harikane, Haruka Kusakabe, Satoshi Kikuta, Yuki Isobe, Yongming Liang, Dongsheng Sun, Thai Tran, and Yi Xu, as well as Hiroya Umeda, Akinori Matsumoto, Kuria Watanabe, Chenghao Zhu, Hiroto Yanagisawa, Minami Nakane, Yui Takeda, Tomokazu Kiyoto, Yuta Kageura, Kana Takechi, and Yoshimi Matsuura, for the stimulating research environment and the enjoyable social gatherings.

I would also like to thank my friends at NAOJ, including Dr. Tomohiro Yoshida, Ryota Ikeda, Shun Ishigami, Shun Hatano, Sena Matsui, Hiroko Okada, Toya Suzuki, and Shiori Koshisaka for daily discussions and friendship. I also thank the members of Prof. Peter Behroozi's group at the University of Arizona, including Dr. Haowen Zhang, Haley Bowden, and Elaheh Hayati, for their warm hospitality during my visit.

I gratefully acknowledge financial support from JST, the establishment of university fellowships towards the creation of science technology innovation, Grant Number JPMJFS2136, and partial support from the Overseas Travel Fund for Students (2024) of the Astronomical Science Program, SOKENDAI. I also wish to thank the staff of the SOKENDAI and the Division of Science for their administrative assistance.

Finally, I would like to express my heartfelt thanks to my parents and my partner. I truly appreciate their warm support and understanding throughout this journey.

List of Publications

This thesis is mainly based on the following three papers:

1. **EMPRESS. XI. SDSS and JWST Search for Local and $z \sim 4-5$ Extremely Metal-poor Galaxies (EMPGs): Clustering and Chemical Properties of Local EMPGs**
   Moka Nishigaki, Masami Ouchi, Kimihiko Nakajima, Yoshiaki Ono, Michael Rauch, Yuki Isobe, Yuichi Harikane, et al.
   The Astrophysical Journal, 952, 11 (2023)
   DOI: 10.3847/1538-4357/accf14
   (This paper corresponds to Chapters 1, 2, 3, and 6)

2. **ChemicalUniverseMachine. I. Uncovering the Cosmic Evolution of Metals in the Galaxy–Interstellar Medium–Circumgalactic Medium Ecosystem**
   Moka Nishigaki, Peter Behroozi, Masami Ouchi, Hong Guo, Rachel S. Somerville, Anna R. Gallazzi, Kimihiko Nakajima, and Kuria Watanabe
   The Astrophysical Journal, 984, 135 (2025)
   DOI: 10.3847/1538-4357/adc43c
   (This paper corresponds to Chapters 1, 5, 6, and Appendix A)

3. **DREAMS. II. Galaxy Demographics from Direct T_e-Based Metallicities at $z \sim 2-10$: Tracing the Evolution of the Mass-Metallicity and Fundamental Relations**
   Moka Nishigaki, Kimihiko Nakajima, Masami Ouchi, Peter Behroozi, Minami Nakane, Yui Takeda, Hiroya Umeda, Hidenobu Yajima, and Hiroto Yanagisawa
   Submitted to The Astrophysical Journal, arXiv:2512.12983
   (This paper corresponds to Chapters 1, 4, and 6)

Contents

Abstract i

Acknowledgements iii

List of Publications v

Chapter 1 Introduction 1
1.1 Background . . . 1
1.2 Fundamental Physics . . . 2
1.2.1 Electron Temperature . . . 2
1.2.2 Ionization Parameter . . . 3
1.2.3 Gas-Phase Metallicity . . . 3
1.2.4 Equilibrium Model . . . 4
1.3 Key Scientific Questions . . . 6
1.3.1 Origin of Local EMPGs and the Metallicity Floor Problem 6
1.3.2 Cosmic Evolution of the Fundamental Metallicity Relation 7
1.3.3 Empirical Model for the Cosmic Evolution of Chemical Enrichment . . . 8
1.4 Structure of This Thesis . . . 9

Chapter 2 Extremely Metal Poor Galaxies in the Local Universe 13
2.1 Selection Method . . . 13
2.1.1 Developing Machine Learning Classifier . . . 13
2.1.2 Testing the Classifier with the SDSS Spectroscopic Data . 16
2.1.3 Comparison between Classifiers . . . 17
2.2 Photometric Selection . . . 18
2.3 Spectroscopic Screening . . . 18
2.3.1 Screening with 3.8m Seimei . . . 18
2.3.2 Screening with 2m Nayuta . . . 19
2.3.3 Results of the Spectroscopic Screening Observations . . . 19
2.4 Spectroscopic Confirmation of Metallicity . . . 19
2.4.1 MagE Observation and Data Reduction . . . 19
2.4.2 Metallicities . . . 22
2.5 EMPG Sample . . . 24
2.6 Chemical Properties . . . 24

Chapter 3 Searching for High-z EMPGs 29

3.1 Extending the Technique to the High-z EMPG Search 29
3.1.1 SED Model 30
3.1.2 JWST Data 30
3.1.3 Candidate Selection 30
3.1.4 Line Ratio Estimation 34
3.2 FOCAS follow-up observations 36

Chapter 4 Chemical Evolution of Galaxies: Observations 39
4.1 Data 39
4.1.1 DREAMS 39
4.1.2 JADES 40
4.1.3 ERO and CEERS 40
4.1.4 Galaxies at $z > 9.5$ 41
4.1.5 Emission-line Flux Measurements 41
4.1.6 Sample Selection 43
4.2 Method 43
4.2.1 Stellar Masses and SFRs 44
4.2.2 Stacking Analysis 46
4.2.3 Uncertainties 48
4.2.4 Photoionization Models 48
4.3 Results 49
4.3.1 Strong Line Ratios 49
4.3.2 Direct Method Metallicity 50
4.3.3 Comparing Direct-metallicity Results to Empirical Strong-line Calibrations 50
4.3.4 Interpreting the Evolution of Strong-line Ratios with Photoionization Models 52
4.3.5 Electron Density 54
4.3.6 Mass-Metallicity Relation 54
4.3.7 Fundamental Metallicity Relation 55

Chapter 5 Chemical Evolution of Galaxies: Modeling 63
5.1 Method 63
5.1.1 Method Overview 63
5.1.2 Multi-phase ISM Overview 64
5.1.3 Dark Matter Simulation 64
5.1.4 UniverseMachine 65
5.1.5 NeutralUniverseMachine 66
5.1.6 Computing Metallicity Evolution 67
5.1.7 Accounting for Halo Growth and Mergers 69
5.1.8 Inflows and CGM Metallicity 69
5.1.9 Well-Mixed ISM Model 71
5.1.10 Multi-phase Metallicity Model 72
5.1.11 Model Fitting and Posteriors 72
5.2 Observational Data 73
5.2.1 Metallicity Measurements in Star-forming Regions 73

5.2.2 Metallicity Measurements in Atomic Gas 75
5.2.3 Metallicity Measurements in the CGM 75
5.3 Results 76
5.3.1 Fitting Results 76
5.3.2 Distribution of Metals between the ISM and CGM 77
5.3.3 Systematic Uncertainties 77

Chapter 6 Discussion 83
6.1 Baryon Cycle in the Local Universe: An Equilibrium State . . . 83
6.2 Baryon Cycle in the Early Universe: Signs of a Non-Equilibrium State 84
6.2.1 Enhanced Gas Cycling via Inflows and Outflows 84
6.2.2 Transition Driven by Cosmic Reionization 84
6.3 Physical Interpretation of the Baryon Cycle Evolution 85
6.3.1 Insights from the ChemicalUniverseMachine on the FMR evolution 85
6.3.2 Evolution of Metal Distribution and the Physical Drivers 89
6.4 The Metallicity Floor: A Legacy of Cosmic Chemical History . . 91
6.5 Further Considerations on Statistical Averages and Individual Dynamical States in Local EMPGs 93
6.6 Impact of Selection Biases 93
6.7 Future Prospects 94

Chapter 7 Summary 99

Appendix 101
A Supplement to Chapter 5 101
A.1 Corner Plot 101
A.2 Other Fitting Functions 101
A.3 Effects of Inflow Metallicity and Mergers 101
A.4 ISM distribution fractions for the well-mixed ISM model . 103
A.5 Effects of Stellar Yield 103

Reference 105

Chapter 1

Introduction

1.1 Background

A fundamental goal of modern astronomy is to elucidate the physical laws that govern the formation and evolution of galaxies. Central to contemporary theories is the paradigm of the baryon cycle: a continuous interplay between the inflow of gas from the cosmic web, the conversion of this gas into stars within the interstellar medium (ISM), and the subsequent outflow of gas and heavy elements (metals) driven by stellar activity. The efficiency and modality of this cycle are thought to be the primary drivers of the observed diversity in galaxy properties (see, e.g., the review by Somerville & Davé 2015). Therefore, developing a comprehensive understanding of the physics of the baryon cycle is paramount to constructing a complete picture of how galaxies evolve over cosmic time (Tumlinson et al. 2017).

However, directly observing the diffuse and often faint gas flows that constitute the baryon cycle is an exceptionally challenging task. This thesis, therefore, focuses on one of the most powerful diagnostic tools available for probing these processes: using the abundance of metals within the ISM, known as the gas-phase metallicity, as a tracer (Maiolino & Mannucci 2019). The metallicity serves as a crucial "fossil record", encoding essential information about a galaxy's past star formation history and its interaction with its environment. Specifically, we focus on the abundance of oxygen, which is a particularly effective tracer for several key reasons. First, oxygen is synthesized predominantly in massive stars and is rapidly returned to the ISM via core-collapse supernovae, making its abundance a direct and immediate measure of a galaxy's recent star formation activity. Second, as one of the most abundant heavy elements produced by stars, oxygen produces strong, prominent emission lines in the rest-frame optical spectrum, making its abundance relatively straightforward to measure observationally. The ISM metallicity is highly sensitive to the other components of the baryon cycle: it is diluted by the inflow of metal-poor gas and reduced by the expulsion of enriched gas in outflows. Consequently, the gas-phase metallicity serves as a crucial physical indicator that encodes the integrated history of star formation and gas flows a galaxy has experienced (Maiolino & Mannucci 2019).

Observational studies have established tight scaling relations between galaxy properties and metallicity. The most well-known of these is the mass-metallicity relation (MZR), a positive correlation between stellar mass and metallicity that is well-

established in the local universe over a wide range of stellar masses from approximately 10^7 to 10^{12} $M_\odot$ (e.g., Lequeux et al. 1979; Tremonti et al. 2004; Mannucci et al. 2010; Andrews & Martini 2013; Curti et al. 2020). This relation has been found to have a secondary dependence on the star formation rate (SFR). At a fixed stellar mass, galaxies with higher SFRs tend to exhibit lower metallicities. This three-parameter correlation is known as the fundamental metallicity relation (FMR; e.g., Ellison et al. 2008; Mannucci et al. 2010; Curti et al. 2020; Baker et al. 2023). The FMR provides a more complete description of the state of a galaxy's ISM and is a key focus of this thesis.

1.2 Fundamental Physics

To properly discuss the gas-phase metallicity of galaxies, it is essential to first understand the physical conditions of the interstellar gas from which metallicity measurements are derived. This section establishes the foundation for the analyses in this thesis by defining the essential physical quantities used to characterize the interstellar medium. We will first explain the physical basis and measurement of electron temperature, ionization parameter, and gas-phase metallicity, all of which are critical for interpreting nebular emission-line spectra. Following these definitions, we will introduce the equilibrium model, a key theoretical framework that provides a physical context for these observational measurements and connects them to the broader processes of galaxy evolution.

1.2.1 Electron Temperature

The electron temperature, denoted as T_e, is a fundamental physical quantity that describes the kinetic energy of free electrons within an H II region, thereby representing the thermal state of the ionized gas. This parameter is of paramount importance as it governs the frequency and energy of collisions between electrons and ions. Consequently, it directly influences the efficiency of collisional excitation, the process by which ions are raised to higher energy levels through collisions. The rate of collisional excitation is a primary determinant of the observed emission-line strengths, making electron temperature one of the most basic physical quantities required for interpreting nebular spectra.

The most reliable method for determining the electron temperature is the direct method, which utilizes the flux ratio of specific forbidden lines that are highly sensitive to temperature. A classic and widely used example, which is also employed in this thesis, involves the O^{2+} ion. The energy levels of this ion are illustrated in Figure 1.1. The method compares the fluxes of emission lines originating from different energy levels. Specifically, the [O III]λ4363 line is produced by the transition from the $^1S_0$ level to the $^1D_2$ level, while the much stronger [O III]λ5007 line is produced by the transition from the $^1D_2$ level to the $^3P_2$ level. In a gas with a higher electron temperature, electrons possess greater kinetic energy, making it more probable that their collisions will excite ions to the higher $^1S_0$ energy state. As a result, the flux ratio of the [O III]λ4363 line to the [O III]$\lambda\lambda$4959,5007 lines serves as an excellent diagnostic

for the electron temperature, as shown in Figure 1.2.

This temperature-sensitive ratio is advantageous because it is nearly independent of the gas density in typical H II regions. This is because the rates of collisional de-excitation are negligible compared to the rates of spontaneous radiative transitions when the electron density is well below the critical density. The critical density of level $^1S_0$ and $^1D_2$ are 6.4×10^5 cm^{-3} and 2.8×10^7 cm^{-3}, respectively. These values are significantly higher than the typical densities of H II regions ($\lesssim 10^3$ cm^{-3}). Therefore, the [O III]λ4363/[O III]λ5007 ratio is primarily a function of temperature, allowing for a robust estimation of the electron temperature of the O^{2+} ion, T_e([O III]).

1.2.2 Ionization Parameter

The ionization parameter, denoted as U, is a dimensionless quantity that describes the ratio of the number density of ionizing photons from stars to the number density of atoms, primarily hydrogen, in the surrounding gas. This parameter is a crucial physical quantity as it fundamentally determines the ionization state of the gas. A higher ionization parameter signifies a more intense radiation field relative to the gas density, which results in a greater fraction of atoms being in higher ionization states. For example, in a region with a high ionization parameter, doubly-ionized oxygen (O^{2+}) will be more prevalent than singly-ionized oxygen (O^+).

As the ionization parameter cannot be measured directly, it is typically inferred indirectly through the use of specific emission-line flux ratios. Ratios of emission lines from ions of different ionization stages are particularly sensitive to the ionization parameter. A widely used diagnostic, which is also employed in Chapter 3.1 of this thesis, is the ratio of the [O III]λ5007 and [O II]λ3727 fluxes, often referred to as the O32 index. The estimation is performed by comparing such observed line ratios to the predictions of photoionization models, of which `Cloudy` is a prominent example. These models compute the expected emission-line ratios over a grid of varying physical conditions, including different ionization parameters and metallicities. The ionization parameter for a given galaxy is then estimated by identifying the model that best reproduces the observed line ratios.

1.2.3 Gas-Phase Metallicity

The gas-phase metallicity is a fundamental physical property that refers to the total abundance of elements heavier than hydrogen and helium within the interstellar medium. It is one of the most essential quantities for discussing the chemical composition of galaxies. Observationally, the abundance of oxygen is typically used as a proxy for the total metallicity, due to its high cosmic abundance and the prominence of its bright emission lines. It is conventionally expressed on a logarithmic scale as the number ratio of oxygen to hydrogen atoms, denoted as $12 + \log(\mathrm{O/H})$. There are two principal methodologies employed for determining the gas-phase metallicity from emission-line spectra.

The first and most physically robust of these is the direct method. This technique relies on a direct measurement of the electron temperature (T_e), as described in Section 1.2.1. The procedure begins with the measurement of emission line fluxes for

ions such as O^+ and O^{2+} from lines like [O II]λ3727 and [O III]$\lambda\lambda$4959, 5007, relative to a hydrogen recombination line such as Hβ. With the electron temperature known, the ionic abundances of these species can be calculated from these flux ratios based on the principles of atomic physics. The total elemental abundance of oxygen is then derived by summing the abundances of the dominant observed ionization states. In practice, these complex calculations are performed using specialized software that incorporates the latest atomic data. A prominent example is PyNeb , a package for nebular physics calculations that is not only utilized in this thesis but is also widely used in previous studies (e.g., Pérez-Montero 2014; Esteban et al. 2014; Onodera et al. 2016; Annibali et al. 2019; Kojima et al. 2020; Nakajima et al. 2022; Curti et al. 2023). The primary limitation of the direct method, however, is its dependence on the detection of faint auroral lines, such as [O III]λ4363, which are necessary for the electron temperature measurement. These lines are often too faint to be detected in metal-rich or distant galaxies, restricting the applicability of this method to a limited subset of galaxies.

To overcome this limitation, a second approach known as strong-line calibrations is widely used. This method enables the estimation of metallicity using only the flux ratios of bright, easily detected emission lines, such as [O II], [O III], and Hβ. The methodology involves establishing an empirical correlation, or calibration, between a specific strong-line ratio, for instance the R23 index which is defined as ([O II]λ3727 + [O III]$\lambda\lambda$4959,5007) / Hβ, and the metallicity derived via the direct method for a well-characterized sample of local galaxies. Once this calibration is established, it can be applied to estimate metallicities for vast samples of galaxies where the faint auroral lines are not detected. While this technique offers the significant advantage of being applicable to large numbers of galaxies across cosmic time, it is known to be subject to considerable systematic uncertainties, with different calibrations yielding metallicity estimates that can differ by up to 0.5 dex (e.g., Kewley & Ellison 2008). Careful consideration of the chosen calibration is therefore essential. This thesis mainly refer to calibrations developed by Nakajima et al. (2022).

1.2.4 Equilibrium Model

Gas-phase metallicity is driven by a complex interplay of physical processes: the accretion of metal-poor gas from the IGM, gas consumption during star formation, ISM enrichment via supernovae and stellar winds, and the recycling of ejected material. By examining the dependence of metallicity on stellar mass (M_*) and star formation rate (SFR), we can place constraints on how gas accretion and outflow rates scale with these properties, thereby gaining vital insights into galaxy growth and evolution.

The "equilibrium" class of chemical evolution models serves as a robust and conceptually accessible analytical framework for physically interpreting these observed relations (e.g., Finlator & Davé 2008; Peeples & Shankar 2011; Davé et al. 2012; Lilly et al. 2013, Fig 1.3). Grounded in the conservation of baryonic mass, these models establish a balance between gas inflows and outflows, the SFR, the recycling of stellar material, and the time evolution of the gas reservoir. This analytical method complements detailed hydrodynamical simulations (e.g., Vogelsberger et al. 2014; Schaye et al. 2015) by providing a transparent lens to decipher the complex physics driving galaxy evolution.

The model is formulated by constructing a differential equation that describes the time evolution of the total gas mass and the total metal mass within a galaxy's ISM. This equation incorporates several key physical processes: the production of new metals from star formation, the removal of metals via galactic outflows, the supply of metal-poor gas through inflows, and the consumption of gas by ongoing star formation. Each of these processes is parameterized as a function of fundamental galaxy properties such as stellar mass and SFR, using terms like the mass-loading factor (λ) and the star formation efficiency (ϵ).

A central assumption of the model is that many galaxies exist in a state of dynamic equilibrium, where these competing processes are balanced such that the net change in the total gas mass over time is relatively small. By solving the chemical evolution equation under this equilibrium assumption, one can derive a theoretical prediction for the gas-phase metallicity that a galaxy should possess. These models have been shown to successfully reproduce the observed MZR and FMR. In this theoretical framework, the MZR is interpreted as a consequence of higher gas fractions and/or more efficient metal removal by outflows at lower stellar masses (Tremonti et al. 2004; Davé et al. 2012; Lilly et al. 2013; Sanders et al. 2023). The equilibrium model thus provides a valuable way to utilize MZR and FMR observations to constrain the interplay of these ongoing processes. This framework provides a valuable tool for interpreting observational data, and a version of this model, as described by Lilly et al. (2013), is utilized in Chapter 2 of this thesis.

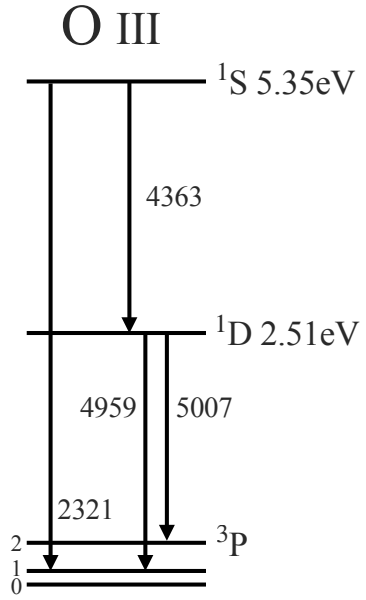

Fig. 1.1 Energy levels of O^{2+} ion and radiative transitions between energy levels (Aller 1984).

1.3 Key Scientific Questions

While the equilibrium state of the baryon cycle is relatively well-characterized in the local universe, a complete picture of how this state was established over cosmic history remains elusive. The overarching scientific goal of this thesis is to illuminate the evolutionary history of the baryon cycle, from the era of the first stars (a.k.a. Population III; Nakajima & Maiolino 2022) to the present-day equilibrium observed in galaxies. To address this ambitious goal, this thesis is structured around three central research questions that bridge observational exploration with theoretical modeling.

1.3.1 Origin of Local EMPGs and the Metallicity Floor Problem

Even in the well-studied local universe, fundamental questions remain, centered on the most primitive galaxies known: extremely metal-poor galaxies (EMPGs Izotov et al. 2019; Kojima et al. 2020). These systems, defined as having gas-phase metallicities less than 10% of the solar value ($12 + \log(\mathrm{O/H}) < 7.69$), are crucial analogs for understanding the early phase of galaxy formation. With properties such as low stellar masses, low metallicities, and high specific star formation rates, they are often regarded as local counterparts to the primordial galaxies predicted to exist at high redshift (e.g., Isobe et al. 2022b). Particularly before the launch of the James Webb Space Telescope (JWST) in 2022 (when the research for this thesis began) observational constraints made it extremely difficult to directly investigate the physical properties of galaxies in the early universe. Specifically, the rest-frame optical lines required for metallicity diagnostics are redshifted into the infrared regime at $z > 4$, often falling outside atmospheric windows and becoming inaccessible to ground-based instruments. Consequently, these local EMPGs served as the most accessible laboratories for studying the initial stages of galaxy formation. However, two major puzzles persist, motivating the work in this thesis.

The first is the apparent absence of Hyper Metal-Poor Galaxies (HMPGs), defined as galaxies with metallicities below 1% of the solar value ($Z < 0.01 Z_\odot$). Despite decades of intensive searches by numerous groups, including systematic studies like the Extremely Metal-Poor Representatives Explored by the Subaru Survey (EMPRESS; e.g., Kojima et al. 2020), no galaxies have been identified with metallicities below 1% $Z_\odot$ in the local universe. This observational fact, which is also seen in the stellar mass-metallicity relation of Local Group dwarf galaxies (e.g., Kirby et al. 2013), stands in stark contrast to predictions from some cosmological hydrodynamic simulations that find galaxies with significantly lower metallicities should exist (e.g., Wheeler et al. 2019; Applebaum et al. 2021, see also Hayashi et al. 2022). This discrepancy raises a critical question: do HMPGs not exist in the local universe, or are they simply too faint or rare to have been detected by current surveys? This issue is often discussed in the context of the "metallicity floor" (Wise et al. 2012; Kirby et al. 2013).

The second puzzle is the physical origin of EMPGs. Two primary scenarios are debated. The first is the "episodic star-formation scenario," which posits that metal-poor gas accretes onto a pre-existing, more metal-rich galaxy, triggering a new star-

burst that temporarily dominates its appearance (e.g., Sánchez Almeida et al. 2014, 2015, 2016; McQuinn et al. 2020). This model is supported by morphological evidence, as many EMPGs are composed of a star-forming metal-poor clump and an extended, diffuse "tail" structure which is often found to be more metal-rich by up to 1 dex. The second is the "first star-formation scenario," where the galaxy is a genuinely young system forming its first generation of stars from primordial gas (e.g., Recchi et al. 2001; Isobe et al. 2022a). While the episodic model is compelling, the chemical abundance picture remains controversial. Certain EMPGs exhibit unique abundance patterns, such as high Fe/O and low N/O ratios, that are difficult to explain via standard chemical enrichment with its associated time delays, but are consistent with enrichment by primordial massive stars like pair-instability supernovae (Isobe et al. 2022a). Understanding the true nature of these local systems is a critical first step toward understanding their primordial counterparts.

1.3.2 Cosmic Evolution of the Fundamental Metallicity Relation

A central question in galaxy evolution is whether the FMR, interpreted as an "equilibrium state" of the baryon cycle, is a universal law that holds across cosmic time. Answering this is critical for understanding how star formation and feedback operated in the early universe. However, until recently, probing the FMR at high redshifts ($z \gtrsim 3$) had been severely limited by observational constraints, particularly the difficulty of accessing rest-frame optical emission lines with ground-based telescopes. The James Webb Space Telescope (JWST), with its unprecedented near-infrared sensitivity and spectroscopic capabilities, has revolutionized this field, enabling detailed studies of galaxies out to $z \sim 10$.

Initial studies with JWST have already revealed tantalizing signs that the FMR may begin to break down at the earliest epochs. For instance, Nakajima et al. (2023), using JWST/NIRSpec data from the ERO and CEERS surveys, and Curti et al. (2024), analyzing JADES data, both found evidence for deviations from the local FMR emerging at redshifts of $z \gtrsim 6-8$. These pioneering works suggest that galaxies in the early universe are systematically more metal-poor than predicted by local relations based on their stellar mass and SFR.

However, these high-redshift studies face significant challenges. They often rely on strong-line metallicity calibrations, which are established using local galaxy samples. These calibrations may introduce systematic biases, as they may not accurately reflect the ISM conditions in early galaxies, which are expected to differ substantially in terms of ionization state, gas density, and elemental abundance ratios (e.g., Cameron et al. 2023; Topping et al. 2024). The more physically robust direct T_e method is observationally expensive, as it requires the detection of faint auroral lines like [O III]λ4363, which are often too faint to detect in individual high-redshift galaxies. Overcoming these limitations to robustly measure the evolution of the FMR is a key frontier in astrophysics. This thesis aims to address these challenges directly by employing a stacking technique on a large sample of JWST spectra, which allows for the detection of the crucial auroral lines and a more reliable measurement of metallicity evolution.

1.3.3 Empirical Model for the Cosmic Evolution of Chemical Enrichment

As observational data on metallicity across cosmic time and in multiple gas phases, including star-forming H II regions (e.g., Tremonti et al. 2004; Mannucci et al. 2010; Andrews & Martini 2013; Erb et al. 2006; Nakajima et al. 2013; Marshall et al. 2008; Sanders et al. 2021; Nakajima et al. 2023; Curti et al. 2024), neutral atomic hydrogen (H I) in Damped Lyman-Alpha (DLA) systems (e.g., Wolfe et al. 2005; Ledoux et al. 2006; Bañados et al. 2019; Berg et al. 2021; Ledoux et al. 2006; Møller et al. 2013), and the diffuse CGM (e.g., Tumlinson et al. 2017; Prochaska et al. 2017), become increasingly abundant, a corresponding need arises for theoretical models capable of interpreting this rich dataset. While various approaches such as analytical models (e.g., Davé et al. 2012; Lilly et al. 2013), semi-analytical models (e.g., Somerville et al. 2008), and hydrodynamical simulations (e.g., Finlator & Davé 2008; Ma et al. 2016; Torrey et al. 2019) have provided invaluable insights, they often rely on strong physical priors regarding complex sub-grid processes like star formation and feedback, which can complicate the interpretation of comparisons with observations.

This thesis, therefore, adopts the powerful and flexible framework of empirical modeling. This approach is designed to infer the underlying relationships between the well-understood formation of dark matter structures and the observed properties of galaxies. Unlike models with strong physical priors, empirical models search flexible parameter spaces to deduce the relationships present in the real Universe. Our work builds upon the successes of the UniverseMachine (Behroozi et al. 2019) and NeutralUniverseMachine (Guo et al. 2023). The UniverseMachine connects galaxy growth to halo growth by parameterizing star formation rates as a function of halo properties and calibrating them against a wide array of observational data, including stellar mass functions and galaxy clustering. The NeutralUniverseMachine extends this framework to model the evolution of H I and H_2 gas masses, constrained by observations of gas mass functions and cosmic gas densities.

This thesis introduces a new model built upon this foundation: the ChemicalUniverseMachine. We parameterize the fractions of newly synthesized metals that are distributed into the multiphase ISM ($f_{\rm H2}$ for the star-forming molecular phase and f_{HI} for the neutral atomic phase) and the CGM. Metal production, supply, and dilution are caused by star formation, galaxy mergers, and gas inflow from the inter-galactic medium, respectively, with rates determined by the (Neutral)UniverseMachine models. Since the amount of gas at each redshift is already constrained empirically by the NeutralUniverseMachine model, net gas inflow and outflow rates are already constrained by the evolution of structure in ΛCDM, so we do not need to assume or parameterize gas regulation physics (including mass loading factors). By constraining these distribution fractions empirically against the full suite of available observational data across all gas phases, we aim to self-consistently infer the physical processes of metal distribution throughout cosmic time.

1.4 Structure of This Thesis

This thesis addresses the fundamental question: "When and how is the equilibrium state of galaxies established?" To answer this, we combine observational studies of the early stages of galaxy formation (Chapters 2–4) with the construction of a physical model to interpret these observations (Chapter 5).

Chapter 2 details the systematic survey of HMPGs/EMPGs in the local universe. We develop and apply a machine learning technique to select photometric candidates from the Sloan Digital Sky Survey (SDSS) imaging data. These candidates are then followed up with ground-based spectroscopy from the Seimei, Nayuta, and Magellan telescopes. Using a large sample of EMPGs, we investigate the physical origins of EMPGs through their chemical properties.

Chapter 3 utilizes the observational window opened by JWST. We extend the successful search methodology developed in Chapter 2 to the early universe, applying it to deep near-infrared imaging from JWST's NIRCam instrument. We identify a sample of compelling EMPG candidates with the primary aim of discovering HMPGs, which are expected to exist in the early universe despite their apparent absence in the local universe. We further conducted spectroscopic follow-up observations using FOCAS on the Subaru Telescope to spectroscopically confirm their redshifts.

Chapter 4 employs large-scale spectroscopic survey data from JWST to investigate the chemical evolution of galaxies from redshift $z = 2$–10. By combining data from multiple public surveys including DREAMS, JADES, and CEERS, we apply a spectral stacking technique to detect faint auroral lines, enabling robust metallicity measurements via the direct T_e method. This analysis allows us to probe the cosmic evolution of key scaling relations, specifically the MZR and the FMR.

In Chapter 5, we present the ChemicalUniverseMachine, a new empirical model that connects the baryon cycle to the distribution of metals in the multiphase interstellar and circumgalactic medium. This model is designed to provide a unified, self-consistent interpretation of our diverse observational results, from the local universe to high redshift, and to explore the physical mechanisms driving the evolution of the baryon cycle.

Finally, Chapter 6 provides a comprehensive discussion of the conclusions drawn from this thesis, and Chapter 7 summarizes the conclusions of this thesis.

Throughout the paper, we assume a Chabrier (2003) initial mass function (IMF). We use a solar metallicity $Z_\odot$ as $12 + \log(\mathrm{O/H}) = 8.69$ (Asplund et al. 2021), and adopt a standard ΛCDM cosmology with $\Omega_\Lambda = 0.7, \Omega_M = 0.3$ and $H_0 = 70$ km s^{-1} Mpc^{-1}. All magnitudes are given in the AB system (Oke & Gunn 1983). Halo masses are defined based on the Bryan & Norman (1998) spherical overdensity definition and denote peak historical halo masses extracted from the merger tree (Mpeak).

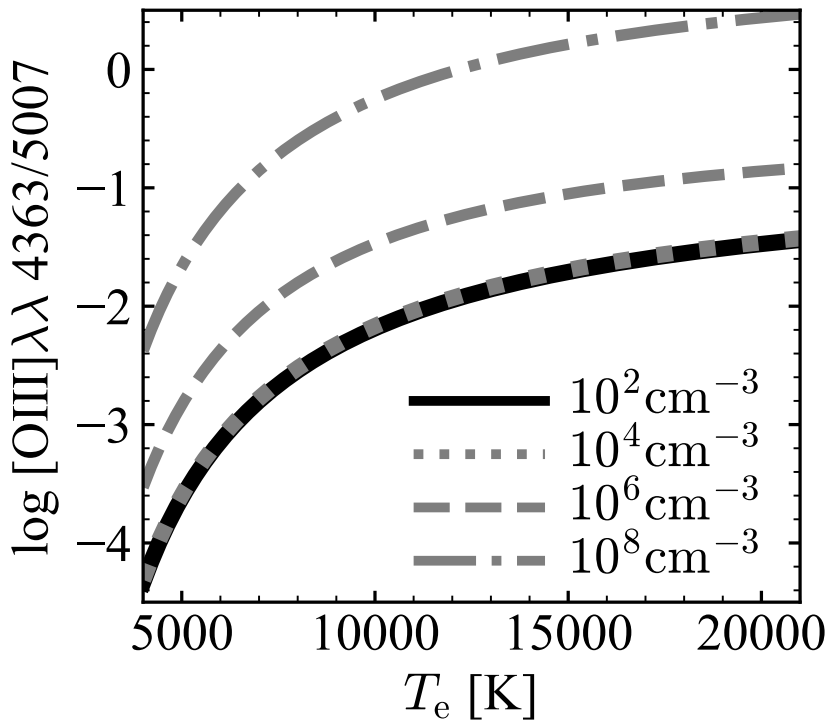

Fig. 1.2 Line ratio of [O III]$\lambda\lambda$4363/5007 as a function of electron temperature. These values are obtained with Pyneb (Luridiana et al. 2015). The black line indicates the relation for electron temperature of typical H II regions (i.e., $n_e = 10^2$ cm^{-3}). The gray dotted, dashed, and dashed-dotted lines represent the relation for electron temperature of $n_e = 10^4$, 10^6, and 10^8 cm^{-3}, respectively, which show the density dependence on the ratio is negligible if the density is significantly lower than critical density (i.e., $n_e \lesssim 10^4$ cm^{-3}).

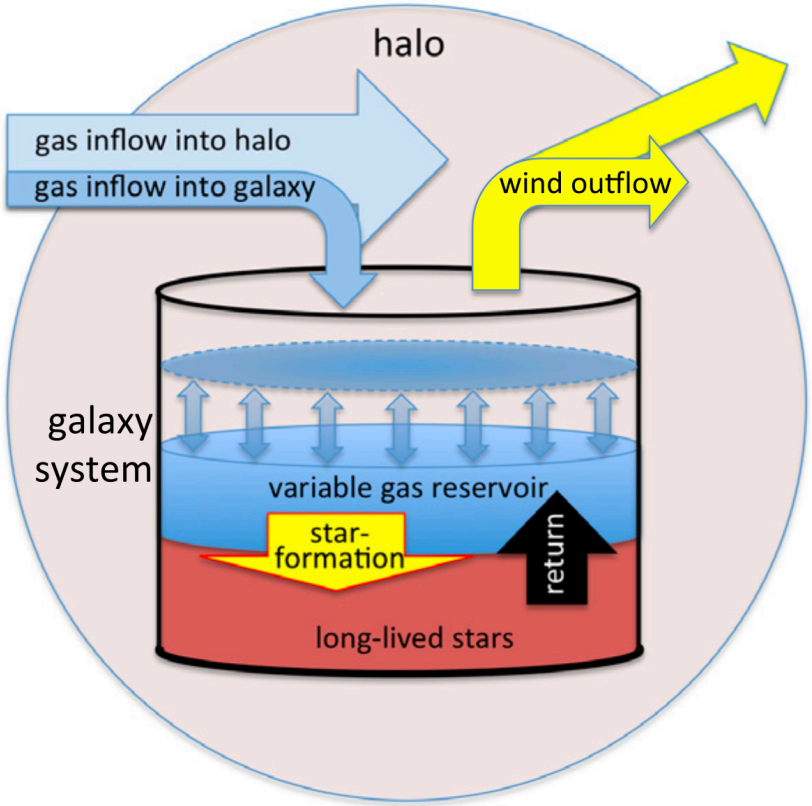

Fig. 1.3 Illustration of basic processes of the equilibrium model (Lilly et al. 2013). Gas flows into the halo, some fraction of which also flows into the galaxy and adds to the gas reservoir. Stars continuously form out of the reservoir and a fraction of which is immediately returned to the reservoir, along with newly produced metals. Some gas is expelled by a wind from the system, and possibly from the halo.

Chapter 2

Extremely Metal Poor Galaxies in the Local Universe

In this chapter, we search for metal-poor galaxy candidates from SDSS imaging data using a machine-learning technique and conducting follow-up spectroscopy. We describe the details of our selection method for identifying local EMPGs in Section 2.1 and present the results of spectroscopic follow-up observations used to measure the metallicities of EMPG candidates. In Section 2.5, we present our sample of spectroscopically-confirmed EMPGs that we use in the subsequent analysis. In Section 2.6, we examine their chemical properties of our EMPG sample.

2.1 Selection Method

This section presents local EMPGs identified by our selection method and spectroscopic observations. We develop our selection method using a machine learning technique, testing the selection method with the SDSS spectroscopic objects (Section 2.1.2). We then apply the selection method to the faint SDSS photometric objects with no spectroscopic identifications to make EMPG photometric candidates (Section 2.2). Because stars with no emission lines contaminate the EMPG photometric candidates, we conduct shallow spectroscopic screening observations for the EMPG photometric candidates via strong emission lines with 2–4m telescopes of Seimei and Nayuta (Sections 2.3.1 and 2.3.2). We perform deep spectroscopy with a large telescope of 6.5m Magellan for the screened candidates to confirm EMPGs via metallicity measurements with faint emission lines detected with Magellan (Section 2.4).

2.1.1 Developing Machine Learning Classifier

In this study, we focus on identifying EMPGs at redshifts $z < 0.03$ with a large rest-frame equivalent width of Hα, denoted as EW$_0$(Hα), of $\gtrsim 800$ Å. Our motivation for this selection criterion is to identify local counterparts of high-z low-mass galaxies with high sSFRs of $\gtrsim 10$ Gyr^{-1} (e.g., Ono et al. 2010; Harikane et al. 2018; Stark et al. 2017). Galaxies with such large EW$_0$(Hα) are thought to have undergone a very early phase of galaxy evolution (e.g., Inoue 2011).

To identify these EMPGs efficiently from objects in a photometric catalog, we

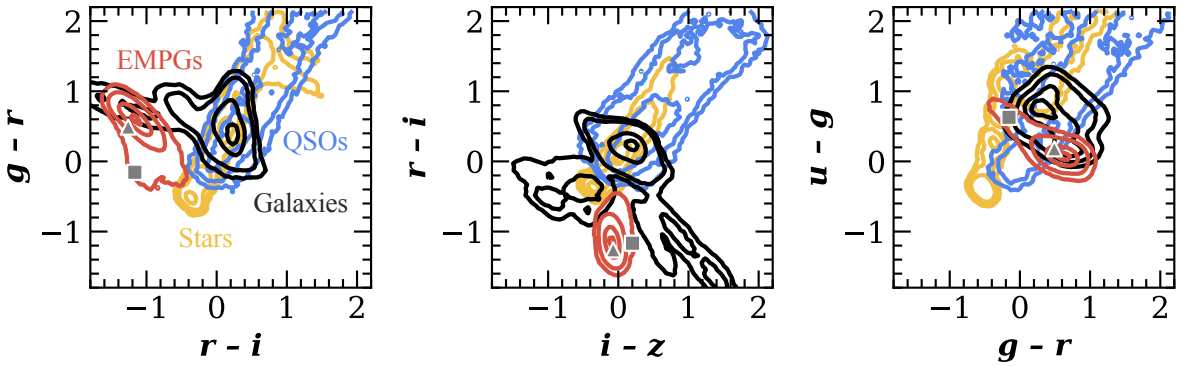

Fig. 2.1 Colors of the mock observational objects on the color-color diagrams of $g-r$ vs. $r-i$ (left), $r-i$ vs. $i-z$ (middle), and $u-g$ vs. $g-r$ (right). The contours show the 1, 10, 50, and 90% level for each object type of EMPGs (red), non-EMPG galaxies (black), stars (yellow), and QSOs (blue). The gray triangle and square indicate the representative points for the HMPG model ($Z = 0.01\ Z_\odot$) and the EMPG model ($Z = 0.1\ Z_\odot$), respectively.

construct a machine-learning classifier using a supervised machine learning algorithm. We utilize LightGBM (Ke et al. 2017), a gradient boosting framework that employs a decision-tree based learning algorithm. The inputs for the classifier are four colors of SDSS *ugriz* bands ($u-g, g-r, r-i$, and $i-z$) for each object, while the output is one of the four object types: EMPG, non-EMPG galaxy, star, or QSO. Our goal is to isolate EMPGs from other object types with minimal contamination. We chose to use colors as inputs to capture information about the spectral energy distribution of each object and avoid biases that may arise from other observable properties like size, morphology, or brightness.

To develop the machine-learning classifier, a training sample is prepared with spectral energy distribution (SED) models, following the approach of Kojima et al. (2020). The training sample comprises synthetic colors derived from SED models of EMPGs, non-EMPG galaxies, stars, and QSOs. These values are computed by convolving the respective SEDs with the broadband filter response functions of the SDSS (Fukugita et al. 1996). The EMPG SED models used in this study are a combination of the EMPG models of Kojima et al. (2020) and Nakajima & Maiolino (2022), whose metallicity ranges are $Z = 0.01$–$0.1\ Z_\odot$ and $Z = 0.001$–$0.01\ Z_\odot$, respectively. The EMPG SED models for metallicities $Z = 0.01$–$0.1\ Z_\odot$ are produced with the SED model code Beagle (Chevallard & Charlot 2016). The Beagle code calculates the stellar continuum and nebular emission using Bruzual & Charlot (2003)'s stellar population synthesis code and the nebular emission library of Gutkin et al. (2016), which is computed using the photoionization code Cloudy (Ferland et al. 2013). A Chabrier (2003) stellar initial mass function (IMF) is applied in the Beagle code. We use five parameters: stellar mass M_*, age $t_{\max}$, metallicity Z, ionization parameter U, and redshift z. These parameters are varied in the following ranges: $\log(M_*/M_\odot) = 4.0 - 9.0$, $\log(t_{\max}/\mathrm{yr}) = 6.00$–$7.75$, $Z = 0.01 - 0.1\ Z_\odot$, $\log U = -2.7 - -2.3$, and $z = 0.01$–0.02, respectively. These ranges of parameters are based on typical values observed in EMPGs and/or theoretical predictions. The EMPG SED models for metallicities $Z = 0.001$–0.01

$Z_\odot$ are based on the SED models presented by Nakajima & Maiolino (2022). These models have metallicities of $Z = [1/1400, 1/140]$ $Z_\odot$, cover stellar ages from 6.0 $<$ log(age/yr) $<$ 7.8 with a step size of 0.2, span ionization parameters $\log U$ from -3.0 to -1.0 with a step size of 0.5, and include dust attenuations of $E(B-V) = [0, 0.01, 0.02, 0.05]$, assuming a Kroupa (2001) IMF with upper mass limits of 100 $M_\odot$ and 300 $M_\odot$. The SED models of non-EMPG galaxies are generated similarly to those of EMPGs with $Z = 0.01$–0.1 $Z_\odot$, but with different parameter ranges, which contains the SED models of high-z ($z = 0.08$–0.40) [O III] emitters. The stellar SED models are based on those of Castelli & Hubrig (2004). QSO SED models are generated from the observed composite spectrum of quasars at $1 < z < 2.1$ (Selsing et al. 2016) by parameterizing three key parameters: the power-law index (α) of the intrinsic near-ultraviolet slope ($f_\lambda \propto \lambda^\alpha$), the V-band dust attenuation optical depth, and redshift. Full details of the SED models are described in Kojima et al. (2020). The numbers of the model SEDs are 4,752, 6,728, 4,240, and 7,250 for the EMPGs, non-EMPG galaxies, stars, and QSOs, respectively. We add random noise to the colors to simulate observational errors, and mock observational objects are produced. The mock EMPGs, non-EMPG galaxies, stars, and QSOs consist of 475,200, 672,800, 42,400, and 72,500 objects, respectively. From each of the four object types, 30,000 mock observational objects are randomly selected to create a training sample of 120,000 objects in total ($= 30,000 \times 4$).

To illustrate the typical color properties of our models, we select two representative SEDs from the EMPG and HMPG populations. The HMPG model is characterized by $Z = 0.01$ $Z_\odot$, $\log(t_{\rm max}/{\rm yr}) = 6.0$, $\log U = -1.5$, $E(B-V) = 0.01$, and $z = 0.01$, while the EMPG model is characterized by $Z = 0.1$ $Z_\odot$, $\log(t_{\rm max}/{\rm yr}) = 6.0$, $\log U = -2.0$, $E(B-V) = 0.01$, and $z = 0.01$. Figure 2.1 shows the colors of the mock observational objects on the color-color diagrams of $g-r$ versus $r-i$, $i-z$ versus $r-i$, and $g-r$ versus $u-g$, with these representative HMPG and EMPG models highlighted as a gray triangle and a gray square, respectively.

We divide the training sample into two sets: 80% for training and the remaining 20% for validation. We train the machine-learning classifier to minimize classification errors on the training data set. LightGBM has several hyperparameters that control the learning process. We focus on five hyperparameters: `num_leaves`, `min_data_in_leaf`, `learning_rate`, `bagging_fraction`, `bagging_freq`, which are important in improving classification performance while avoiding over-fitting to the training data set using an automatic hyperparameter optimization software framework, optuna[*1]. The meanings of these hyperparameters can be found in the LightGBM documentation[*2]. Default values are used for the remaining hyperparameters. To avoid bias towards a particular model, we select 20 hyperparameter sets that show comparable performance. We then design our classifier to select EMPG candidates predicted by at least a certain number of models. We set this threshold at 3 because it yields the highest number of true EMPGs selected in the validation data set. To evaluate the performance of the classifier, we calculate both accuracy and F1 score on the validation data set, which both metrics yielding results over 0.99. Here, the accuracy is deter-

[*1] `https://optuna.org/`

[*2] `https://lightgbm.readthedocs.io/en/v3.3.4/pythonapi/lightgbm.LGBMClassifier.html`

mined by the total number of correctly identified EMPGs (true positives) as well as the correctly identified non-EMPG galaxies, stars, or QSOs (true negatives), which are then divided by the total number of classifications. This can be expressed as $(N_{TP}+N_{TN})/N_{tot}$, where N_{TP} (N_{TN}) represents the number of actual EMPGs (non-EMPG galaxies, stars, or QSOs) that are correctly classified as such. The F1 score is also a widely used metric for evaluating binary classification tasks, and it is calculated as the harmonic mean of precision and recall. Precision measures the proportion of true positives out of the total number of samples that are classified as EMPGs, while recall measures the proportion of true positives out of the total number of actual EMPGs.

2.1.2 Testing the Classifier with the SDSS Spectroscopic Data

Before we apply the machine learning classifier to the SDSS photometric data, we test whether the classifier correctly isolates EMPGs from other-type objects (non-EMPG galaxies, stars, or QSOs) with the SDSS objects having spectroscopic classifications (i.e., galaxy, star, or QSO) and emission-line measurements, as well as optical photometry. Using the SDSS DR16 (Ahumada et al. 2020) data, we select objects whose photometric measurements are brighter than the SDSS limiting magnitudes (95% completeness for point sources), $u < 22.0$ mag, $g < 22.2$ mag, $r < 22.2$ mag, $i < 21.3$ mag, and $z < 21.3$ mag. We remove objects with errors larger than 0.2 mag in u, g, r, i, and z bands. Note that we use `Modelmag` in the SDSS DR16 data. We remove an object with "0" (i.e., unclean) for the clean flag in `PhotoObjALL` of the SDSS DR16 data that has a photometric measurement problem, such as a duplication, deblending/interpolation, suspicious detection, and detection at the edge of an image. We also remove an object with $r - i > 0$ whose color is apparently different from EMPGs (left and middle panels of Figure 2.1) to save computational power. By these selections, we have constructed a catalog with the spectroscopy and photometry composed of 24,419 sources in total (998 galaxies, 22,583 stars, and 838 QSOs) that is referred to as the SDSS test catalog.

We apply our machine-learning classifier to the SDSS test catalog, and obtain 40 EMPG candidates. We check the spectroscopic classifications (i.e., galaxy, star, and QSO) of these EMPG candidates, and identify 38 candidates as galaxies and 2 candidates as stars. Of the 38 galaxies, we find detailed spectroscopic studies for 18 galaxies in the literature (Kniazev et al. 2003; Izotov et al. 2006, 2012; Thuan & Izotov 2005; Guseva et al. 2007; Nakajima et al. 2022). We obtain direct-method metallicity measurements based on [O III]λ4363 of these 18 galaxies. We find that 6 out of 18 galaxies are EMPGs ($12+\log(O/H) = 7.22$–7.65). The other 12 galaxies are also actively star-forming galaxies with intense emission lines but have slightly higher metallicities ($12 + \log(O/H) = 7.72$–8.37). Hereafter, we refer to such a moderately metal-poor galaxy as a "MPG". For the remaining 20 galaxies, we estimate metallicities with strong line ratios as described below. We obtain N2 ([N II]λ6584/Hα), R3 ([O III]λ5007/Hβ), and R23 (([O III]$\lambda\lambda$5007, 4959 + [O II]λ3727)/Hβ) indices using the flux measurements in the SDSS `galSpecLine` catalog, only for sources with S/N > 3 for all emission lines used. Using the metallicity calibrations of Nakajima et al. (2022), we determine the high- or low- metallicity branch (i.e., whether $12+\log(O/H)$

$\gtrsim 8$ or not) from N2 index, and then we estimate metallicities from R23 index if available, otherwise R3 index. We find that 5 are EMPGs ($12 + \log(\mathrm{O/H}) \simeq 7.54$–$7.60$), 13 are MPGs ($12 + \log(\mathrm{O/H}) \simeq 7.69$ – 8.61), and the remaining one is metal-rich galaxies ($12 + \log(\mathrm{O/H}) \simeq 8.80$). Taken together, the purity is 28% (11/40), which is lower than the accuracy in the training (Section 2.1.1) but comparable to the results of spectroscopic observations in Kojima et al. (2020), 30% (3/10).

2.1.3 Comparison between Classifiers

In this section, we present a comparison of the selection performance of our machine-learning classifier with those of Kojima et al. (2020) and Nakajima et al. (2022). These three methods share the same idea; they isolate EMPGs based on photometric data using color excess expected for EMPGs at $z < 0.03$ with $EW_0(H\alpha) > 800$ Å. One difference is the choice of a machine learning algorithm. While Kojima et al. (2020) and Nakajima et al. (2022) use CNN, we use lightGBM, which is widely used for classification tasks on tabular data. We can see how much the performance would change with a classifier using a different algorithm other than CNN. Another difference is the training data. Both Nakajima et al. (2022) and this work use the SED models of galaxies with $Z < 0.01\ Z_\odot$ (i.e., HMPGs) in the training data, while Kojima et al. (2020) does not, which allow us to search for HMPGs.

We investigate whether these three classifiers produce differences in a selection performance. We select EMPG candidates using the classifier of Kojima et al. (2020) (Nakajima et al. (2022)) from the SDSS test catalog described in Section 2.1.2. A total of 47 (40) objects are selected as EMPG candidates, of which 42 (38) are galaxies and 5 (2) are stars based on the SDSS spectroscopic classifications. We obtain direct-method metallicities for 20 (18) out of the 42 (38) galaxies from the literature (Kniazev et al. 2003; Izotov et al. 2006, 2012; Thuan & Izotov 2005; Guseva et al. 2007; Nakajima et al. 2022), and find that 6 (7) are EMPGs and 14 (11) are MPGs. For the remaining 22 (20) objects, we estimate metallicities using the strong line method as described in Section 2.1.2, resulting in 4 (4) being EMPGs, 15 (13) being MPGs, and 3 (3) being metal-rich galaxies. The overall purity of the classifiers of Kojima et al. (2020) and Nakajima et al. (2022) are 21% (10/47) and 28% (11/40), respectively. The results are summarized in Table 2.1.

Compared to Kojima et al. (2020), our and Nakajima et al. (2022)'s classifiers reduce the rate of stellar contamination from 10% to 5%, while maintaining a similar number of correctly selected EMPGs. Although there is some overlap in the EMPGs selected by each classifier, we select 3 (2) EMPGs that are not selected by the classifier of Kojima et al. (2020) (Nakajima et al. (2022)), while the classifier of Kojima et al. (2020) (Nakajima et al. (2022)) select 2 (2) EMPGs that are not selected by ours. Combining multiple models may offer the advantage of increasing the total number of correctly selected EMPGs.

Table 2.1 The number of EMPGs, other galaxies, and stars that are selected by the each mahine-learning classifier from the SDSS test catalog.

	This study	Kojima et al. 2020	Nakajima et al. 2022
EMPGs	11 (6)	10 (6)	11 (7)
Other galaxies	27 (12)	32 (14)	27 (11)
Stars	2	5	2
Total	40	47	40

Note—The number of EMPGs and MPGs with the direct-method metallicity measurements are noted in brackets.

2.2 Photometric Selection

In this section, we select EMPG candidates from an SDSS photometric catalog with our machine-learning classifier, which works well with the SDSS test catalog (Section 2.1.2). Based on the SDSS DR16 (Ahumada et al. 2020) photometric data with no SDSS spectroscopic measurements, we select objects that satisfy the criteria of the SDSS limiting magnitudes, magnitude errors, flags, and colors as described in Section 2.1.2. We obtain 569,825 sources in total from the SDSS DR16 photometric data that is referred to as the SDSS photometric catalog. We apply our machine-learning classifier to the SDSS photometric catalog and 734 objects are classified as EMPGs. We check the images of the 734 candidates and eliminate sources that obviously have an unclean photometric data but not fully removed with the flags. By this visual inspection, we obtain 134 EMPG candidates, whose i-band magnitudes are in the range of $i \simeq 14 - 20$ mag.

2.3 Spectroscopic Screening

We carried out spectroscopic screening observations with 3.8m Seimei and 2m Nayuta telescopes for the 134 candidates to remove contamination of stars via strong emission lines and select candidates for deep spectroscopy.

2.3.1 Screening with 3.8m Seimei

We carried out spectroscopic observations for 130 EMPG candidates with KOOLS-IFU (Matsubayashi et al. 2019) on the Seimei telescope (Kurita et al. 2020) during 2020 December and 2021 March (PI: Y. Isobe). We used the VPH-blue grism with a wavelength range of 4100–8900 Å and a spectral resolution power of $R \sim 500$. The exposure times were 600 s. The sky was clear during the observations with seeing sizes of around $2''$.

Data reduction is conducted using the software provided on the website of KOOLS-

IFU[*3]. The reduction and calibration processes include the bias subtraction, flat-fielding, sky subtraction, wavelength calibration, and flux calibration. A 1D spectrum is made by combining spectra of the region where a target flux is detected. A 1D error spectrum contains read-out and photon noise of sky and object emission. One of the reduced spectra obtained by the Seimei observations is shown in Figure 2.2. We note that the flux of the [O III]λ5007 line in Figure 2.2 may be saturated. However, the flux value is not important since the purpose of the screening observations is to isolate galaxies with strong emission lines.

2.3.2 Screening with 2m Nayuta

We carried out spectroscopic observations for four EMPG candidates with MALLS on the Nayuta telescope on 2021 April (PI: Y. Isobe). We used the 150 mm^{-1} grating with a slit width of 1.2 arcsec. Our observations had a wavelength coverage of 3700 − 9500 Å and a spectral resolution of $R \sim 600$. The exposure time was 1200 s. The seeing size during the observation was around $4''$. Data reduction is conducted in a standard manner with IRAF. One of the reduced spectra obtained by the Nayuta observations is shown in Figure 2.3.

2.3.3 Results of the Spectroscopic Screening Observations

Of the 134 sources, 104 (77) objects have a strong emission line around wavelength of Hα (Hβ and [O III]λ5007). Some sources display strong Hβ, [O III]λ5007, and Hα emission lines but lack an apparent detection of [N II]λ6583, indicating a low [N II]λ6583/Hα ratio ($\log([\mathrm{N\,II}]\lambda6583/\mathrm{H}\alpha) \lesssim -1$) and corresponding to low metallicities ($Z \lesssim 0.5\ Z_\odot$; Nakajima et al. 2022). We select 10 sources that are observable with the Magellan telescope during our scheduled observation dates and do not have significant [N II]λ6583 detections indicating low metallicities. Out of these 10 sources, eight were identified during the Seimei observations, and the remaining two were identified during the Nayuta observations.

2.4 Spectroscopic Confirmation of Metallicity

2.4.1 MagE Observation and Data Reduction

We carried out deep spectroscopy to confirm EMPGs via metallicity measurements with faint emission lines detected with the Magellan Echellette Spectrograph (MagE; Marshall et al. 2008) on the Magellan telescope. We observed the 10 EMPG candidates from the results of screening observations (Section 2.3.3). The observing nights were 2021 July 10, October 9, 2022 April 27, and 28 (PI: M. Rauch). We used the echellette grating with the $0\farcs70 \times 10''$ or $1\farcs0 \times 10''$ slits. The exposure times were 300 s or 600 s, depending on luminosities of targets. The MagE spectroscopy covered $\lambda \sim 3100 - 10000$ Å with a spectral resolution of $R \sim 4000$. We observed the standard

[*3] http://www.kusastro.kyoto-u.ac.jp/~kazuya/p-kools/reduction-201806/index.html

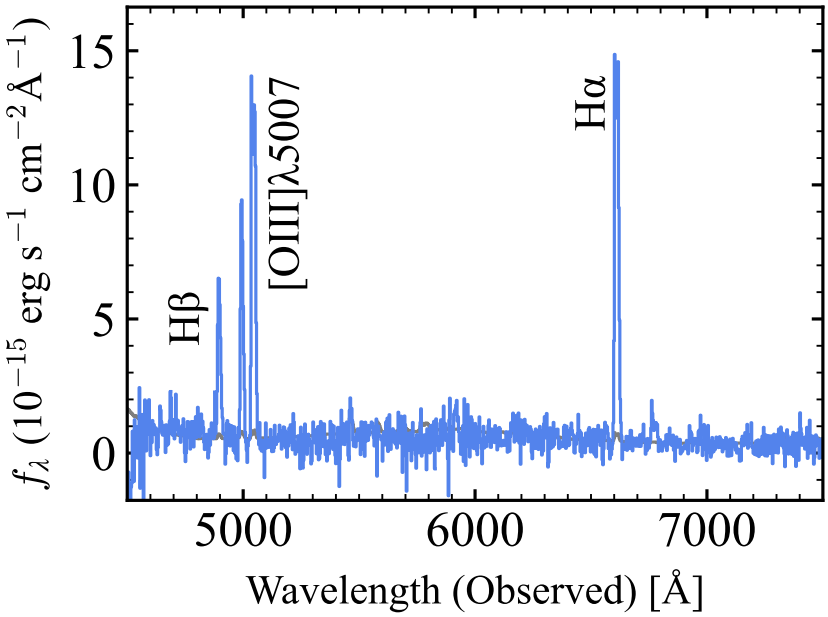

Fig. 2.2 Spectrum of J1034–0221 obtained by Seimei observations. The gray line indicates the error spectrum.

Table 2.2 Summary of MagE Observations.

ID	R.A. (hh:mm:ss)	Dec. (dd:mm:ss)	Slit (arcsec)	Exposure (s)
(1)	(2)	(3)	(4)	(5)
J1034–0221	10:34:01.0	–02:21:50.4	1.0	600
J1244+2828	12:44:41.3	+28:28:04.4	1.0	600
J1305+2852	13:05:46.9	+28:52:03.0	1.0	600
J1432+0611	14:32:07.3	+06:11:16.2	0.70	300
J1526+1610	15:26:11.6	+16:10:00.8	1.0	600
J1604+1459	16:04:18.0	+14:59:37.5	1.0	600
J1616+1453	16:16:29.8	+14:53:24.1	1.0	600
J1637+1729	16:37:42.0	+17:29:50.7	0.70	300
J1804+0008	18:04:19.6	+00:08:05.8	0.70	300
J2136+0414	21:36:58.0	+04:14:00.0	1.0	300

Note—(1) ID. (2) Right ascension in J2000. (3) Declination in J2000. (4) Slit width. (5) Exposure time.

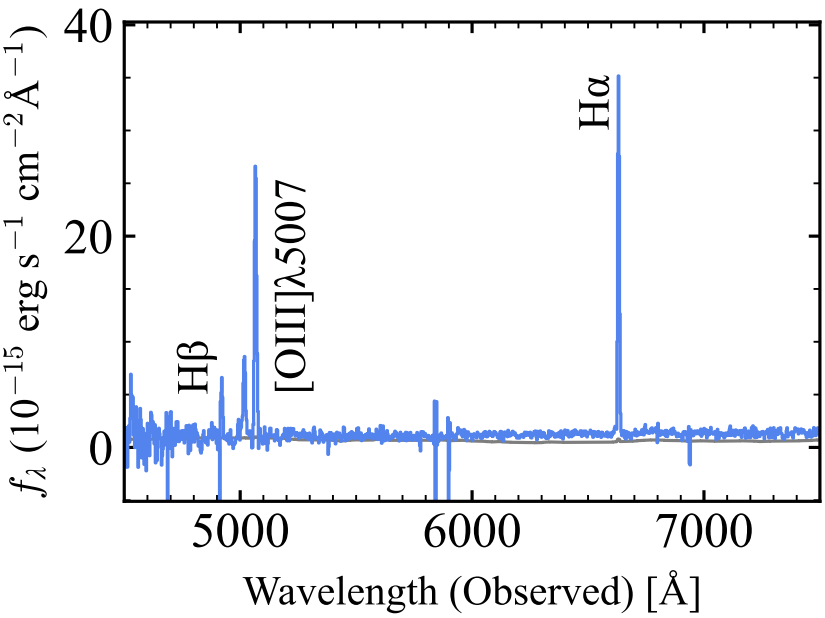

Fig. 2.3 Spectrum of J1804+0008 obtained by Nayuta observations.

stars CD329927 and Feige56. All the nights were clear with a typical seeing sizes of $\sim 0''.7$.

We perform the reduction of MagE data using PypeIt[*4] (Prochaska et al. 2020), an open-source Python-based spectroscopic data reduction pipeline. PypeIt executes standard processing steps, including bias subtraction, flat fielding, sky subtraction, wavelength calibration, cosmic-ray rejection, and 1D spectral extraction. Following the method outlined in Xu et al. (2022), we modify the sky model in regions with bright extended emission lines (e.g., Hα and [O III]λ5007), where the default fitting often fails. In these pixels, we enforce a flat sky background by averaging the values from nearby pixels. We note that the majority of the emission lines analyzed here are free from significant skyline contamination. We show MagE spectra as well as SDSS images of the 10 candidates in Figure 2.4.

Line fluxes are measured by fitting a Gaussian profile superimposed on a constant continuum. The measured values are corrected for Galactic extinction following the Schlafly & Finkbeiner (2011) map and the Cardelli et al. (1989) extinction curve. The dust attenuation is derived from the Balmer decrement (Hα, Hβ, Hγ, and Hδ), adopting the Calzetti et al. (2000) attenuation law, following the early EMPRESS work (Isobe et al. 2022a; Nakajima et al. 2022). The obtained $E(B-V)$ values are listed in Table 2.3. These $E(B-V)$ values are higher than those we assume in SED fitting in Section 2.1 but comparable to those of the EMPGs identified in Kojima et al. (2020). Flux errors are estimated from 1D error spectra extracted by PypeIt. Table 2.4 lists the key line fluxes normalized by Hβ and their 1σ errors for the 10 candidates.

We confirm that the 10 candidates do not show an obvious active galactic nucleus

[*4] `https://github.com/pypeit/PypeIt`

Table 2.3 Physical properties of the 10 EMPG candidates.

ID	Redshift	$E(B-V)$ (mag)	$T_e([O\,III])$ (10^4K)	$n_e([O\,II])$ (cm^{-3})	$12+\log(O/H)$
(1)	(2)	(3)	(4)	(5)	(6)
J1034–0221	0.00718	0.254 ± 0.005	1.67 ± 0.01	55 ± 22	7.668 ± 0.003
J1244+2828	0.00337	0.221 ± 0.007	1.13 ± 0.01	17 ± 6	8.201 ± 0.013
J1305+2852	0.02698	0.222 ± 0.007	0.94 ± 0.02	232 ± 6	8.376 ± 0.025
J1432+0611	0.00831	0.154 ± 0.008	1.40 ± 0.01	70 ± 6	7.945 ± 0.010
J1526+1610	0.01540	0.327 ± 0.012	1.21 ± 0.02	78 ± 2	8.163 ± 0.019
J1604+1459	0.01517	0.225 ± 0.010	1.33 ± 0.02	95 ± 5	7.962 ± 0.022
J1616+1453	0.01588	0.221 ± 0.013	0.98 ± 0.03	24 ± 2	8.311 ± 0.042
J1637+1729	0.01363	0.254 ± 0.014	1.58 ± 0.02	72 ± 6	7.765 ± 0.015
J1804+0008	0.00629	0.507 ± 0.007	1.90 ± 0.01	144 ± 74	7.504 ± 0.005
J2136+0414	0.01714	0.213 ± 0.014	1.96 ± 0.50	48 ± 63	7.388 ± 0.399

Note—(1) ID. (2) Redshift. (3) Dust attenuation of gas. (4) Electron temperature of O^{2+}. (5) Electron density of O^+. (6) Gas-phase metallicity.

(AGN) activity based on the BPT (Baldwin et al. 1981) diagram (Kauffmann et al. 2003), as presented in Figure 2.5. We note that we can not exclude the possibility of the existence of a metal-poor AGN with the BPT diagram (e.g., Kewley et al. 2013)

2.4.2 Metallicities

We identify the temperature-sensitive line of [O III]λ4363 in all of the 10 EMPG candidates, which allows us to derive gas-phase metallicities with the direct T_e method as described below. We use the tool PyNeb (Luridiana et al. 2015) to estimate electron temperatures, electron number densities, and ion abundances. We use [O III]λ4363/[O III]$\lambda\lambda$4959,5007 ratio and [O II]$\lambda\lambda$3726/3729 ratio to estimate electron temperatures and electron number densities of O^{2+} ($T_e([O\,III])$ and $n_e([O\,II])$) using the PyNeb package `getCrossTemDen`. The temperature of O^+, $T_e([O\,II])$, is estimated from $T_e([O\,III])$ using an empirical relation of Garnett (1992). We assume that the density of the gas is uniform throughout. We derive the abundances of O^+/H^+ with the [O II]$\lambda\lambda$3726,3729 / Hβ ratio and $T_e([O\,II])$, and O^{2+}/H^+ with the [O III]$\lambda\lambda$4959,5007 / Hβ ratio and $T_e([O\,III])$, using the PyNeb package `getIonAbundance`. We ignore O^{3+} and higher-order oxygen ions as in previous work (e.g., Isobe et al. 2022a). The measured oxygen abundances as well as redshift, $E(B-V)$, and $T_e([O\,III])$ are summarized in Table 2.3.

The derived metallicities are in the range of $Z = 0.05$–$0.5\ Z_\odot$ as similarly identified in earlier studies of EMPRESS (Kojima et al. 2020; Isobe et al. 2022a; Nakajima et al. 2022). Three of the ten EMPG candidates have a metallicity below the 10% solar value ($12 + \log(O/H) < 7.69$) and are thus confirmed to be EMPGs, but still no object with metallicity below $0.01\ Z_\odot$ is found. The remaining seven sources are turned out to have slightly higher metallicities (0.1–$0.5\ Z_\odot$).

Table 2.4 Extinction-corrected emission-line fluxes of the 10 EMPG candidates

ID	[O II]λ3727	[O II]λ3729	Hδ	Hγ	[O III]λ4363	Hβ	[O III]λ4959	[O III]λ5007	Hα	[N II]λ6583
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
J1034-0221	30.4 ± 0.6	43.2 ± 0.6	25.6 ± 0.2	46.4 ± 0.2	11.8 ± 0.1	100.0 ± 0.3	165.3 ± 0.3	483.2 ± 0.8	272.2 ± 0.9	2.1 ± 0.0
J1244+2828	74.1 ± 0.5	109.2 ± 0.4	25.0 ± 0.4	44.8 ± 0.3	4.5 ± 0.2	100.0 ± 0.5	158.8 ± 0.7	478.0 ± 0.9	276.7 ± 0.5	3.9 ± 0.2
J1305+2852	73.6 ± 0.4	89.8 ± 0.2	22.8 ± 0.4	45.9 ± 0.4	2.0 ± 0.2	100.0 ± 0.5	124.4 ± 0.6	374.2 ± 1.1	280.4 ± 0.6	14.9 ± 0.1
J1432+0611	64.0 ± 0.5	90.0 ± 0.3	24.3 ± 0.4	45.5 ± 0.4	8.7 ± 0.3	100.0 ± 0.6	184.1 ± 0.6	514.6 ± 1.1	271.7 ± 0.6	4.3 ± 0.1
J1526+1610	61.5 ± 0.1	85.7 ± 0.2	21.9 ± 0.6	44.0 ± 0.8	7.1 ± 0.4	100.0 ± 0.8	204.3 ± 1.0	607.3 ± 1.8	271.2 ± 0.9	6.4 ± 0.2
J1604+1459	69.8 ± 0.3	96.1 ± 0.4	23.2 ± 0.5	43.5 ± 0.6	6.5 ± 0.5	100.0 ± 0.7	149.3 ± 0.4	442.6 ± 0.5	269.0 ± 0.3	5.9 ± 0.1
J1616+1453	91.3 ± 0.3	133.8 ± 0.1	22.8 ± 0.7	46.2 ± 0.7	1.9 ± 0.3	100.0 ± 0.9	102.1 ± 1.3	308.3 ± 1.1	281.3 ± 0.8	21.0 ± 0.3
J1637+1729	27.7 ± 0.1	38.9 ± 0.2	28.6 ± 0.9	47.9 ± 1.0	12.1 ± 0.5	100.0 ± 0.9	184.3 ± 0.7	554.5 ± 1.9	282.8 ± 0.8	2.6 ± 0.1
J1804+0008	13.6 ± 0.8	18.2 ± 0.7	26.3 ± 0.2	45.6 ± 0.4	14.9 ± 0.2	100.0 ± 0.5	176.0 ± 0.7	470.3 ± 1.0	271.8 ± 0.8	1.0 ± 0.0
J2136+0414	14.5 ± 1.1	20.7 ± 1.3	25.6 ± 0.9	44.8 ± 0.8	12.7 ± 0.5	100.0 ± 0.9	130.4 ± 0.9	387.9 ± 1.5	269.1 ± 0.7	0.9 ± 0.1

Note—(1) ID. (2) – (11) Extinction-corrected flux and 1σ error normalized by the Hβ flux and error.

2.5 EMPG Sample

Here we describe our EMPG sample. Including the three EMPGs newly identified in this study (Sec. 2.1), EMPRESS has delivered 14 EMPGs so far below $Z < 0.1$ $Z_\odot$whose stellar masses are particularly low ($M_* \sim 10^4$–10^8 $M_\odot$; Kojima et al. 2020; Isobe et al. 2021; Nakajima et al. 2022; Xu et al. 2022). Adding the other 89 local EMPGs that are compiled from the literature (Nakajima et al. 2022), we use the sample of 105 EMPGs in the following statistical analysis. This is the largest sample constructed below $Z < 0.1$ $Z_\odot$ fully based on the direct temperature method. Basic key properties of the 105 EMPGs, such as metallicity, stellar mass (i-band magnitude), and SFR (UV magnitude and Balmer emission) are summarized in (Nakajima et al. 2022).

2.6 Chemical Properties

We investigate the relation among stellar mass, metallicity, and sSFR of our EMPGs to understand the chemical properties of these galaxies. To obtain M_* and SFR of our EMPG sample, we exclude galaxies whose rest-frame i-band photometry and Hα or Hβ fluxes are not available from the literature. We use 80 EMPGs in total.

We estimate stellar masses in the manner similar to Isobe et al. (2021), using the SED model code, Beagle (Chevallard & Charlot 2016). We run the Beagle code by changing the parameters of stellar mass M_* and maximum stellar age $t_{\rm max}$ in the range of $\log(M_*/M_\odot) = 4.0 - 10.0$ and $\log(t_{\rm max}/{\rm yr}) = 4.0$–$8.0$, respectively. We derive a mass-luminosity relation (linking M_* and absolute i-band luminosity) through a linear fit to the properties obtained from our model SEDs. These models are constructed assuming a fixed metallicity of $12+\log({\rm O/H}) = 8.2$, a constant star-formation history, a dust-free environment, and a Chabrier (2003) IMF. We estimate stellar mass of our EMPGs from their i-band magnitudes using this relation.

SFRs are calculated based on the dust-corrected Hα fluxes, assuming the Kennicutt (1998) calibration with a Chabrier (2003) IMF. If the Hα line is saturated, we use a dust-corrected Hβ line instead and use the intrinsic line ratio of Hα/Hβ = 2.86 for the case B recombination for typical Hii regions (i.e., $T_{\rm e} = 10^4$ K and $n_{\rm e} = 10^2$ cm^{-3}; Osterbrock & Ferland 2006).

We show the mass-metallicity relation (MZR) of our EMPGs in Figure 2.6, comparing with the star-forming galaxies at $z \sim 0$ (Andrews & Martini 2013) and $1 < z < 4$ (Sanders et al. 2020). The data of Andrews & Martini (2013) are the composite spectra that are binned in both M_* and SFR, while the data of Sanders et al. (2020) are individual galaxies. All of the data have the direct-method metallicity measurements.

We use analytic chemical evolution models to understand whether our EMPGs satisfy the balance between inflow and outflow, just like typical local galaxies. We model metallicities of $z \sim 0$ SDSS galaxies by employing the ideal gas-regulator model with steady inflow and outflow (Lilly et al. 2013). In this model, the equilibrium value of metallicity $Z_{\rm eq}$ is expressed as

Table 2.5 Best-fit parameters of Equation 2.1.

$\log y$	λ_{10}	a	${\varepsilon_{10}}^{-1}$(Gyr)	b
(1)	(2)	(3)	(4)	(5)
9.09 ± 0.17	0.55 ± 0.42	-0.44 ± 0.071	0.85 ± 0.47	-0.43 ± 0.07

Note— (1) Yield in units of $12 - \log(\mathrm{O/H})$. (2) Coefficient of the mass-loading factor in Equation 2.2. (3) Exponent of the mass-loading factor in Equation 2.2. (4) Coefficient of the star formation efficiency in Equation 2.3. (5) Exponent of the star formation efficiency in Equation 2.3.

$$Z_{\mathrm{eq}} = Z_0 + \frac{y(1-R)}{(1-R)+\lambda+\varepsilon^{-1}\left\{(1-R)(1+\beta-b)\mathrm{sSFR}-\frac{1.2}{t}\right\}} \tag{2.1}$$

where λ is the mass-loading factor, $R = 0.4$ is a return fraction, ε is the star formation efficiency, $\beta = -0.1$ is the logarithmic slope of the mass dependence of the sSFR, and t is set to 13.8 Gyr. As in Lilly et al. (2013), we assume that both ε and λ are represented by power laws in the stellar mass of the galaxy, i.e.,

$$\lambda = \lambda_{10} {m_{10}}^{a} \tag{2.2}$$

$$\varepsilon = \varepsilon_{10} {m_{10}}^{b} \tag{2.3}$$

where m_{10} is the stellar mass in units of 10^{10} $M_\odot$. We fit using Equation (2.1) to $Z(M_*, \mathrm{SFR})$ data for SDSS galaxies at $z \sim 0$ obtained by Andrews & Martini (2013). We have five free parameters, ε_{10}, λ_{10}, a, b, and y in Equation (2.1). We assume the infall metallicity Z_0 to be 0. We fit to the data using the `curve_fit` function from the Python package scipy. The best-fit parameters are given in Table 2.5. The dashed lines in Figure 2.6 represent the obtained model of $Z(M_*, \mathrm{sSFR})$ at the fixed sSFR of -1.5, -0.5, 0.5, 1.5, and 2.5 Gyr^{-1}. As this model does not explain a metallicity floor, we introduce a non-zero infall metallicity to consider IGM metal enrichment. Without changing the other parameters, the infall metallicity of $Z_0 = 0.0034$ $Z_\odot$, instead of 0, minimize the mean squared error between the $Z(M_*, \mathrm{SFR})$ of the model and that of the $z \sim 0$ SDSS galaxies. The model with $Z_0 = 0.0034$ $Z_\odot$ is shown by the solid lines in the Figure 2.6 and used in the following discussions.

We divide our EMPGs into two groups by their metallicities with $12 + \log(\mathrm{O/H}) = 6.69$–$7.38$ and $12 + \log(\mathrm{O/H}) = 7.38$–$7.69$. The mean and standard deviation of the metallicity, stellar mass, and SFR in each group are $12 + \log(\mathrm{O/H}) = 7.20 \pm 0.12$ (7.56 ± 0.08), $\log(M_*/M_\odot) = 5.13 \pm 0.94$ (6.45 ± 0.75), and $\log \mathrm{SFR} = -2.44 \pm 1.33$ (-1.35 ± 2.24) for the group with $12 + \log(\mathrm{O/H}) = 6.69$–$7.38$ (7.38–7.69). The metallicity of the model at the same stellar mass and SFR, $Z(M_*, \mathrm{SFR})$, are $Z(\log(M_*/M_\odot) = 5.13, \mathrm{SFR} = -2.44) = 7.15$ and $Z(\log(M_*/M_\odot) = 6.45, \mathrm{SFR} = -1.35) = 7.60$. We find that the prediction on the low-metallicity regime of the model, which is fitted to the galaxies with higher metallicities, agree with the mean values of the EMPGs within the scatter. This result suggests that on average, our EMPGs are in equilibrium i.e., satisfying the balance between gas inflow and outflow.

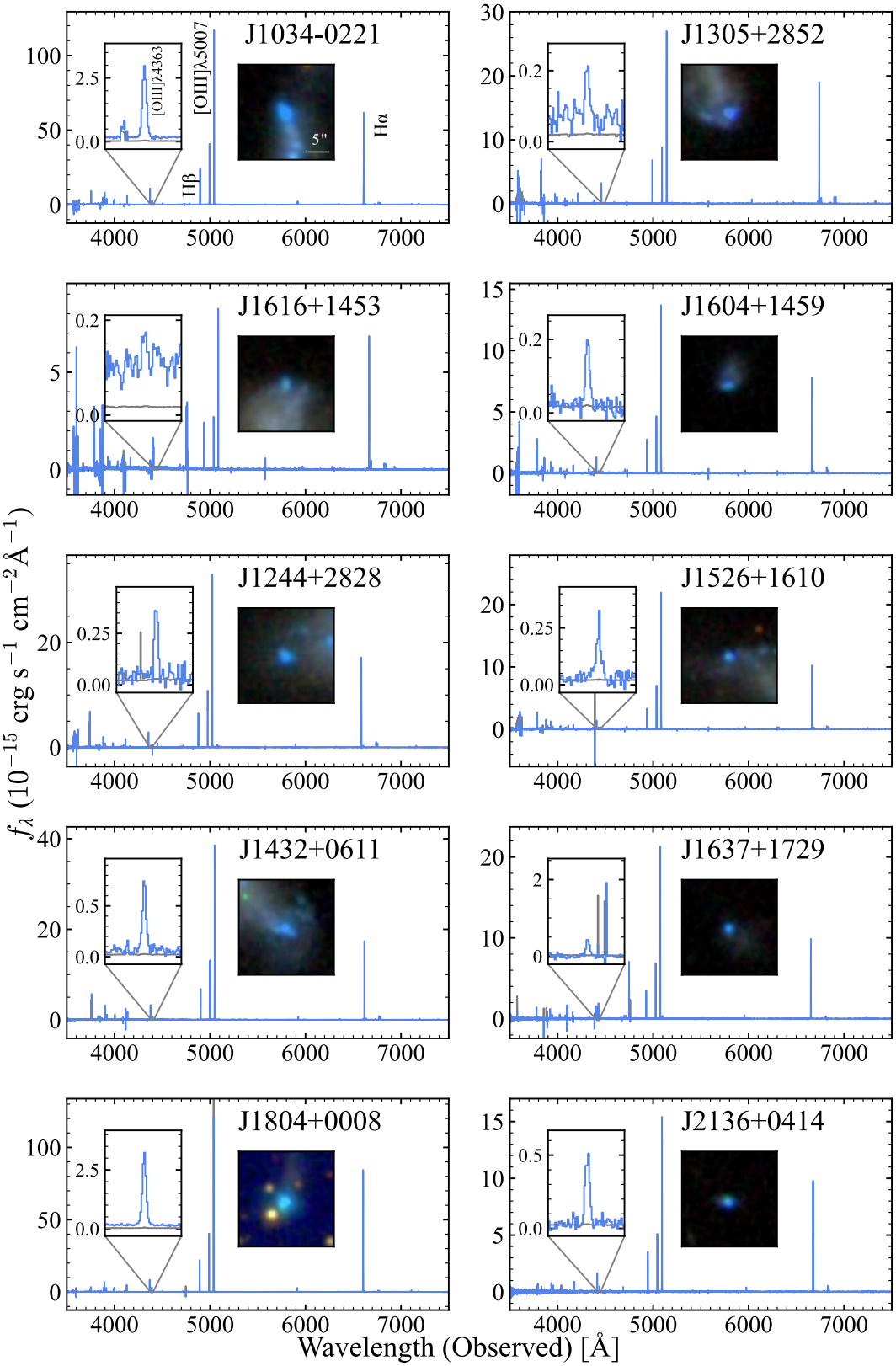

Fig. 2.4 MagE spectra of the ten EMPG candidates. The gray lines indicate error spectra. The left-hand inset of each panel presents an enlarged view of the spectrum around [O III]λ4363. The right-hand inset of each panel shows $20'' \times 20''$ cutout *gri*-composite images from SDSS.

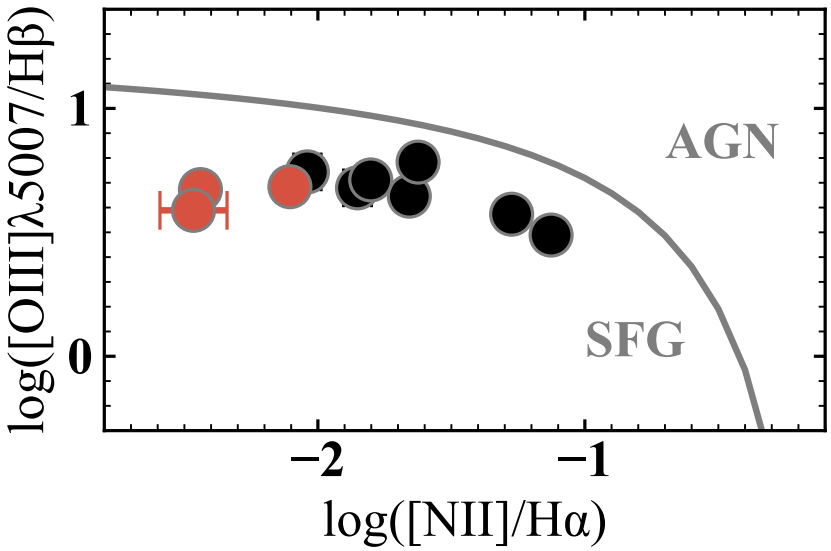

Fig. 2.5 Ten MagE objects on the BPT diagram. The flux ratios are derived from extinction-corrected line flux of MagE spectra. The red (black) points indicate objects with metallicities below (above) $0.1Z_{\odot}$. The gray line present the demarcation line defined by Kauffmann et al. (2003) which divides regions dominated by star-forming galaxies (SFGs) and AGNs.

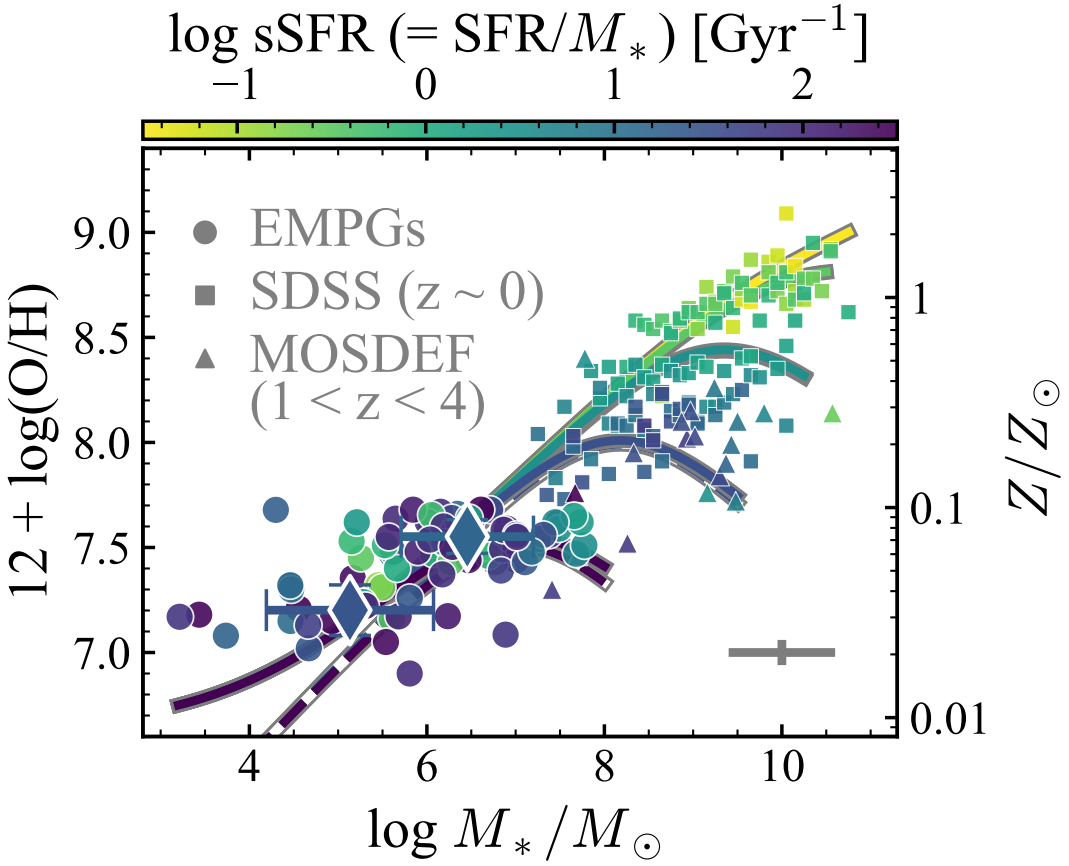

Fig. 2.6 Mass-metallicity relation, color-coded by sSFR. The circles and diamonds indicate individual and average values binned by metallicities of our EMPGs, respectively. We also show comparison samples of $z \sim 0$ galaxies from SDSS (squares; Andrews & Martini 2013) and high-z ($1 < z < 4$) galaxies from MOSDEF survey (triangles; Sanders et al. 2020). The solid (dashed) lines represent the best-fit model to the $z \sim 0$ SDSS galaxies at the fixed sSFR of -1.5, -0.5, 0.5, 1.5, and 2.5 Gyr^{-1}, with the infall metallicity of $Z_0 = 0.0034\ Z_\odot$ ($Z_0 = 0$).

Chapter 3

Searching for High-z EMPGs

We extend our technique to a high-z EMPG search using the first data release of deep JWST/NIRCam imaging to search for metal-poor galaxy candidates at $z \sim 4$–5.

3.1 Extending the Technique to the High-z EMPG Search

The preceding chapter presented a systematic search for extremely metal-poor galaxies in the local universe and discussed their physical origins. A key result of that investigation, consistent with previous searches, was the non-detection of any so-called Hyper Metal-Poor Galaxies (HMPGs) with metallicities below one percent of the solar value. The existence of this observational "metallicity floor" is a long-standing puzzle in the local universe. However, it remains unknown whether this is a universal physical constraint that holds throughout cosmic time or a phenomenon exclusive to the chemically evolved, present-day cosmos. To answer this question, it is necessary to push the observational frontier into the early universe.

Galaxies in the early universe are predicted to have, on average, lower metallicities, making it a more probable environment to find HMPGs and perhaps even more primordial systems that have only recently been enriched by the first stars. Until now, however, it has been technically prohibitive to measure the rest-frame optical emission lines that are indispensable for accurately determining metallicity in such distant galaxies.

This situation has been transformed by the advent of the JWST. The unprecedented infrared sensitivity and spatial resolution of JWST have made it possible to observe the rest-frame optical light of distant galaxies, which has been redshifted into near-infrared wavelengths, at unparalleled depths. With this capability, we finally possess the powerful tools necessary to directly probe the chemical composition of galaxies in the early universe.

This chapter leverages this new observational window by extending the broadband color excess method, which was developed and validated in the previous chapter, to the high-redshift universe. Based on the same physical principles used to select EMPG candidates with SDSS filters in the local universe, we will now use the near-infrared filters of JWST/NIRCam to efficiently select EMPG candidates at a redshift of approximately $z \sim 4 - 5$. The primary objective of this chapter is to test for the existence of the metallicity floor at these early cosmic times, thereby providing new

constraints on the physics of the first chemical enrichment processes in galaxies.

3.1.1 SED Model

To investigate if we can use NIRCam photometric data to isolate EMPGs from other type objects of metal-rich galaxies, QSOs, and stars, we use SED models of EMPGs, metal-rich galaxies, QSOs, and stars used in Section 2.1. For the EMPG models, in addition to those used in Section 2.1, we employ the SED models of PopIII (Nakajima & Maiolino 2022) with metallicities of $Z = 0$, stellar ages of $6.0 \leq \log(\mathrm{age/yr}) \leq 7.8$ with a step of 0.2, ionization parameters of $-2.0 \leq \log U \leq -0.5$ with a step of 0.5, dust attenuations of $E(B-V) = [0,\ 0.01,\ 0.02,\ 0.05]$, and a Salpeter (1955) IMF with mass ranges of 1–100 $M_\odot$, 1–500 $M_\odot$ and 50–500 $M_\odot$(Schaerer 2003). We also use non-PopIII EMPG models of Nakajima & Maiolino (2022) with metallicities of $Z = [1/1400,\ 1/140,\ 0.01,\ 0.02,\ 0.05]\ Z_\odot$, stellar ages of $6.0 < \log(\mathrm{age/yr}) < 7.8$ with a step of 0.2, ionization parameters of $-3.0 \leq \log U \leq -1.0$ with a step of 0.5, dust attenuations of $E(B-V) = [0.,\ 0.01,\ 0.02,\ 0.05]$, and a Kroupa (2001) IMF with upper mass limits of 100 $M_\odot$ and 300 $M_\odot$. We obtain colors of NIRCam bands with the model SEDs, which include random noise corresponding to the observational errors, in similar procedures as described in Section 2.1.1. Figure 3.1 shows color-color diagrams of the bands of F200W, F277W, F356W, and F444W. The red contours in Figure 3.1 represent EMPG models at $z \simeq 3.8$–4.8 (i.e., $z \sim 4$–5) whose Hβ+[O III] and Hα fall in F277W and F356W, respectively, while continuum with no strong lines is sampled by F200W and F444W. As shown in Figure 3.1, $z \sim 5$ EMPG candidates can be selected by the color-excess technique.

3.1.2 JWST Data

We analyze four JWST/NIRCam datasets taken by the Early Release Observations (ERO; Pontoppidan et al. 2022) and Early Release Science (ERS) programs: ERO SMACS J0723, ERO Stephan's Quintet, ERS Cosmic Evolution Early Release Science (CEERS; Finkelstein et al. 2017, 2022b), and ERS GLASS (Treu et al. 2017, 2022). The total area used in our analysis is 87.4 arcmin^2. The 5σ limiting magnitudes in the F356W band ranges between 28.9 and 29.9 mag. We use the photometric catalogs constructed by Harikane et al. (2022). The data reduction and photometry are described in Harikane et al. (2022). Following Harikane et al. (2022), we measure the object colors with the 0.″3-diameter aperture magnitude in PSF-matched images. For simplicity, we use SExtactor `MAG_AUTO` as a total magnitude.

3.1.3 Candidate Selection

We select $z \sim 4$–5 EMPG candidates based on broadband colors by adopting the following color criteria:

$$\mathrm{F356W} - \mathrm{F444W} < -0.8 \tag{3.1}$$

$$\mathrm{F277W} - \mathrm{F356W} > 0.3 \tag{3.2}$$

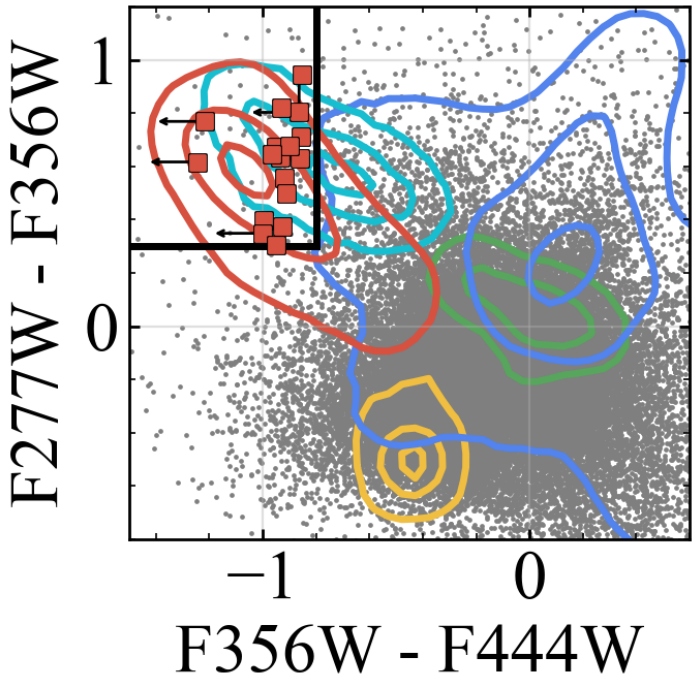

Fig. 3.1 Color-color diagrams of F277W−F356W vs. F356W−F444W. The gray points show NIRCam sources which are detected at $> 3\sigma$ in all three bands of the each panel. The red contours illustrate the distribution of EMPGs at $z \simeq 3.8$–4.8 with metallicities of $Z = 0$–$0.05\ Z_\odot$ based on the SED model, while the other colors correspond to other populations of metal rich galaxies at $z \simeq 3$–8 (blue), QSOs at $z \simeq 3$–8 (green), and galactic stars (yellow). The cyan contours show [O III] emitters at $z \simeq 6.6$–6.7. The black lines in the left panel are our color criteria to select the EMPG candidates at $z \sim 4$–5. The selected 17 candidates are marked with the red squares.

In calculation of colors, a magnitude fainter than a 3σ limiting magnitude is replaced by the 3σ limiting magnitude. We further require $> 3\sigma$ detections within $0\farcs2$-diameter circular apertures in the F356W band. Note that we do not rely on a machine-learning technique because the number of JWST sources that we need to visually inspect is small even if we adopt simple color cuts due to the smaller size of the JWST source catalogs than that of SDSS. In the future with more data, a machine-learning technique would be useful for the high-z EMPG search as well.

We visually inspect images of the candidates to remove spurious sources and sources affected by bright neighbors and diffraction spikes. We then compare SEDs of the candidates, including bluer bands (F200W, F150W, F115W, and F090W if available), with an EMPG spectral model at $z \simeq 4.6$ with $Z = 0$, $\log U \simeq -2.0$, a stellar age of $\sim 10^6$ yr, and $E(B-V) \simeq 0.01$ (as shown in Figure 3.3). The EMPG model is normalized by average magnitudes in the F277W, F356W, and F444W bands.

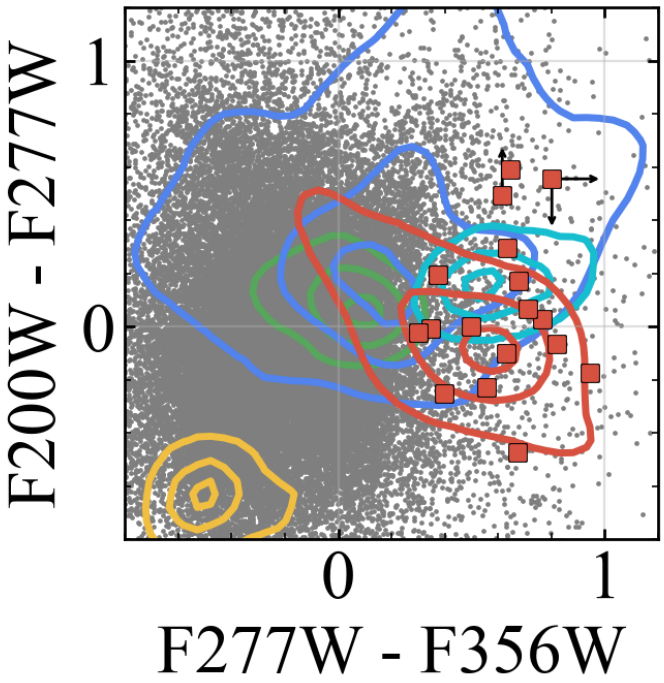

Fig. 3.2 Same as Figure 3.1, but for F200−F277W vs. F277W−F356W.

Additionally, we use multi-band images obtained by the Hubble Space Telescope (HST) ACS in the SMACS J0723 and CEERS fields, which are available on the websites of RELICS (Coe et al. 2019)[*1] and CEERS[*2], respectively. We perform photometry for the candidates on the HST images in the F606W and F814W bands of the CEERS fields, and the F435W, F606W, and F814W bands of the SMACS J0723 field. We remove sources whose SEDs are not consistent with the EMPG model.

One of the main sources removed from the candidate list is a [O III] emitter at $z \simeq$ 6.6–6.7 because it shows strong [O III] emission in F356W and no strong emission in F444W. The black (gray) line in Figure 3.4 (Figure 3.3) represents an [O III] emitter spectral model at $z \simeq 6.7$ with $Z = 0.2\ Z_\odot$, $\log U \simeq -2.5$, stellar age of $\sim 10^{6.2}$ yr, and $E(B-V) \simeq 0.01$. While the redshift range that the [O III] emitter model satisfies our color criteria is narrower ($\Delta z \sim 0.1$) than EMPGs ($\Delta z \sim 1$), strong [O III] emission-line emitters are abundant at $z \gtrsim 6$ (e.g., Matthee et al. 2022) and can be contaminants in our sample. We distinguish between $z \sim$ 4–5 EMPGs and $z \sim 6.7$ [O III] emitters based on the Lyman break at $\sim$ 600–700 nm and $\sim$ 930 nm, respectively. The expected magnitude in the F090W band of the [O III] emitter model spectrum is 0.5 mag fainter than that of the EMPG model spectrum due to

[*1] https://archive.stsci.edu/prepds/relics
[*2] https://ceers.github.io/releases.html

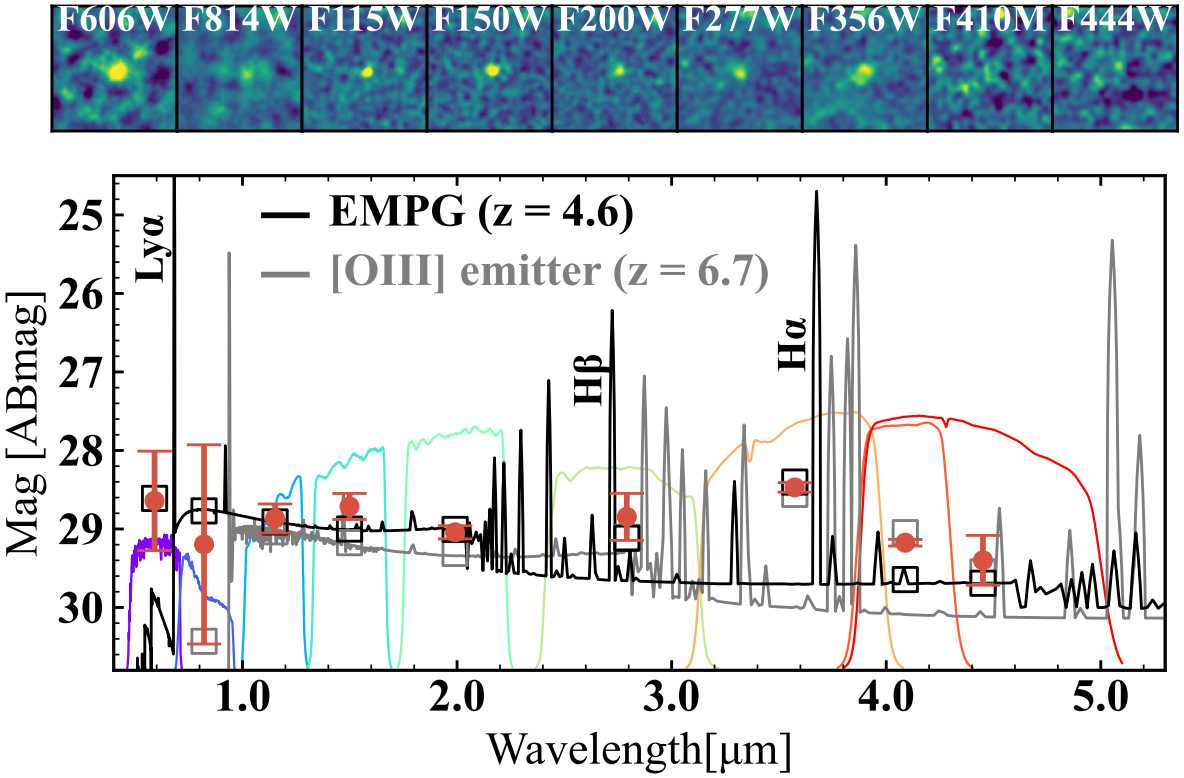

Fig. 3.3 SED of one of the JWST EMPG photometric candidates (red circles with error bars) at R.A. = 14h 19m 31.699s, Dec. = +52d 50m 39.660s (J2000). The candidates present enhanced F356W and F277W photometry, which are explained by the intense Hα and Hβ and modest [O III] lines, as indicated by an EMPG model SED at $z \simeq 4.6$ (black line and squares). Note that F606W photometry of the EMPG model is boosted by strong Lyα emission. the gray line and squares show SED of an [O III] emitter model spectrum at $z \simeq 6.7$, which mimic the colors of $z \sim$ 4–5 EMPGs. The [O III] emitter model SED cannot explain the F606W photometry of the candidate. Colored lines show NIRCam and HST filter throughputs of F606W, F814W, F115W, F150W, F200W, F277W, F356W, F410M, and F444W, from left to right. The upper nine panels show NIRCam and HST images ($1\farcs5 \times 1\farcs5$).

the Lyman break. Therefore, we remove sources with drops in F090W. We also use HST multi-band images covering wavelengths shorter than $\sim$ 930 nm to remove $z \sim 6.7$ [O III] emitters. We remove objects that are not detected in these HST bands from the candidate list because EMPGs at $z \sim$ 4–5 would be brighter than the 2σ limiting magnitudes according to our model predictions. Figure 3.4 illustrates one of the potential [O III] emitters that satisfies our color criteria but is removed from the candidate list.

Finally, we find 17 EMPG candidates that exhibit characteristic colors similar to $z \sim$ 4–5 EMPGs. Figure 3.3 shows the EMPG spectral model and SED for one of the 17 candidates. The signal-to-noise ratio within $0\farcs2$-diameter circular apertures in the F356W band for all 17 candidates is greater than 4. Among the 17 candidates, one is marginally detected in the F606W (2.6σ) and F814W (1.5σ) bands, which

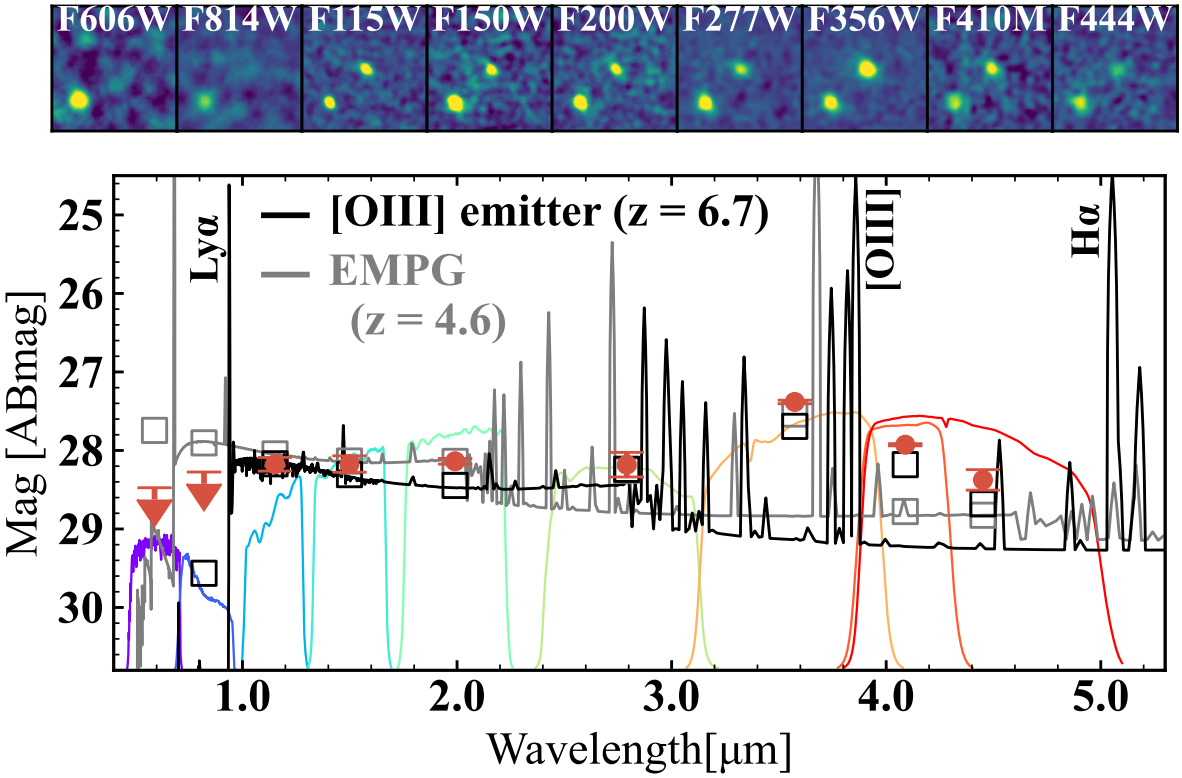

Fig. 3.4 Same as Figure 3.3, but for a possible [O III] emitter at R.A. = 14h 19m 26.971s, Dec. = +52d 48m 37.980s (J2000). The black (gray) line and squares show SED of an $z \simeq 6.7$ [O III] emitter ($z \simeq 4.6$ EMPG) model spectrum.

are consistent with the SED of the EMPG model (Figure 3.3). Of the remaining 16 candidates, eight have F090W photometry, showing flat colors in the F090W, F150W, and F200W bands. However, we cannot place constraints on the remaining 8 candidates using either F090W photometry or HST images because of the shallow detection limits of the HST images, which are at around $\sim$ 26–29 mag, and the lack of F090W photometry. While a spectroscopic confirmation is necessary to further investigate the contamination, we retain the 17 candidates in the $z \sim$ 4–5 EMPG sample in this paper based on the NIRCam and HST photometry.

3.1.4 Line Ratio Estimation

We estimate Hβ+[O III] and Hα line fluxes of the candidates with F277W and F356W, respectively, with the continuum estimated with F444W, in the same manner as Isobe et al. (2021). The candidates show large color excesses in the F356W band of F356W $-$ F444W ~ -1, which are mainly caused by strong Hα lines. We regard the observed F444W-band flux density $f_{\rm F444W}$ as tracers of the stellar continuum with a negligible effect of strong emission lines. We also assume that the flux densities per unit frequency of stellar continuum is constant over the wavelength range of Hβ,

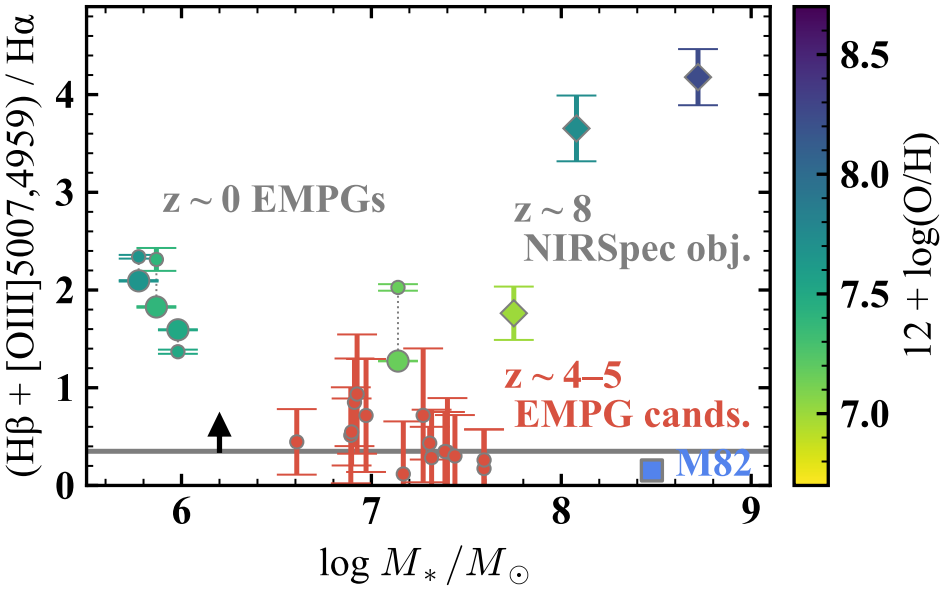

Fig. 3.5 Strong line ratio of (Hβ+[O III]$\lambda\lambda$5007, 4959)/Hα as a function of stellar mass. The $z \sim$ 4–5 EMPG candidates are shown with the red points with errors of the line ratio. The green points and circles denotes line ratios of the photometric estimates and the spectroscopic results, respectively, for $z \sim 0$ EMPGs from this thesis (Section 2.1) and the literature (Izotov & Thuan 1998). The diamonds are the $z \simeq$ 4–8 galaxies identified by NIRSpec spectroscopy (Nakajima et al. 2023). The $z \sim 0$ EMPGs and NIRSpec objects are color-coded according to metallicities based on the direct T_e method. The gray horizontal line represents the theoretical lower limit with negligible metal emission of [O III], i.e., Hβ/Hα, for the case B recombination at $T_e = 10^4$ K and $n_e = 10^2$ cm^{-3}.

[O III], and Hα. We then calculate the Hα flux $f_{H\alpha}$ as follows:

$$f_{H\alpha} = (f_{F356W} - f_{F444W}) \times \frac{c}{\lambda_{F356W}} \times \Delta\lambda_{F356W}, \tag{3.3}$$

where f_{F356W}, λ_{F356W}, and $\Delta\lambda_{F356W}$ represent the observed F356W-band flux density, the central wavelength of the F356W-band filter, and the width of the F356W-band filter, respectively. Errors of the $f_{H\alpha}$ is calculated from photometric errors. We estimate the sum of Hβ+[O III] fluxes and their errors in the same manner as the one of the Hα fluxes. Figure 3.5 presents (Hβ+[O III])/Hα ratios as a function of stellar mass derived by the mass-luminosity relation used in Section 2.6.

For comparison, we photometrically estimate line ratios of (Hβ+[O III])/Hα for $z \sim 0$ EMPGs, including the three newly identified EMPGs (Section 2.1) and I Zw18 NW (Izotov & Thuan 1998), one of the well-known EMPGs, in the same manner as the JWST EMPG photometric candidates. The Hβ+[O III] and Hα line fluxes are estimated with SDSS g- and r-band photometry, respectively, with the continuum estimated with i-band photometry. Errors of the estimated line fluxes are calculated from photometric errors. Compared to the spectroscopic results of the line ratios for these three EMPGs (Section 2.4) and I Zw18 NW (Izotov & Thuan 1998), the spectroscopic results agree with the photometric estimates within 0.05 – 0.20 dex.

We confirm that the method of the photometric estimates does not produce the small (Hβ+[O III])/Hα ratios by estimation systemtics. We can also compare our results with those obtained from high-z galaxies in a similar redshift range, whose optical emission lines have been spectroscopically identified using JWST/NIRSpec. We present 10 galaxies at $z = 4$–8.5 in Figure 3.5 that are found from the three major public spectroscopy programs of ERO, GLASS, and CEERS with reliable metallicity determinations based on [O III]λ4363 (Nakajima et al. 2023). We estimate the Hα flux from the Hβ flux for the NIRSpec objects assuming the line ratio of Hα/Hβ = 2.86 for the case B recombination at $T_{\rm e} = 10^4$ K and $n_{\rm e} = 10^2$ cm^{-3}.

Comparing the $z \sim 4$–5 EMPG candidates with the $z \sim 0$ EMPGs and the $z \simeq 4$–8 NIRSpec galaxies in Figure 3.5, we find that the $z \sim 4$–5 EMPG candidates exhibit exceptionally small (Hβ+[O III])/Hα ratios. Interestingly, 14 out of the 17 candidates have very low (Hβ+[O III])/Hα ratios consistent with negligible [O III] emission (i.e., Hβ/Hα; gray horizontal line in Figure 3.5), indicative of PopIII-like galaxies. Although dust extinctions are not corrected for the ratios of JWST EMPG candidates as well as $z \sim 0$ EMPGs and NIRSpec objects, the ratios increase only by 0.05 dex by the typical extinction values for local EMPGs of $E(B-V) = 0.1$, which is much smaller than the error bars (~ 0.4 dex).

3.2 FOCAS follow-up observations

We performed deep spectroscopic follow-up observations for the eight EMPG candidates at $z \sim 4$–5 identified in the JWST CEERS field. This program (PI: M. Nishigaki) aimed to detect strong Lyα emission lines expected from these metal-poor, young star-forming galaxies, which would allow us to spectroscopically confirm their redshifts and analyze their interstellar medium properties.

The observations were conducted using the Faint Object Camera and Spectrograph (FOCAS; Kashikawa et al. 2002) mounted on the Subaru Telescope on UT 2023 June 23 and 24 in the Multi-Object Spectroscopy (MOS) mode. We used the VPH650 grating combined with the Y47 order-cut filter and a slit width of 0.″7. This configuration provided a spectral coverage of $\lambda = 5300$–7700 Å with a spectral resolution of $R \sim 1000$–2000. We utilized two MOS masks for the targets (Figure 3.6). The on-source integration times were 2.7 hr (8×1200 s) for the first mask (Mask 1) and 3.0 hr (9×1200 s) for the second mask (Mask 2). Observational conditions were clear, with a typical seeing of $\sim 0.''7$. The spectrophotometric standard star HZ 44 was also observed for flux calibration. Data reduction is performed in a standard manner using IRAF/FOCASRED and Python, including bias subtraction, flat-fielding, wavelength calibration, and sky subtraction.

Visual inspection of the reduced 2D spectra reveals no significant emission lines at the expected spatial positions for any of the targets. Figure 3.7 shows the spectrum of one of the targets. To quantify the non-detections, we estimate the 5σ limiting line fluxes of our observations. We randomly place spectral apertures in regions free from sky emission lines and derive the noise level. The resulting 5σ limiting line fluxes are 5.3×10^{-18} erg s^{-1} cm^{-2} for Mask 1 and 3.0×10^{-18} erg s^{-1} cm^{-2} for Mask 2, respectively. The non-detection of Lyα emission allows us to place upper limits

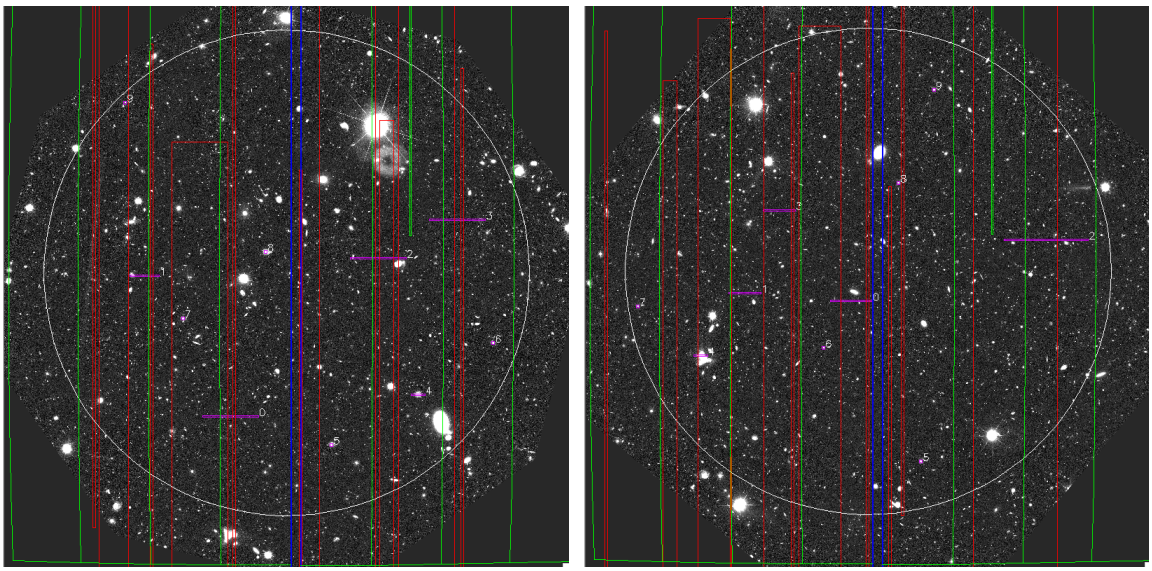

Fig. 3.6 A screenshot of the mask design using the Subaru/FOCAS Mask Design Program (MDP) for Mask 1 (left) and Mask 2 (right). The effective FOV is indicated by the white circle, while the green outlines and the vertical blue lines represent the CCD boundaries and the chip gap, respectively. The positions of the slit apertures are marked by pink segments. The red rectangles display the expected spectral coverage areas on the detectors.

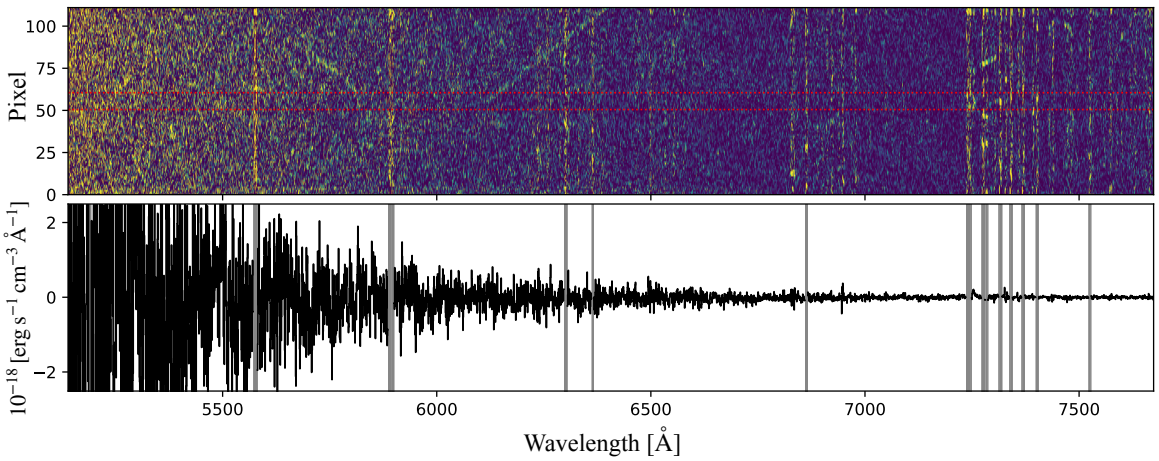

Fig. 3.7 Example spectra of one of the EMPG candidates. **Top**: The reduced 2D spectrum. The horizontal red dotted lines indicate the spatial aperture of $\sim \pm 1''$ centered on the target used for extraction. **Bottom**: The extracted 1D spectrum. No significant emission lines are detected in either panel.

on the Lyα escape fraction, $f_{\rm esc}^{\rm Ly\alpha}$. We first estimate the intrinsic Lyα flux for each candidate. We utilize the observed photometric excess in the JWST long-wavelength bands (F356W − F444W), which acts as a proxy for Hα emission strength relative to the continuum. Assuming Case B recombination physics Osterbrock & Ferland (2006), we convert the inferred Hα fluxes into expected intrinsic Lyα fluxes. The predicted intrinsic fluxes range from 1.7×10^{-17} to $3.2 \times 10^{-17}\,{\rm erg\,s^{-1}\,cm^{-2}}$.

By comparing these expected intrinsic fluxes with our observational limiting fluxes, we derive the upper limits on the Lyα escape fraction. We find that $f_{\rm esc}^{\rm Ly\alpha}$ is constrained to be < 0.1–0.3 for these targets. The derived upper limits are consistent with previous studies suggesting that the typical Lyα escape fraction for star-forming galaxies at $z \sim 4$–5 is relatively modest, ranging from $\sim 5\%$ to 20% (e.g., Hayes et al. 2011; Begley et al. 2024). Given that these targets are candidates for extremely metal-poor galaxies, we assume the dust content is minimal. Thus, the low escape fraction is likely caused by efficient resonant scattering by neutral hydrogen (H I) gas rather than dust extinction. Alternatively, we note that uncertainties in the photometric redshifts could also place the emission line outside our spectral window.

Chapter 4

Chemical Evolution of Galaxies: Observations

In Section 4.1, we describe the data used in this chapter. Section 4.2 outlines the methods, including the stacking procedure applied to the spectra. Section 4.3 presents the results, covering emission line ratios, metallicity measurements, ionization parameters, the mass-metallicity relation (MZR), and the FMR.

4.1 Data

In this study, we make use of spectroscopic and imaging datasets obtained from multiple JWST programs, including DREAMS, JADES, ERO, and CEERS. These programs collectively provide medium-resolution NIRSpec spectra that cover the key rest-frame optical emission lines, along with complementary NIRCam photometry. The combination of these datasets allows us to construct a large and homogeneous sample of galaxies across a wide redshift range.

4.1.1 DREAMS

We make use of observations from the program GO-4750 (PI: K. Nakajima), titled Deep Reconnaissance of Early Assemblies of Metal-poor Star formation (DREAMS). DREAMS targets the lensing cluster field MACS J0416, enabling access to intrinsically faint galaxies through gravitational magnification. The spectroscopic observations were obtained with JWST/NIRSpec using the MSA and the medium-resolution grating configurations G140M/F070LP and G395M/F290LP, with a resolving power of $R \sim 1000$. The DREAMS survey and its data reduction are partially described in Nakajima et al. (2025), with a full description to be presented in a forthcoming paper (Nakajima et al. in preparation). Among the objects observed in DREAMS, five galaxies (IDs 30001, 30002, 30003, 31001, and 40005; Nakajima et al., in preparation) at redshifts $z \sim 8$ cover the full wavelength range from [O II]λ3727 to [O III]λ5007, so that all key emission lines used for the metallicity determination (i.e., [O II]λ3727, Hγ, [O III]λ4363, Hβ, and [O III]λ5007) are available.

Photometric data for these objects are taken from other JWST programs. We adopt the reduced catalogs published by Ma et al. (2024), Willott et al. (2024), and

Diego et al. (2024). Magnification factors are derived from lensing models of MACS J0416 constructed using the techniques described in Oguri (2010); Oguri et al. (2021); Kawamata et al. (2018), with field-specific information taken from Fu et al. (2025). The magnification factors for the DREAMS objects in our sample are 1.38, 1.42, 1.42, 1.05, and 8.4 for IDs 30001, 30002, 30003, 31001, and 40005, respectively.

4.1.2 JADES

We also use both spectroscopic and photometric data from the third data release of the JWST Advanced Deep Extragalactic Survey (JADES; Eisenstein et al. 2023; Bunker et al. 2024; D'Eugenio et al. 2025). This release provides deep imaging and NIRSpec spectroscopy in the GOODS-S and GOODS-N fields, offering the largest number of galaxies in our sample and covering a broad range of redshifts and stellar masses. The spectroscopic component includes medium-depth and deep observations taken with the NIRSpec microshutter assembly (MSA), covering a spectral range of 0.6–5.3 μm. These observations employ both the low-resolution prism ($R \sim 30$–300) and all three medium-resolution grating configurations: G140M/F070LP, G235M/F170LP, and G395M/F290LP, which provide spectral resolutions of $R \sim 1000$. We use the medium-grating spectra to derive spectral features. All one-dimensional spectra were reduced by the JADES team as part of the official data release (D'Eugenio et al. 2025), and spectroscopic redshifts based on these spectra are also provided.

Photometric data are taken from the official JADES photometric catalog, which was also compiled and released by the survey team (Rieke et al. 2023). The catalog includes observations in up to 14 filters, depending on the field coverage, including NIRCam bands of F090W, F115W, F150W, F200W, F277W, F356W, and F444W. These broad-band data are essential for deriving stellar population properties such as stellar mass and star formation rate via spectral energy distribution (SED) fitting.

We select galaxies for analysis by cross-matching the photometric and spectroscopic catalogs and excluding sources flagged for data reduction problems (DR_flag). We further limit our sample to galaxies with spectroscopic redshifts determined from emission lines, retaining only those classified as flag A (based on high-resolution spectra) or flag B (based on low-resolution spectra).

We also require that each spectrum covers the full wavelength range from [O II]λ3727 to [O III]λ5007, ensuring that all of the key emission lines used for the metallicity determination ([O II]λ3727, Hγ, [O III]λ4363, Hβ, and [O III]λ5007) are available. We confirm that these lines are not affected by the instrumental gap of the grating. From the JADES catalog, a total of 822 galaxies satisfy these criteria.

4.1.3 ERO and CEERS

Our analysis further incorporates spectra and catalogs provided by Nakajima et al. (2023), which combine observations from the Early Release Observations (ERO; Finkelstein et al. 2022a, Proposal ID 2736) and the Cosmic Evolution Early Release Science Survey (CEERS; Finkelstein et al. 2022b, Proposal ID 1345).

The ERO program targeted the SMACS 0723 lensing cluster field. Spectroscopic observations were obtained with JWST/NIRSpec using the MSA and the medium-

resolution grating configurations G235M/F170LP and G395M/F290LP, providing a resolving power of $R \sim 1000$. Among the sources in this field, five galaxies at redshifts $z > 5$ show prominent [O III] and Hβ emission lines, and all five cover the full set of key emission lines used in this study ([O II]λ3727, Hγ, [O III]λ4363, Hβ, and [O III]λ5007). These galaxies have also been observed with JWST/NIRCam in the F090W, F150W, F200W, F277W, F356W, and F444W filters, for which we adopt the photometric catalog described in Harikane et al. (2023a).

The CEERS program provides NIRSpec MSA spectroscopy in the blank field of the EGS. The survey includes both prism ($R \sim 100$) and medium-resolution ($R \sim 1000$) spectra obtained with the G140M, G235M, and G395M gratings. The CEERS medium-resolution spectra catalog contains 74 galaxies at redshifts $z > 4$ with secure detections of [O III] and Hβ emission lines. From this catalog, we use 43 galaxies that cover the full set of key emission lines used in this study (from [O II]λ3727 to [O III]λ5007) and have available NIRCam photometry data. CEERS also offers complementary JWST/NIRCam imaging with the F115W, F150W, F200W, F277W, F356W, F410M, and F444W filters, for which we again use the photometric reductions of Harikane et al. (2023a).

4.1.4 Galaxies at $z > 9.5$

At redshifts beyond $z > 9.5$, the [O III]λ5007 line is redshifted out of the JWST/NIRSpec spectral coverage, which makes it impossible to estimate electron temperatures and thus determine gas-phase metallicities using direct $T_{\rm e}$ method. To explore metallicity evolution further into the early universe, we instead compile measurements from the literature.

Table 4.1 summarizes a compilation of galaxies at $z > 9.5$ with metallicity estimates derived using alternative methods. These include the T_e method when [O III]λ4363 is detected, as well as strong-line diagnostics such as R23, O32, R3, and Ne3O2, depending on the available line detection in each object. The table also includes stellar mass and star formation rate of each galaxy, both of which are derived from SED fitting.

The sample includes five galaxies spanning $z = 9.5$–12.3, and currently represents the earliest known examples of galaxies with individual metallicity measurements. These objects provide crucial benchmarks for probing chemical enrichment during the first ∼500 million years of cosmic history.

In our analysis and discussion of chemical evolution at $z > 9.5$, we use the median values of redshift, stellar mass, and star formation rate derived from this compiled sample to ensure consistency and comparability with the lower-redshift population.

4.1.5 Emission-line Flux Measurements

We first measure fluxes for individual spectra obtained with medium-resolution gratings. Gaussian profiles are fitted to key emission lines including Hα, Hβ, Hγ, and [O III]$\lambda\lambda$5007,4959, utilizing the noise spectrum for weighting. All analysis is conducted on the vacuum wavelength scale. Flux uncertainties are estimated by summing the spectral noise in quadrature within a window of ±FWHM centered on the

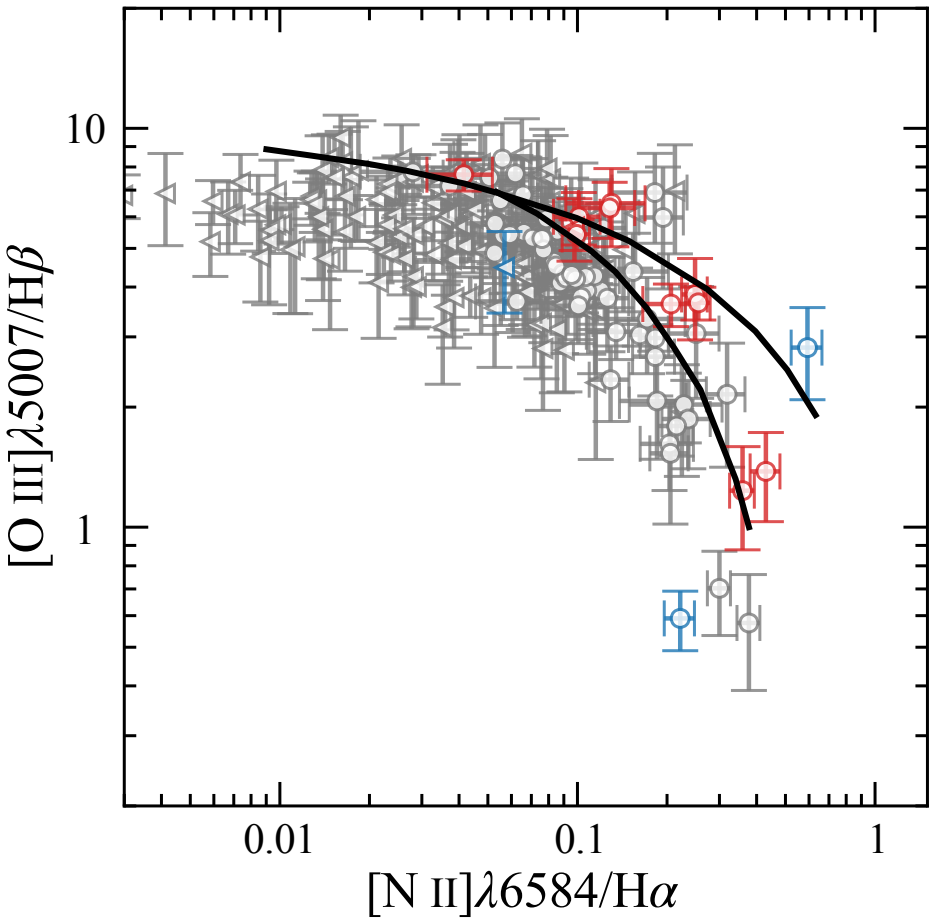

Fig. 4.1 The BPT diagram ([O III]λ5007/Hβvs. [N II]λ6584/Hα) used for the classification of galaxies and the exclusion of AGN. We plot all galaxies from the parent catalog for which the relevant emission lines are covered and at least one line in each ratio is detected with S/N $>$ 3. Circles represent galaxies where all four lines are detected. Triangles indicate 3σ upper or lower limits for galaxies where only one line in a given ratio is detected. The solid curves show the AGN/SFG demarcation lines of Kewley et al. (2001) (upper) and Kauffmann et al. (2003) (lower). Galaxies are color-coded according to their classification. Those identified as AGN based on these criteria are shown in red and are excluded from our final sample. Galaxies exhibiting broad Hα emission are shown in blue and are excluded from our analysis as broad-line AGN, regardless of their position on the diagram. The remaining galaxies, classified as SFGs and used in our analysis, are shown in gray.

Gaussian peak. For some objects, the central wavelengths of the fitted Gaussian peaks deviate from those predicted by the JADES catalog redshifts, with offsets of up to $\sim$20 Å. Most of these redshifts were determined using PRISM spectra, which have lower spectral resolution than the medium-resolution grating data. For the following analysis, we adopt redshifts measured from the Gaussian fitting of the [O III]λ5007 line for all objects.

4.1.6 Sample Selection

For our analysis, we first select galaxies in which both the Hβ and [O III]λ5007 emission lines are detected with signal-to-noise ratios greater than three.

The presence of AGN-driven ionization can bias metallicity calibrations that are optimized for star-forming galaxies and also complicate the determination of stellar masses and star formation rates. To mitigate these effects, we exclude galaxies showing evidence of AGN activity. We first remove sources exhibiting broad-line emission in Hα or Hβ. This includes seven objects in the JADES sample reported by Maiolino et al. (2024), two additional JADES objects we identified (IDs 39435 and 209777), two CEERS objects reported by Harikane et al. (2023b) (IDs 00397 and 02782), and one DREAMS object (ID 30001), which will be presented in Takechi et al. (in preparation). We also exclude ERO 06355 due to potential AGN activity, which is supported by the detection of [Ne IV]λ2423 reported in Brinchmann (2023). In addition, we apply the classical BPT diagnostic diagram based on the [O III]/Hβ and [N II]/Hα line ratios as shown in Figure 4.1. Among galaxies at $z < 6.9$, where Hα falls within the NIRSpec wavelength coverage, we remove 15 further objects that satisfy the AGN selection criteria of Kewley et al. (2001) and Kauffmann et al. (2003) as show in Figure 4.1.

In summary, our final sample consists of 292 galaxies, which are used in the following analysis. The DREAMS survey, though comprising modest sample size of four galaxies, provides a critical contribution to our investigation of the high-redshift frontier. By combining our data, we increase the total number of galaxies at $z > 7.8$ with NIRSpec spectroscopy to 11; the physical properties of these individual sources are listed in Table 4.1. For comparison, previous compilations from ERO, GLASS, and CEERS by Nakajima et al. (2023) and from JADES DR1 by Curti et al. (2024) contained seven and three galaxies in this redshift range, respectively, with uncertainties that made it difficult to clearly discern evolutionary trends. The DREAMS data therefore play a key role by providing robust constraints in a regime where previous studies were limited by large uncertainties.

4.2 Method

The aim of this study is to determine the average spectral properties of galaxies at different redshifts and to characterize their typical physical conditions, such as emission line ratios and electron temperature (T_e)-based metallicities. Since [O III]λ4363 auroral lines required for direct T_e metallicity measurements are faint, spectral stacking is essential to enhance the signal-to-noise ratio and enable reliable detection. To minimize selection biases and ensure a meaningful comparison across cosmic time, we also construct a subsample divided within fixed ranges of stellar mass and star formation rate (hereafter, the fixed M_*–SFR subsample). In this section, we describe the procedures used to estimate stellar masses and star formation rates for individual galaxies. We then detail our stacking methodology and the treatment of uncertainties in the derived physical quantities. Finally, we introduce the photoionization models

employed to interpret the observed line ratios in terms of underlying physical parameters such as ionization parameters.

4.2.1 Stellar Masses and SFRs

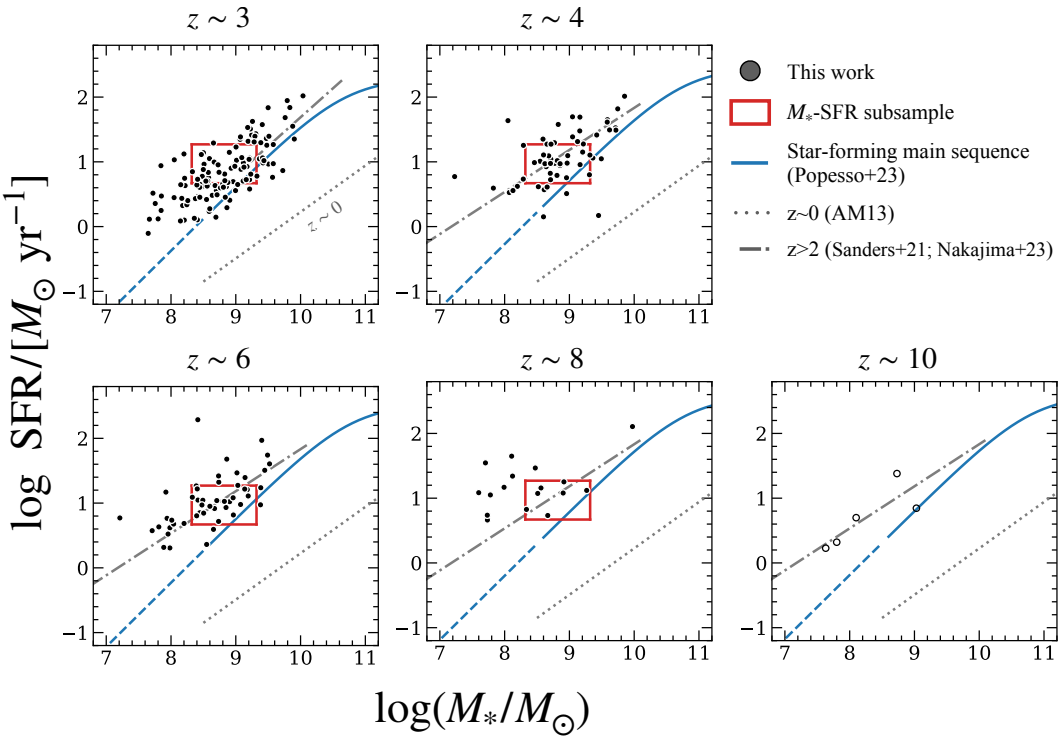

Fig. 4.2 Star formation rate versus stellar mass for the galaxy sample used in this study, shown in five redshift bins: z = 2-4, 4-5.5, 5.5-7, 7-9.5, and z = 9.5-12. Filled gray dots represent our sample galaxies, with stellar masses derived from Prospector SED fitting, and SFRs calculated from dust-corrected Hβ luminosities. In the highest-redshift panel, open gray dots indicate literature data points. Note that the methods used to estimate stellar mass and SFR differ across these studies. The blue solid line in each panel shows the predicted star formation main sequence from the simulation by Popesso et al. (2023) at the corresponding redshift. For comparison, we show samples used in previous FMR studies: the gray dotted line shows the local ($z \sim 0$) relation from Andrews & Martini (2013), while the gray dashed and dot-dashed lines correspond to $z \sim$ 4-10 observations from Nakajima et al. (2023) and Curti et al. (2023), respectively. In the $z \sim 3$ panel, gray symbols represent the sample from Sanders et al. (2021). Red rectangles, shown in all panels, denote the selection criteria for the subsample, which is designed to minimize the sample bias and effects of FMR definition by comparing galaxies within a consistent region of the SFR-M_* plane.

We estimate stellar masses and SFRs for all our samples using a consistent methodology. The stellar masses are derived by fitting the SED models to NIRCam photometry, while the SFRs are calculated from emission-line fluxes measured from the NIRSpec spectra.

For the JADES sample, the SED fitting inputs are photometric data from the JADES catalog (Rieke et al. 2023), specifically total magnitudes derived from circular aperture photometry on PSF-convolved images at the F444W resolution, with aperture corrections applied. Photometric uncertainties are determined using randomly placed apertures across the images. When available, PRISM spectra are simultaneously fit alongside the photometry. Before fitting, the spectra are corrected for slit losses by normalizing them to match the photometric fluxes, ensuring consistency between spectroscopic and photometric data. For the CEERS and ERO samples, we adopt the stellar masses from the Nakajima et al. (2023) catalog, which were derived using NIRCam photometry. For the DREAMS sample, we perform SED fitting on the publicly available NIRCam photometry. Because DREAMS galaxies are gravitationally lensed, the resulting stellar masses and SFRs are corrected for magnification (see Section 4.1.1 for details). Our fitting procedure, described below, is identical to that of Nakajima et al. (2023), ensuring a consistent analysis.

We perform SED fitting using the Bayesian inference code Prospector (Johnson et al. 2021), following the methodology described in Harikane et al. (2023a) and Nakajima et al. (2023). Model spectra are generated with the Flexible Stellar Population Synthesis (FSPS; Conroy et al. 2009; Conroy & Gunn 2010) package, incorporating MIST isochrones Choi et al. (2016). We assume a Chabrier (2003) initial mass function (IMF), the Calzetti et al. (2000) dust attenuation law, and IGM absorption following Madau (1995). A flexible, nonparametric star formation history is adopted, where the first bin spans 0–10 Myr and the remaining four are evenly spaced in log time from 10 Myr to a lookback time corresponding to $z = 30$, with constant SFR assumed within each bin. In the fitting procedure, we vary the total stellar mass, V-band dust attenuation τ_V, and the star formation history as free parameters, while fixing the stellar metallicity at $Z = 0.2\,Z_\odot$. We assume a continuity prior for the star formation history and flat priors over the following ranges: $0 < \tau_V < 2$ and $6 < \log(M_*/M_\odot) < 12$. The posterior distributions are sampled using Markov Chain Monte Carlo (MCMC) sampling with the EMCEE algorithm (Foreman-Mackey et al. 2013a), based on the minimum χ^2.

We derive SFRs for all galaxies from the dust-corrected Hβ emission-line luminosities, following the same procedure as Nakajima et al. (2023). The dust extinction for each galaxy is individually estimated from the Balmer decrement (Hα/Hβ or Hγ/Hβ). We assume the Gordon et al. (2003) attenuation law and Case B recombination with an electron temperature of $T_e = 17{,}500$ K. The choice of Balmer line is based on its availability and signal-to-noise ratio (S/N). To derive the color excess, $E(B-V)$, we use the line with the higher S/N between Hα (if covered) and Hγ, provided its S/N is greater than 3. For galaxies where neither Hα nor Hγ meets this S/N threshold, we apply a correction using the average $E(B-V)$ derived from the stacked spectrum from the corresponding redshift bin (Figure 4.9 (d)). After correcting the Hβ flux for extinction, we calculate SFRs using the Kennicutt (1998) relation, adjusted for a Chabrier (2003) IMF via the conversion factor from Madau & Dickinson (2014). Figure 4.2 summarizes the distributions of stellar mass and SFR obtained for our sample.

4.2.2 Stacking Analysis

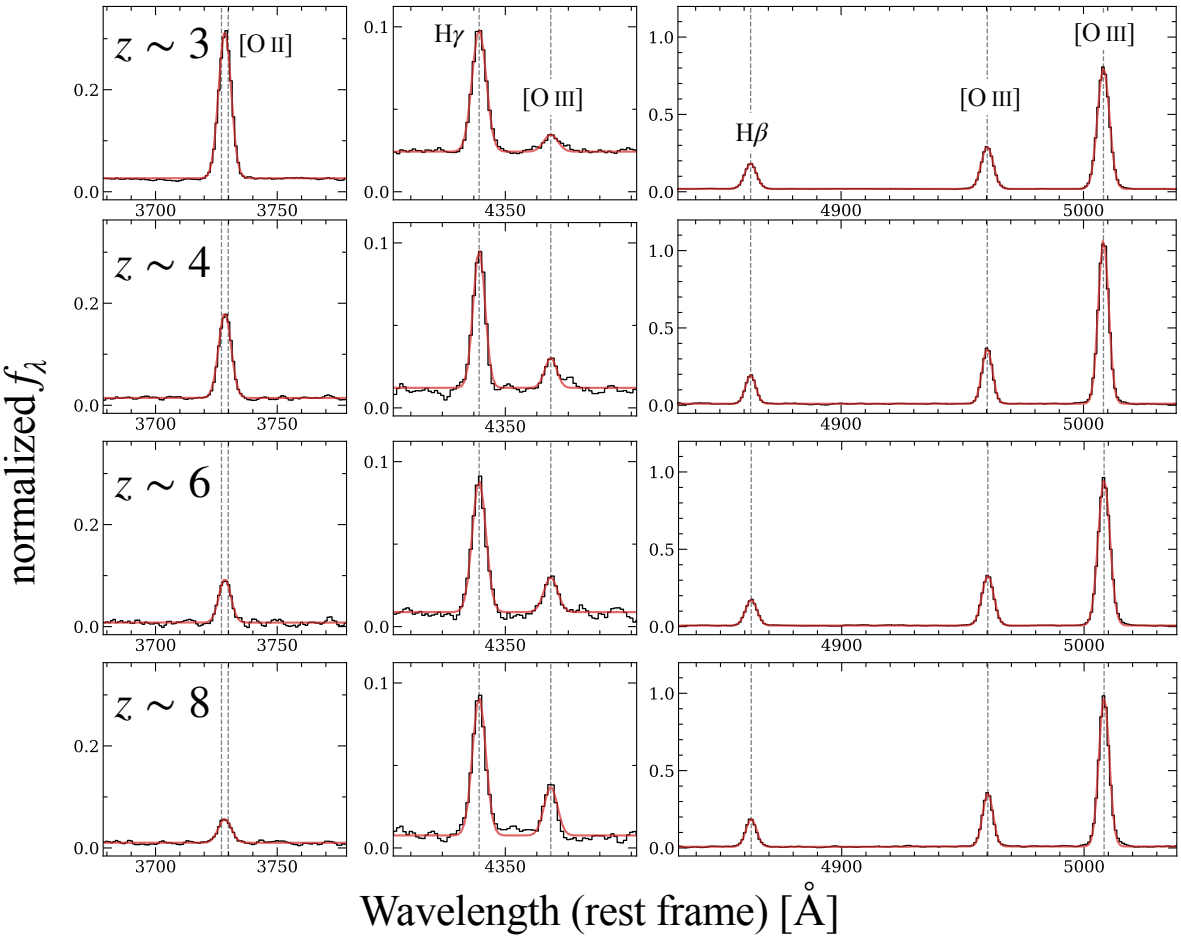

Fig. 4.3 Stacked spectra used in this study, shown for four redshift bins ($z \sim 3$, 4, 6, and 8; from top to bottom) and three wavelength ranges (from left to right): [O II]$\lambda\lambda$3727 doublet; Hγ and [O III]λ4363; and Hβ, [O III]λ4959, and [O III]λ5007. Gray vertical lines mark the expected positions of the relevant emission lines. Black curves show the stacked spectra normalized by the Hβ flux of each galaxy before stacking. Red curves indicate the Gaussian fits used for emission-line flux measurements.

We stack the medium grating spectra of our sample in redshift bins to improve the signal-to-noise ratio. This allows us to derive an average metallicity for each bin, based on the assumption that galaxies within it share broadly similar physical properties, particularly metallicity and emission-line ratios.

We divide the sample into four redshift bins spanning $z = 2$ to $z = 9.5$, containing $N = [122, 60, 49, 18]$ galaxies with median redshifts of $z = [3.1, 4.4, 6.1, 7.9]$, as summarized in Table 4.2. Before stacking, the individual spectra are shifted to the rest-frame and linearly interpolated onto a common wavelength grid with $\Delta\lambda = 1$ Å in linear wavelength space. This grid spacing is smaller than the typical spectral

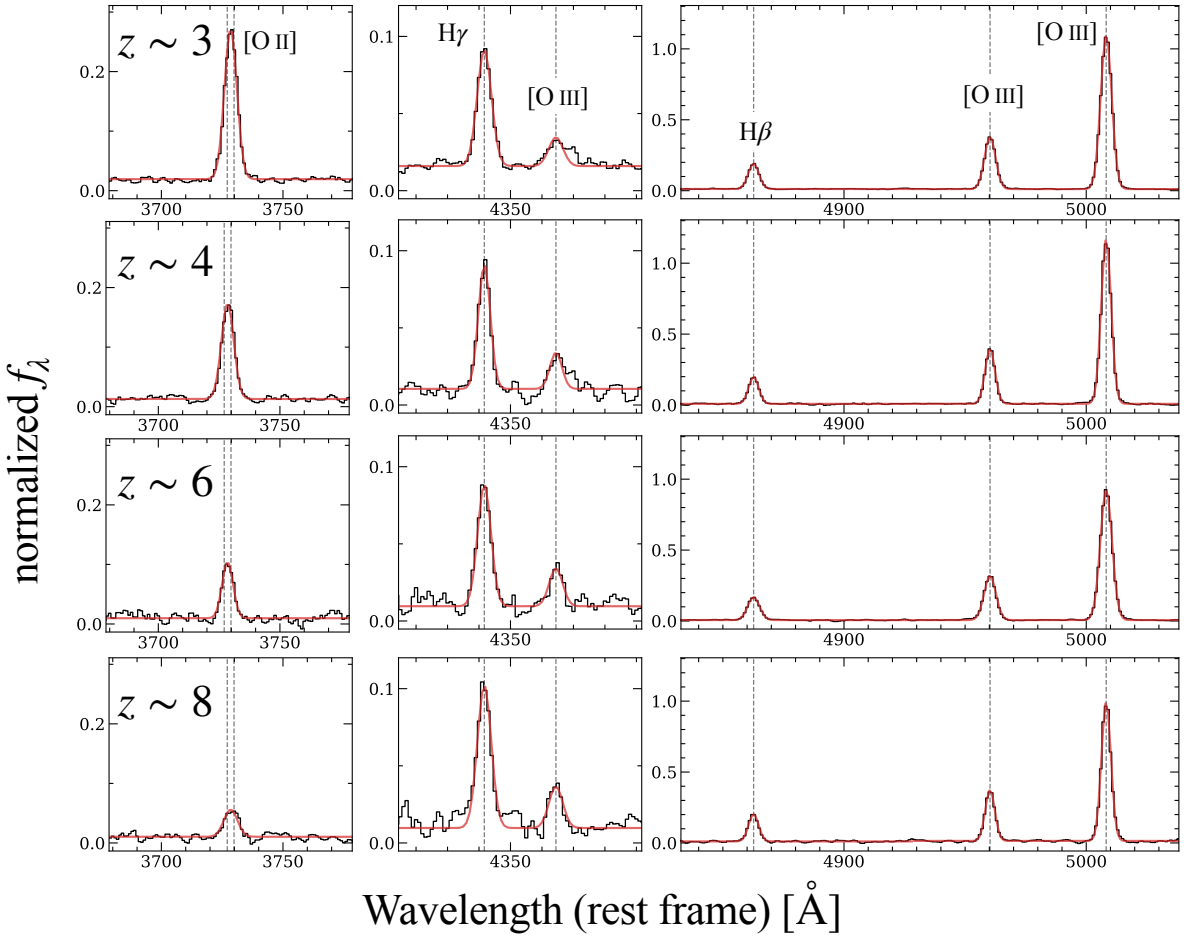

Fig. 4.4 Same as Figure 4.3, but for sample in a fixed M_*-SFR subsample.

resolution of the rest-frame, medium-resolution spectra, meaning the process generally involves interpolation. As the emission lines are well-resolved with widths broader than 1 Å, this interpolation conserves their total flux. We confirmed that using a coarser grid of 2 Å has a negligible impact on our final results. The each spectrum is normalized by its Hβ flux. We then calculate the weighted mean spectrum by applying weights given by the inverse of the S/N of the Hβ line flux for each galaxy. To reduce the impact of outliers, we apply a sigma clipping procedure at each wavelength point, removing data points deviating by more than three standard deviations from the median. When bad pixels coincided with the positions of emission lines, those pixels are masked before stacking.

We measure the emission-line fluxes and ratios of the stacked spectra using the same procedures as for individual spectra as described in section 4.1.5. We correct for dust extinction in the stacked spectra using the Balmer decrement. For the $z \sim 3, 4$, and 6 stacks, we utilize the Hα/Hβ ratio. Note that the Hα emission in these stacks is derived from the subset of galaxies ($> 80\%$ of the sample) where the line falls within detector coverage, with no contribution from sources located in detector gaps or at $z > 6.9$. For the $z \sim 8$ stack, where Hα is redshifted out of the NIRSpec range, we use the Hγ/Hβ ratio. We adopt the Gordon et al. (2003) attenuation law and assume

Case B recombination with an electron temperature of $T_e = 17,500$ K. The resulting stacked spectra are shown in Figure 4.3.

In addition to redshift binning, we also create composite spectra within each redshift bin by constructing the fixed M_*–SFR subsample, defined as rectangular regions in the M_*–SFR plane (Figure 4.2). This subsampling aims to reduce the dependence on the choice of FMR formulation and to mitigate potential biases related to sample selection and galaxy intrinsic properties. The numbers of galaxies in these subsamples are $N = [38, 34, 22, 7]$, with median redshifts of $z = [3.3, 4.5, 6.2, 7.7]$, also listed in Table 4.2. These composite spectra are generated and analyzed following the same procedure as above, and are presented in Figure 4.4.

4.2.3 Uncertainties

Uncertainties on derived properties such as emission-line ratios, electron temperatures, and metallicities are estimated using a bootstrap resampling technique. In each realization, we perturb the individual science spectra according to their associated error spectra, resample the galaxy population within each stacking bin with replacement allowing for duplication, and generate stacked spectra following the procedure described above. Emission-line ratios and metallicities are then remeasured from each perturbed stack. We perform 1000 such realizations, and the uncertainties on each measured quantity are estimated as the half-width of the central 68th percentile of the resulting distribution. This approach accounts for both measurement uncertainties and sample variance in the stacked spectra.

4.2.4 Photoionization Models

To interpret the observed emission-line properties, we employ a grid of photoionization models using the Cloudy code (version c23; Ferland et al. 2017). This approach provides a physically motivated framework to connect the observed line ratios, particularly [O III]λ5007/[O II]λ3727 (O32), [O III]λ5007/Hβ (R3), and [O II]λ3727/Hβ (R2), with the underlying physical conditions of the ionized gas. Specifically, we use the model grid to estimate the ionization parameter that reproduces the observed O32 ratio at a given metallicity and to test the consistency of our full set of observed line diagnostics.

We construct the model grid by varying two key parameters: the gas-phase metallicity and the ionization parameter. The grid spans four discrete metallicities, $Z = [0.1, 0.2, 0.5, 1.0]\, Z_\odot$, and six ionization parameters, ranging from $\log U$ =-3.0 to -0.5 in steps of 0.5 dex. The hydrogen gas density is fixed at $n_\mathrm{e} = 300\,\mathrm{cm}^{-3}$. For the ionizing radiation source, we adopt the stellar continuum from the BPASSv2.3 models for a binary burst population (Eldridge et al. 2017; Stanway & Eldridge 2018). We assume a stellar population with an age of 1 Myr and set the stellar metallicity to match the gas-phase metallicity of each grid point. For the oxygen abundances, we assume a dust depletion factor of 0.6 (Jenkins 1987). The calculations are terminated when the electron fraction falls below 0.01, ensuring that the line-emitting region is fully encompassed.

4.3 Results

4.3.1 Strong Line Ratios

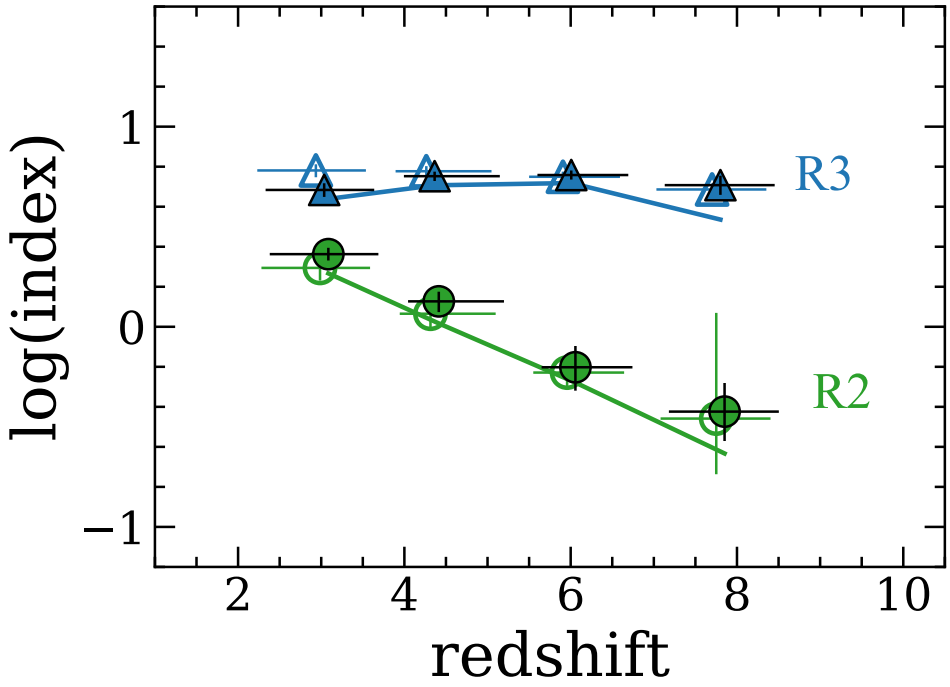

Fig. 4.5 Redshift evolution of strong-line indices measured from stacked spectra. The horizontal axis shows redshift ($z \sim 3$, 4, 6, and 8), and the vertical axis shows the logarithmic values of the line ratios: R3 ([O III]λ5007/Hβ); blue triangles) and R2 ([O II]λ3727/Hβ); green circles). Filled symbols represent measurements from the full sample at each redshift, while open symbols correspond to subsamples matched in stellar mass and SFR. The lines indicate predictions from Cloudy photoionization models, calculated using the observed gas-phase metallicity and ionization parameter ($\log U$) derived for each redshift bin from the full sample.

We measure line fluxes for the stacked spectra, including [O II] λ3727, [O III]λ4363, Hβ, and [O III] $\lambda\lambda$5007,4959. The fitting procedures are same as for individual objects described in Section 4.1.5. We then derive line ratios such as R3, R2, and O32 after dust correction. Figure 4.5 presents the evolution of the strong-line indices R3 and R2 as a function of redshift, derived from stacked spectra at $z \sim 3$, 4, 6, and 8. The R3 index remains nearly constant across this redshift range, while R2 index decreases, indicating that [O II] emission becomes relatively weaker at earlier times compared to both Hβ and [O III]. These trends persist even when comparing the fixed M_*–SFR subsample, suggesting that they are not driven solely by global galaxy properties but rather reflect conditions in the ISM. We discuss the physical interpretation of this trend with a photoionization model in section 4.3.4.

4.3.2 Direct Method Metallicity

All stacked spectra exhibit significant detections of the [O III]λ4363 auroral line (S/N>3), enabling us to determine the gas-phase oxygen abundance using the direct electron temperature (T_e) method. Specifically, we measure the S/N of the line to be 4.9, 5.0, 5.0, and 6.7 in our four redshift bins ($z \sim 3$, 4, 6, and 8, respectively). The line remains significantly detected even in our fixed M_*–SFR subsample, with S/N values of 4.5, 5.0, 3.5, and 5.0.

We first estimate the electron temperature of the O^{2+} emitting region, T_e([O III]), from the [O III] (λ4363/λ5007) line ratio, assuming an electron density of 300 cm^{-3}. This calculation is performed using the getTemDen routine in the PyNeb package (Luridiana et al. 2015). The adopted value of 300 cm^{-3} for the electron density is consistent with that inferred from the [O II] doublet in the stacked spectra at all redshifts (4.3.5). We also confirm that the assumed density has a negligible effect on the derived metallicities, as varying it between 10 and 1000 cm^{-3} changes log(O/H) by less than 0.03 dex. The electron temperature of the O^+ emitting region, T_e([O II]), is estimated from T_e([O III]) using the empirical relation provided by Izotov et al. (2006). Ionic abundances are then computed using the getIonAbundance task in PyNeb: O^+/H^+ from the [O II] $\lambda\lambda$3727 to Hβ ratio and T_e([O II]), and O^{2+}/H^+ from the [O III] $\lambda\lambda$4959,5007 to Hβ ratio and T_e([O III]). We do not include contributions from higher ionization states (e.g., O^{3+}/H^+), following the approximation adopted in Izotov et al. (2006), and we find no evidence for HeIIλ4686 emission in the stacked spectra. The derived T_e([O III]) values and oxygen abundances are summarized in Table 4.2 and displayed in Figure 4.9 (a) and (b), respectively.

4.3.3 Comparing Direct-metallicity Results to Empirical Strong-line Calibrations

Many previous studies of high-redshift galaxies have relied on empirical metallicity indicators based on strong emission lines. To enable comparisons with these earlier works and to evaluate the applicability of strong-line calibrations to individual galaxies in the early universe, metallicities are also derived using strong-line ratios measured from the stacked spectra. Four widely used diagnostics, ([O III]$\lambda\lambda$4959, 5007+[O II]λ3727)/Hβ (R23), O32, R3, and R2, are calculated in each redshift bin. These ratios are then compared to the oxygen abundances obtained from the direct method to assess the validity of existing empirical calibrations.

Figure 4.6 shows the resulting relationships between each strong-line ratio and the metallicity derived via the direct method, along with comparison curves from the literature. Empirical relations from Nakajima et al. (2022), Curti et al. (2020), Sanders et al. (2024), and Hirschmann et al. (2023) are included in the figure. The relations from Nakajima et al. (2022) and Curti et al. (2020) are based on local galaxies, while Sanders et al. (2024) derived their relations using JWST observations of galaxies at z = 2–9. Hirschmann et al. (2023) used a cosmological simulation to establish their calibration. Nakajima et al. (2022) also introduced a method to

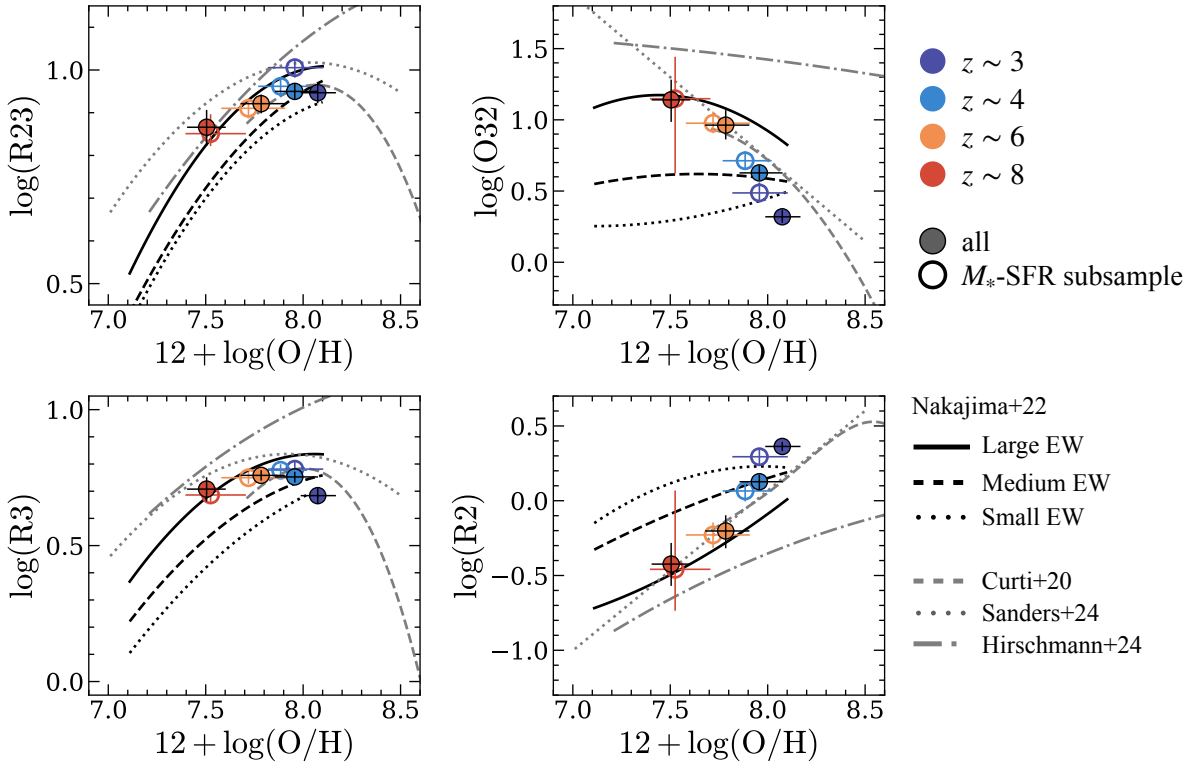

Fig. 4.6 Comparison between strong-line ratios and metallicities derived from the direct method. Each panel shows a different emission-line diagnostic: top-left shows R23 = [O III]$\lambda\lambda$4959, 5007+[O II]λ3727)/Hβ, top-right shows O32 = [O III]λ5007/[O II]λ3727, bottom-left shows R3 = [O III]λ5007/Hβ, and bottom-right shows R2 = [O II]λ3727/Hβ. Colored data points represent results from stacked spectra at different redshifts: $z \sim 3$ (dark blue), $z \sim 4$ (light blue), $z \sim 6$ (orange), and $z \sim 8$ (red). Filled symbols show the full sample at each redshift, while open symbols indicate sub-samples binned by stellar mass and SFR. Lines show metallicity calibrations from the literature. Black solid, dashed, and dotted lines correspond to the empirical relations from Nakajima et al. (2022) for local galaxies with large (EW(Hβ) $>$ 200 Å), medium (100 $<$ EW(Hβ) $<$ 200 Å), and small (EW(Hβ) $<$ 100 Å) Hβ equivalent widths, respectively. The gray dashed line shows the calibration from Curti et al. (2020) based on local galaxies. The gray dotted line represents the relation derived from JWST galaxies at z = 2–9 by Sanders et al. (2024), and the gray dot-dashed line shows the relation predicted by cosmological simulations from Hirschmann et al. (2023).

account for the ionization state of the gas using Hβ equivalent width (EW) as a proxy, motivated by the known correlation between EW(Hβ) and ionization-sensitive ratios such as O32. Their calibrations are provided for different EW(Hβ) bins: EW $<$ 100 Å, 100 Å$<$ EW $<$ 200 Å, and EW $>$ 200 Å.

For each redshift bin in this study, the average EW(Hβ) of the stacked spectra is measured to be 53 Å, 100 Å, 140 Å, and 140 Å at $z \sim$ 3, 4, 6, and 8, respectively. Following the procedure described in Nakajima et al. (2022), the O32 ratio is first used to determine the appropriate metallicity branch (high or low), and the metallic-

ity is then estimated from the R23 ratio using the corresponding relation. At $z \sim 3$, 4, and 6, the strong-line-based metallicities agree well with the results from the direct method. At $z \sim 8$, however, the strong-line calibrations predict systematically higher metallicities than the direct method. This trend is also reported in previous studies ((e.g., Nakajima et al. 2023; Sanders et al. 2024)), which pointed out that the dependence on EW(Hβ) becomes weaker for galaxies at very high redshift. When using the EW $>$ 200 Å calibration, the metallicities and O32 values become more consistent with those derived from the direct method at $z \sim 8$.

Overall, none of the existing calibrations consistently reproduces the observed relationships between strong-line ratios and direct-method metallicities for all four diagnostics across all redshifts. Caution is therefore needed when applying these calibrations to interpret the metallicities of high-redshift galaxies.

4.3.4 Interpreting the Evolution of Strong-line Ratios with Photoionization Models

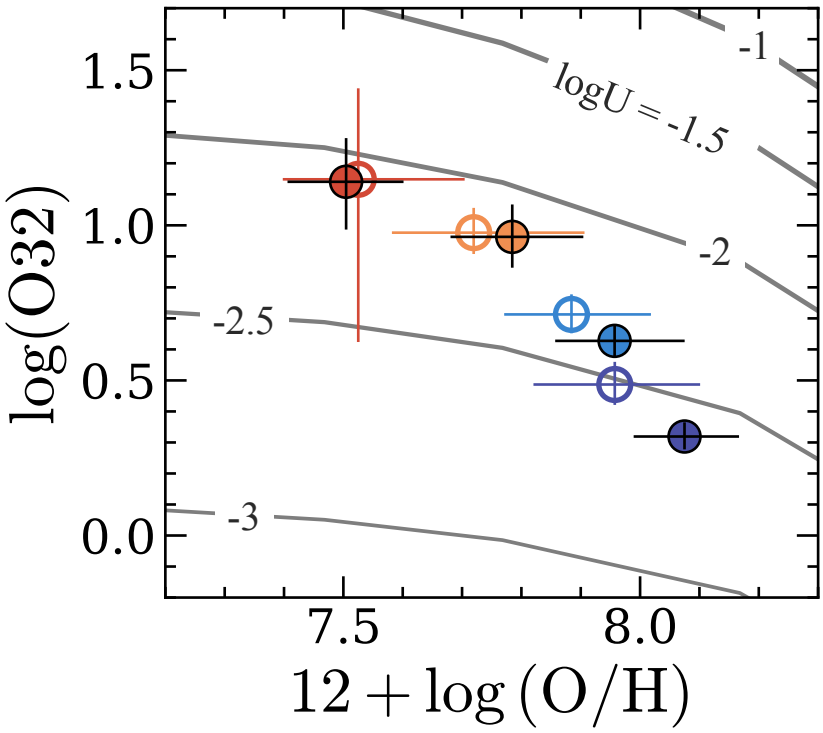

Fig. 4.7 Comparison of observed O32 indices with CLOUDY photoionization model predictions as a function of gas-phase metallicity. The gray curves show model predictions at different ionization parameters log U, ranging from -3 (thinnest) to -1.0 (thickest), computed for metallicities $Z = 0.1$, 0.2, 0.5, and 1 $Z_\odot$. Colored circles indicate observed O32 values at $z \sim 3$, 4, 6, and 8, where the colors correspond to redshift as in Figure 4.6. Filled symbols represent the full sample at each redshift, and open symbols correspond to fixed M_*-SFR subsample. By comparing the observed data points to the model grid, we estimate the ionization parameter log U for each redshift bin.

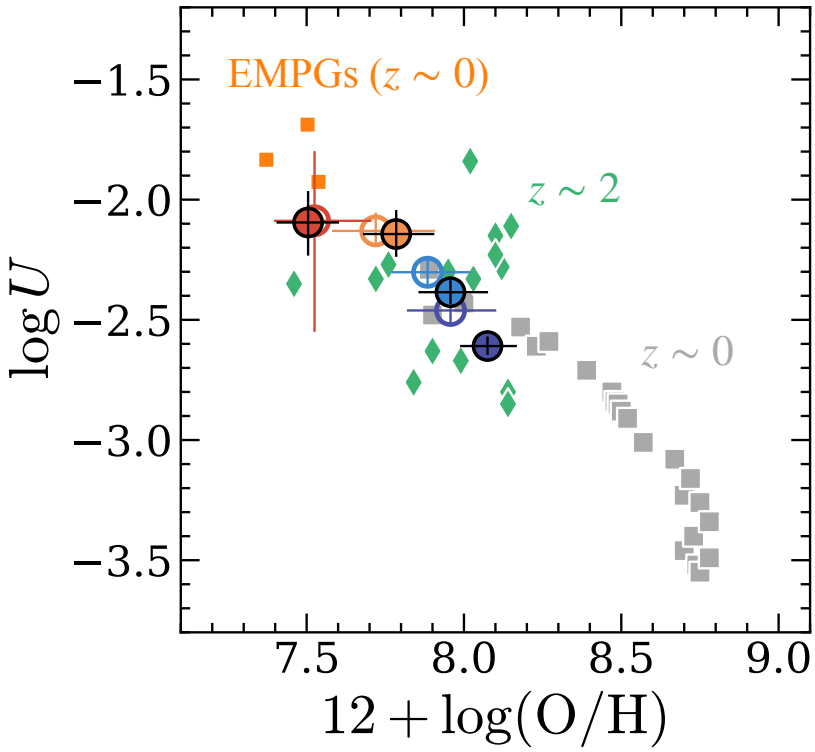

Fig. 4.8 Relationship between ionization parameter and gas-phase metallicity. The horizontal axis shows gas-phase metallicity derived using the direct T_e method, while the vertical axis shows the ionization parameter $\log U$, estimated via photoionization modeling. Colored circles represent stacked galaxy samples at redshifts $z \sim 3$, 4, 6, and 8. Filled circles indicate stacks of the entire sample at each redshift, while open circles show stacks of the fixed M_*–SFR subsample. For comparison, gray squares denote the $z \sim 0$ SDSS results (Andrews & Martini 2013), orange squares represent the three new EMPGs identified in Chapter 2, and green diamonds show the $z \sim 2$ results from the MOSDEF survey (Sanders et al. 2020). Our sample traces the low-metallicity end of the anti-correlation between metallicity and $\log U$ observed at $z \sim 0$ and $z \sim 2$.

As shown in Section 4.3.1, our stacked samples at $z \sim 3$, 4, 6, and 8 show that the R3 index remains roughly constant with redshift, whereas R2 decreases and O32 increases (Figure 4.5). To assess whether these changes can be explained by variations in metallicity and ionization parameters, we employ photoionization models constructed using the Cloudy code, as described in Section 4.2.4.

We first estimate the ionization parameter $\log U$ by computing O32 on a metallicity–ionization parameter grid generated with the photoionization model described in Section 4.2.4 (Figure 4.7). Interpolating this grid with the observed O32 and metallicity values yields ionization parameters for each redshift bin. The resulting values of $\log U$ increase from approximately -2.6 at $z \sim 3$ to -2.1 at $z \sim 8$ (Figure 4.9 (c)).

We present the relation between the ionization parameter and metallicity in Figure 4.8. Previous studies have established an empirical anti-correlation, observed at $z \sim 0$ (Andrews & Martini 2013) and up to $z \sim 2$–3 (Sanders et al. 2020), in which galaxies with lower metallicities tend to have higher ionization parameters. To enable

a consistent comparison, we applied the same estimation method to the three local EMPGs identified in Chapter 2, utilizing their T_e-based metallicities and observed O32 ratios. Our high-redshift sample and the local EMPGs lie at the low-metallicity, high-ionization end of the relation, extending the trend observed at $z \sim 0$–2. This finding indicates that the elevated ionization parameters in high-redshift galaxies are likely a consequence of their low metallicities, rather than evidence for an intrinsic evolution of $\log U$ with redshift.

We next test if our derived physical properties can self-consistently explain the observed trends. Using the estimated metallicities and ionization parameters for each redshift bin as inputs to our photoionization models, we predict the corresponding R3 and R2 values. The model predictions, overplotted in Figure 4.5, show good agreement with the observations. This consistency demonstrates that the observed evolution in strong-line ratios is primarily driven by the joint evolution of gas-phase metallicity and ionization parameter.

The near-constancy of R3 arises from a balance between two competing effects associated with the high-redshift environment: a higher ionization parameter and a lower overall oxygen abundance. The elevated ionization parameter increases the fraction of oxygen in its doubly-ionized state (O^{++}), which acts to boost the [O III] emission. Simultaneously, the lower total oxygen abundance reduces the number of available oxygen ions, which suppresses the emission. These two competing effects largely counteract each other, resulting in a relatively stable R3 across redshift. In contrast, the decline in R2 is driven by the decreasing abundance of O^+ ions. This decrease is caused by both a lower total oxygen abundance and the efficient conversion of O^+ to O^{++} in the increasingly high-ionization conditions at higher redshifts, reaching $\log U \sim -2$ at $z \sim 8$.

4.3.5 Electron Density

We estimate the electron density of our sample from the [O II]λ3727 doublet and examine its evolution with redshift. Although the [O II] doublet is not fully resolved in the medium-resolution data, we fit the blended profile with a double-Gaussian model, fixing the line widths and central wavelengths. The lower-resolution stacks contain more objects and therefore yield more stable measurements, and we adopt these results as our fiducial values. The derived electron densities are shown in Figure 4.9 (d), indicating a mild increase in density with redshift.

Given the large uncertainties in the electron density estimates, we adopt a fixed value of $n_e = 300\,\mathrm{cm}^{-3}$ in the CLOUDY modeling (Section 4.2.4) and in the metallicity measurements (Section 4.3.2). Although the [O II]-based measurements allow for densities up to $n_e \sim 1000\,\mathrm{cm}^{-3}$, assuming this higher value in our analyses does not change the results beyond the measurement uncertainties.

4.3.6 Mass-Metallicity Relation

Figure 4.10 shows the mass–metallicity relation of our sample. Each data point represents a redshift bin ($z \sim 3$, 4, 6, 8, 10), with the stellar mass defined as the median of individual galaxy masses within each bin. For $z \sim 3-8$, metallicities are derived

using the direct method applied to stacked spectra. At $z \sim 10$, where [O III]λ5007 is redshifted out of the NIRSpec coverage and the direct method cannot be applied consistently, we use the median of metallicities from individual galaxies derived using either the direct or strong-line methods (see also Section 4.1.4).

These measurements are broadly consistent with previous studies of the MZR at $z = 4$–10 (Nakajima et al. 2023; Curti et al. 2024). At a fixed stellar mass, we find a clear trend of chemical evolution. The metallicity decreases monotonically from $z \sim 3$ to $z \sim 8$, with a particularly steep decline emerging at $z \gtrsim 8$. While the $z \sim 10$ measurement relies on a heterogeneous mix of methods and exhibits large dispersion, the overall trend points to a rapid chemical enrichment phase in the early universe. This significant evolution of the MZR at fixed stellar mass raises the question of whether this trend is driven solely by the higher star formation rates typical of high-redshift galaxies (i.e., within the framework of the Fundamental Metallicity Relation), which we address in the following sections.

4.3.7 Fundamental Metallicity Relation

To further investigate the metallicity evolution, we examine the fundamental metallicity relation (FMR), which incorporates both stellar mass and star formation rate. The FMR is a potentially more fundamental relation than the MZR alone, as it is known that metallicity depends on SFR at fixed mass. Moreover, FMR-based comparisons can offer insights into galaxy evolution and help mitigate selection biases due to differences in sSFR across redshifts.

In Figure 4.11, we present the difference between the observed metallicities and those predicted by the FMR of Andrews & Martini (2013), $\Delta(\mathrm{O/H}) = (\mathrm{O/H})_{\mathrm{obs}}$–$(\mathrm{O/H})_{\mathrm{pred}}$, as a function of redshift. Here, $(\mathrm{O/H})_{\mathrm{obs}}$ is measured directly from the stacked spectra. To derive $(\mathrm{O/H})_{\mathrm{pred}}$, we calculate the representative stellar mass and SFR for each bin by taking the weighted average of the individual galaxies, using their Hβ signal-to-noise ratios as weights consistent with our stacking procedure. These weighted mean properties are then substituted into the FMR relation. The redshift values and their uncertainties represent the median and 68th percentile ranges of individual galaxy redshifts. The central values and error bars of $\Delta(\mathrm{O/H})$ are derived from the same bootstrap procedure described above. The $z \sim 10$ result is based on individual measurements.

To ensure a fair comparison across different redshifts and to mitigate potential selection effects, we show results from the fixed M_*–SFR subsample (black symbols). By controlling the physical properties of galaxies, we can more clearly trace the intrinsic evolution of the relation. We find that the subsample exhibits a gradual downward trend, with a tentative negative offset appearing as early as $z \sim 4$ ($> 1\sigma$). This deviation becomes more pronounced at higher redshifts, reaching a significant offset at $z \sim 8$.

To assess the robustness and universality of this evolutionary trend, we then compare these results with the full sample (red circles). As shown in Figure 4.11, the offsets derived from the full sample are remarkably consistent with those of the fixed M_*–SFR subsample within the 1σ uncertainties across all redshift bins. While the full sample remains statistically consistent with the local FMR up to $z \sim 6$, it shows

a definitive and significant negative deviation at $z \sim 8$ and $z \sim 10$ ($\Delta(\mathrm{O/H}) < -0.2$ dex). The fact that both the subsample and the full sample yield consistent offsets reinforces the conclusion that the deviation at $z \gtrsim 8$ is a robust feature, independent of the specific sample selection. This indicates a potential FMR break in the early universe, where galaxies are systematically more metal-poor for their M_* and SFR than predicted by the local relation.

For comparison, Figure 4.11 displays the FMR offset derived using metallicities from the R23 strong-line index on our stacked spectra (blue circles). The R23-based metallicities show no significant deviation from the local FMR even at $z \sim 8$, in stark contrast to the direct-method results. This suggests that using local strong-line calibrations at high redshift can lead to an overestimation of metallicity (as discussed in 4.3.3), potentially masking the true evolutionary trends. This reinforces the necessity of using the direct method to accurately probe the baryon cycle in the early universe.

Additionally, we plot measurements for individual galaxies as gray dots. Their metallicities are derived using the direct T_e method if the [O III]λ4363 line is detected with S/N > 3, and from the R23 index otherwise. While these individual measurements exhibit large scatter, their median values in each redshift bin also show a consistent values of $\Delta(\mathrm{O/H})$ with the primary result obtained from our stacked spectra.

We also explore the impact of the FMR definition by performing the same analysis using the FMR of Curti et al. (2020), as shown in Figure 4.12. In contrast to the Andrews & Martini (2013) case, this version predicts systematically higher metallicities, with $\Delta(\mathrm{O/H})$ values of -0.2 dex already apparent at $z \sim 3$, reaching -0.4 dex at $z \sim 8$. The discrepancy likely stems from the differences in the parameter space covered by the local calibration samples. To demonstrate this, Figure 4.13 presents the projection of the FMR onto the $\mu_\alpha \equiv \log M_* - \alpha \log \mathrm{SFR}$ plane, where α quantifies the strength of the SFR dependence. In the left panel using $\alpha = 0.66$ from Andrews & Martini (2013), the vast majority ($\sim$93%) of our sample lies within the parameter space directly constrained by their analysis ($\mu_{0.66} \gtrsim 7.5$). Even for the $z \sim 8$ bin, where the average falls slightly into the unexplored regime, approximately half of the individual galaxies remain within the explored parameter space. Furthermore, the fixed M_*-SFR subsample (open circles) consistently fall within the explored region at all redshifts. Conversely, in the right panel, using $\alpha = 0.55$ from Curti et al. (2020), only 34% of our sample falls within the parameter space directly explored by their analysis ($\mu_{0.55} \gtrsim 8.5$), meaning the relation relies on extrapolation for high-redshift galaxies. These comparisons confirm that the Andrews & Martini (2013) relation is well-suited for our study, as its calibration sample includes the low-mass, actively star-forming galaxies characteristic of the high-redshift universe. Additionally, the strong SFR dependence ($\alpha = 0.66$) inherent to the Andrews & Martini (2013) formulation is consistent with the high specific star formation rates observed at these redshifts; indeed, Curti et al. (2020) report a similarly strong dependence ($\alpha \approx 0.65$) for their high-sSFR population. Therefore, we adopt the Andrews & Martini (2013) FMR as the appropriate baseline for assessing metallicity evolution in this study.

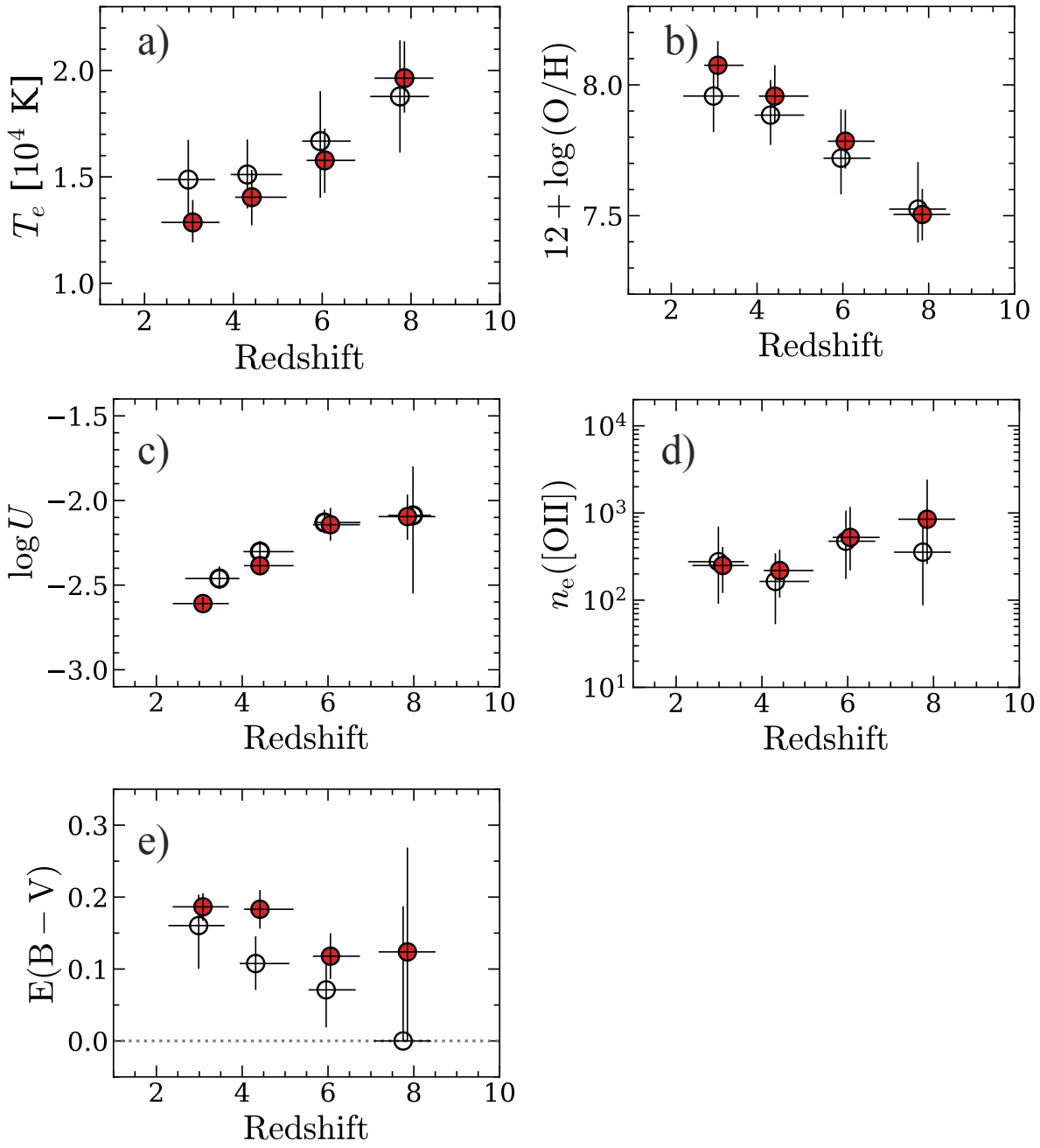

Fig. 4.9 Redshift evolution of key physical properties derived from our sample. The panels show: (a) electron temperature, (b) gas-phase metallicity, (c) ionization parameter, (d) electron density, and (e) nebular dust attenuation $E(B-V)$. All parameters are estimated from stacked spectra in each redshift bin. Filled red circles indicate measurements from the full sample at each redshift, while open black circles correspond to the fixed M_*–SFR subsample.

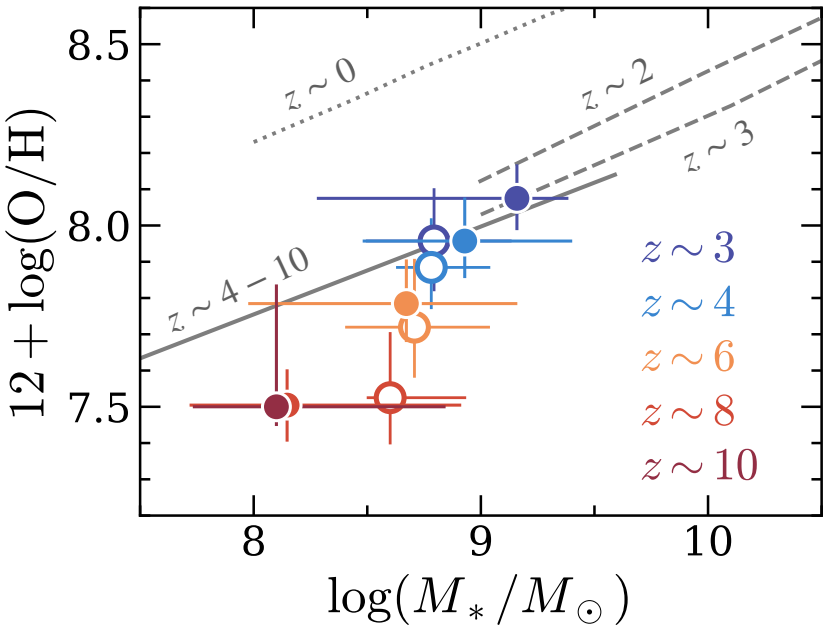

Fig. 4.10 Mass–metallicity relation (MZR). Colored data points represent results from stacked spectra at different redshifts: $z \sim 3$ (dark blue), $z \sim 4$ (light blue), $z \sim 6$ (orange), and $z \sim 8$ (red). Filled symbols show the full sample at each redshift, while open symbols indicate sub-samples binned by stellar mass and SFR. Solid and dashed lines show metallicity calibrations from the literature. For $z \sim 2$–8, metallicities are derived from stacked spectra using the direct method, and the error bars show the 68th percentile range obtained via bootstrap resampling. The stellar masses are estimated individually through SED fitting, and the error bars indicate the 68th percentile distribution of the sample at each redshift. At $z \sim 10$, we compile individual metallicity measurements from the literature and show their distribution. We also include previous MZR results for comparison: the gray dotted line shows the local relation at $z \sim 0$ (Andrews & Martini 2013), the gray dashed lines correspond to $z \sim 2$ and $z \sim 3$ results from Sanders et al. (2020), and the gray solid line represent $z \sim 4$–10 observations from Nakajima et al. (2023). Our results suggest a gradual decrease of metallicity at fixed stellar mass towards higher redshifts.

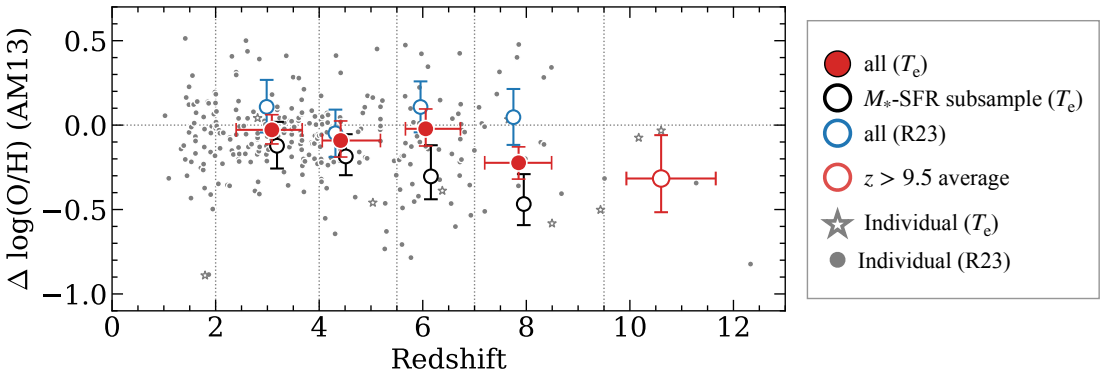

Fig. 4.11 Redshift evolution of the deviation from the Fundamental Metallicity Relation (FMR), defined as the difference between the observed metallicity and that predicted by the Andrews & Martini (2013) FMR given the stellar mass and SFR. Filled red circles indicate our main results at $z \sim 3$, 4, 6, and 8, derived using stacked spectra and the direct method. The open red circle at $z \sim 10$ represents the average of individual measurements from the literature. Black circles show results obtained by stacking galaxies in a fixed M_*-SFR range, while blue circles denote metallicities estimated using the R23 strong-line method. Gray symbols show individual galaxies, metallicities derived from the R23 index (dots) or the direct method where possible (stars). A systematic deviation from the local FMR is observed, which becomes most pronounced at $z \gtrsim 8$.

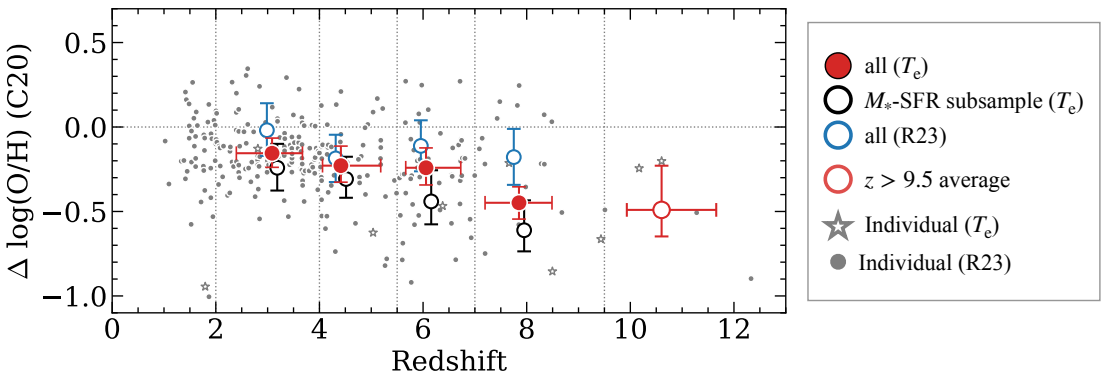

Fig. 4.12 Same as the Figure 4.11, but for the Curti et al. (2020) FMR.

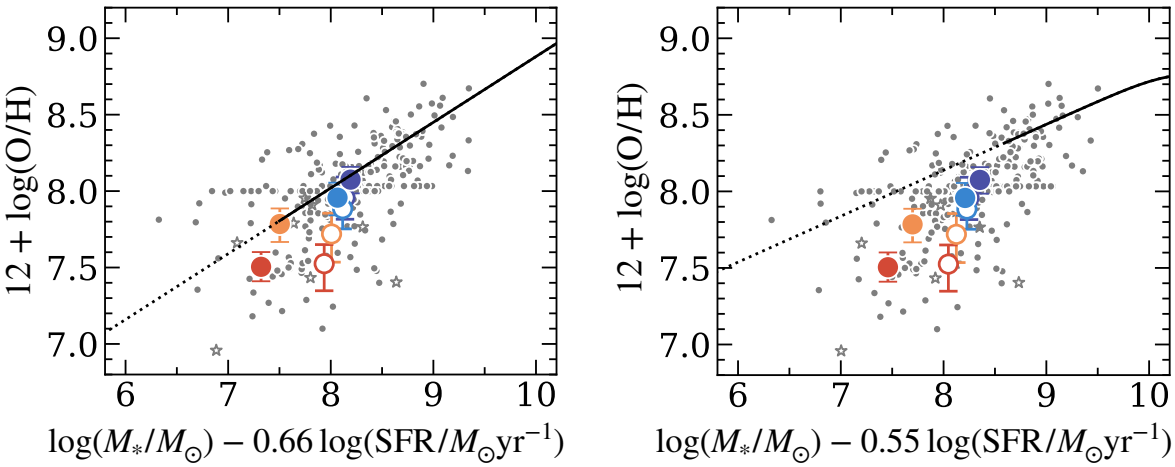

Fig. 4.13 Projection of the FMR onto the μ_α–metallicity plane, where $\mu_\alpha = \log M_* - \alpha \log \mathrm{SFR}$. The colored symbols represent our stacked results at $z \sim 3$–8, with the same legend as Figure 4.6. **Left:** Projection using $\alpha = 0.66$ based on the Andrews & Martini (2013) FMR. The black solid curve indicates the range of $\mu_{0.66}$ effectively explored by the Andrews & Martini (2013) sample, while the dotted curve represents the extrapolation to lower $\mu_{0.66}$. Data points at $z \lesssim 6$ and all fixed M_*–SFR subsample lie within the explored region. At $z \sim 8$, while the average falls slightly into the extrapolated regime, roughly half of the individual galaxies remain within the explored space. **Right:** Projection using $\alpha = 0.55$ based on the Curti et al. (2020) FMR. The solid curve shows the constrained range and the dotted curve shows the extrapolation. In this projection, majority of our samples ($z \sim 3$–8) reside in the unexplored parameter space, requiring extrapolation of the local relation.

Table 4.1 Physical properties of individual galaxies at $z > 7.8$ used in this study.

Object name	Redshift	$\log(M_*/M_\odot)$	SFR $[M_\odot\,\mathrm{yr}^{-1}]$	$12+\log(\mathrm{O/H})$	Ref.
CEERS 01027	7.82	8.11	44	8.00	(1)
DREAMS 31001	7.85	7.73	5.5	7.34	(1)
DREAMS 40005	7.85	6.59	0.4	8.00	(1)
JADES 21842	7.98	8.34	6.7	7.73	(1)
JADES 1899	8.28	8.90	12	8.00	(1)
DREAMS 30002	8.34	7.73	4.6	8.00	(1)
DREAMS 30003	8.34	8.67	5.4	7.98	(1)
JADES 20213084	8.49	7.70	35	7.80	(1)
ERO 04590	8.50	7.60	12	6.96^a	(1)
CEERS 01019	8.68	9.97	126	7.87	(1)
JADES 265801	9.43	8.51	12	7.43^a	(1)
RXJ2129-z95	9.51	7.63	1.7	7.48^b	(2)
MACS0647-JD	10.17	8.10	5.0	7.79^a	(3)
GNz-11	10.60	8.73	24	7.91^a	(4)
GS-z11-1	11.28	7.84	2.1	7.50^c	(5)
GHZ2	12.33	9.03	7.0	7.40^d	(6)

Notes. The listed properties are for two distinct samples of high-redshift galaxies. The galaxies at $z = 7.8 - 9.5$ are part of our $z \sim 8$ stacked sample. The stellar masses, SFRs, and oxygen abundances are derived in this thesis. The galaxies at $z > 9.5$ are not included in our stacking analysis as their [O III]λ5007 line is redshifted out of the NIRSpec wavelength coverage. The properties are compiled from the literature. The oxygen abundances ($12+\log(\mathrm{O/H})$) for individual galaxies in our stack sample are derived using the R23 calibration from Nakajima et al. (2022), except for annotated objects for which we use the direct temperature (T_e) method. For the $z > 9.5$ literature sample, the metallicity derivation methods are adopted from their respective original publications: (*a*) direct temperature (T_e) method, (*b*) R23 and O32 calibrations, (*c*) Ne3O2 calibration, and (*d*) R3 calibration.

References. (1) this thesis; (2) Williams et al. (2023); (3) Hsiao et al. (2024); (4) Álvarez-Márquez et al. (2025); (5) Scholtz et al. (2025); (6) Zavala et al. (2024).

Table 4.2 Galaxy properties derived from our stacked spectra using the entire sample in each redshift bin, and a mean of individual $z > 9.5$ galaxies from the literature.

Sample	Redshift	N	$\log(M_*/M_\odot)$	SFR $[M_\odot\,\mathrm{yr}^{-1}]$	$12+\log(\mathrm{O/H})$	$\log U$	T_e $[10^4\,\mathrm{K}]$
$z\sim3$ stack	2–4 (3.09)	122	9.16	29	$8.07^{+0.08}_{-0.09}$	$-2.61^{+0.03}_{-0.04}$	$1.29^{+0.09}_{-0.10}$
$z\sim4$ stack	4–5.5 (4.41)	60	8.93	20	$7.96^{+0.10}_{-0.12}$	$-2.38^{+0.04}_{-0.04}$	$1.40^{+0.13}_{-0.12}$
$z\sim6$ stack	5.5–7 (6.06)	49	8.67	59	$7.78^{+0.10}_{-0.12}$	$-2.14^{+0.09}_{-0.09}$	$1.58^{+0.15}_{-0.15}$
$z\sim8$ stack	7–9.5 (7.85)	18	8.15	18	$7.50^{+0.10}_{-0.09}$	$-2.10^{+0.13}_{-0.13}$	$1.96^{+0.16}_{-0.17}$
$z>9.5$ mean	10.60	5	7.95	1.62	7.78	–	-

Notes. The Redshift column for our stacked samples indicates the full redshift range and the median redshift in parentheses. The quoted uncertainties represent 16th and 84th percentiles.

Chapter 5

Chemical Evolution of Galaxies: Modeling

This thesis presents a model of metallicity evolution in the multiphase ISM and the CGM of galaxies at $z = 0$–5 and discusses the metal distribution within these regions. We describe our methodology in Section 5.1. Section 5.2 presents the observational data that we use. Our main results are shown in Section 5.3.

5.1 Method

5.1.1 Method Overview

Our model (Fig. 5.1) constrains the evolution of average ISM and CGM metallicities as a function of halo mass and redshift. To understand why we can constrain the metal budget of galaxies empirically, it is helpful to consider an approximation to the full differential equations. The total mass of metals produced in a galaxy can be approximately determined as the product of the stellar mass (M_*) and the yield (y), where the yield here represents the fractional mass of metals injected into the ISM per unit mass of stars formed. Defining the parameter $f_{\rm ISM}$ to be the fraction of metals returned and mixed into the ISM (as opposed to ejected into the CGM or beyond), we can approximately relate the ISM metallicity ($Z_{\rm ISM}$) to the gas mass ($M_{\rm ISM}$) as

$$Z_{\rm ISM} \sim \frac{\langle f_{\rm ISM} \rangle \; y \; M_*}{M_{\rm ISM}}. \tag{5.1}$$

Hence, given measurements of a galaxy's stellar mass, total ISM gas mass, and gas-phase metallicity, we could estimate the average metal distribution fraction over its lifetime, without the need to build an empirical model. If we had such measurements for a range of galaxies, such that we cover a range of halo masses and redshifts, there is then sufficient information to determine how the ISM distribution fraction varies with halo mass and redshift.

In this thesis, we solve the full metallicity evolution equation instead of the approximation above, accounting for mergers and inflowing gas. However, our constraints come from the same basic observational information. In detail, stellar and ISM gas

properties, including stellar and ISM gas masses as functions of halo mass and redshift (i.e., $M_*(M_\mathrm{h}, z)$ and $M_\mathrm{ISM}(M_\mathrm{h}, z)$), are well-constrained by previous models. We utilize the best-fitting UniverseMachine (Behroozi et al. 2019) model for galaxy stellar properties (Section 5.1.4) and the best-fitting NeutralUniverseMachine (Guo et al. 2023) model for ISM gas properties (Section 5.1.5), both calculated on dark matter halos in a dark matter simulation (Section 5.1.3). These models are briefly described in the following three sections. Anchored by these models, we constrain ISM metallicity as a function of halo mass and redshift (i.e., $Z_\mathrm{ISM}(M_\mathrm{h}, z)$) with observational constraints on ISM and CGM metallicity, free parameters for f_ISM, and an assumption for the yield y. In our models, a fraction $1 - f_\mathrm{ISM}$ of the metal mass is ejected into the CGM. In other words, f_ISM quantifies the proportion of metals distributed between the ISM and CGM. Table 5.1 summarizes observational constraints and redshifts of both previous and our models.

5.1.2 Multi-phase ISM Overview

The most common approach to measure ISM metallicity is to derive metallicity from emissions of ionized gas in star-forming regions. However, these measurements may not fully represent the metallicity in the entire ISM, as the gas exists in other phases, including molecular and atomic. It is believed that stars form from molecular gas, and as stars within molecular clouds evolve, they emit ionizing radiation, creating H II regions. Consequently, we assume that both ionized gas and molecular gas coexist in star-forming regions and that metals are well-mixed in these regions. On the other hand, neutral gas is believed to be stored in the surrounding region, as suggested by observations (Walter et al. 2008; Neeleman et al. 2017) and simulations (Bird et al. 2014; Rahmati & Schaye 2014). The mixing of metals within star-forming regions (H II and H_2) with H I regions and the CGM remains not well understood.

In this thesis, we present two models: one adopts a simpler assumption that metals are well-mixed within the entire ISM (Section 5.1.9), and another employs a more realistic approach where metals in H_2 and H I regions are treated separately (Section 5.1.10). For both models, we use observational constraints on CGM metallicity, as well as those on ISM metallicity, to gain insights into the mixing of metals between the ISM and CGM. In the latter model, we include constraints from metallicity measurements in H I regions, offering empirical insight into metal mixing within the ISM.

We refer to the ISM mass as the total mass of molecular and atomic gasses, while the mass of ionized gas is ignored as it is negligibly small compared to the total ISM mass (McKee & Ostriker 1977). The CGM is referred to as the region outside of the ISM and inside of the halo virial radius, and the mass of the CGM is determined by the mass difference between the total baryonic halo mass (estimated as the product of the halo mass and the cosmic baryon fraction) and the stellar+ISM mass. We refer to the IGM as the regions outside of the halo virial radius.

5.1.3 Dark Matter Simulation

The UniverseMachine utilizes the *Bolshoi–Planck* simulation (Klypin et al. 2016; Rodríguez-Puebla et al. 2016) to derive halo properties and assembly histories. Per-

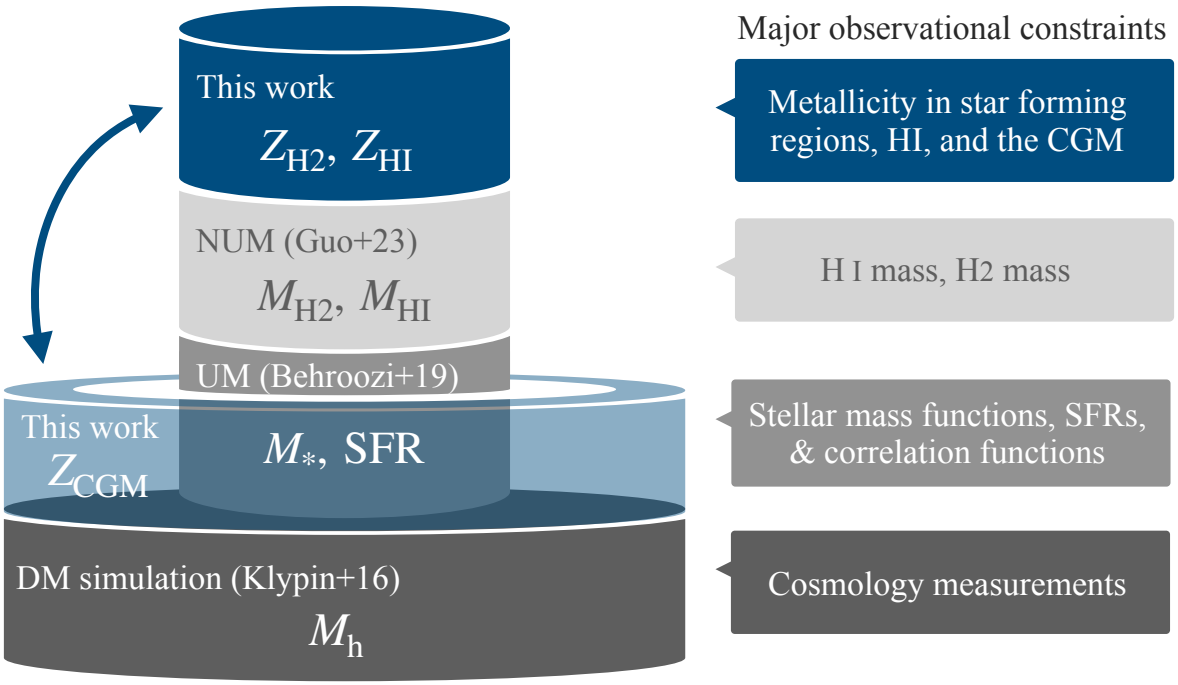

Fig. 5.1 Schematic representation of model hierarchical structure. The metallicities in the ISM (i.e., H I and H_2 regions) and CGM are constrained in this thesis. Dark matter halo properties, such as halo masses and assembly, are provided from the dark matter simulation (Klypin et al. 2016; Rodríguez-Puebla et al. 2016). Stellar properties, including star formation rates (SFRs) and stellar masses, are obtained from the UniverseMachine (Behroozi et al. 2019). For the ISM mass and its time evolution, we use the constraints from the NeutralUniverseMachine (Guo et al. 2023). The major observational constraints for each model are shown on the right side.

formed with the ART code (Kravtsov et al. 1997; Kravtsov & Klypin 1999), this simulation spans a periodic comoving volume of (250 h^{-1} Mpc)^3 containing 2048^3 particles. It achieves high mass and force resolutions of $1.6 \times 10^8 h^{-1} M_\odot$ and 1 h^{-1} kpc, respectively, with time outputs distributed across 180 snapshots equally spaced in $\log(a)$. The underlying cosmology is a flat ΛCDM model consistent with *Planck*15 results ($h = 0.678$, $\Omega_m = 0.307$, $\sigma_8 = 0.823$, $n_s = 0.96$; Planck Collaboration et al. 2016). Halo finding and merger tree construction are carried out using Rockstar (Behroozi et al. 2013b) and Consistent Trees (Behroozi et al. 2013c), respectively.

5.1.4 UniverseMachine

The UniverseMachine (Behroozi et al. 2019) is an empirical framework that establishes the connection between galaxy growth and halo growth with observational constraints spanning a redshift range of $z = 0$–10. This model parameterizes galaxy SFRs as a function of halo potential well depth, redshift, and assembly history for individual halos in the dark matter simulation. The galaxy stellar mass is self-consistently derived from the star formation history along a halo' s assembly and merger history

for each individual halo. The parameter space posterior distribution is determined using a MCMC method by predicting the observable SFRs and stellar mass abundances for a given point in parameter space and then comparing with real observations across a wide range of redshifts. The major observational constraints for the UniverseMachine include stellar mass functions, specific SFRs, and galaxy auto- and cross-correlation functions.

The UniverseMachine output is complied as a catalog,[*1] providing halo and galaxy properties for individual halos at each timestep, including halo mass, SFR, and stellar mass. We extract the average SFR and stellar mass at a given halo mass and redshift. For the definition of halo masses, we use peak historical virial masses extracted from the merger tree ($M_{\rm peak}$).

5.1.5 NeutralUniverseMachine

The NeutralUniverseMachine (Guo et al. 2023) is an empirical model framework that constrains the evolution of atomic hydrogen gas (H I) and molecular hydrogen gas (H_2) masses, developed on the halos in the UniverseMachine catalog. The net gas inflow rates are governed by the evolution of structure in ΛCDM. Figure 5.2 presents the ratios of net gas inflow rates to SFRs as functions of stellar mass and redshift. We calculate the net inflow rates by taking the difference in average gas mass for star-forming galaxies between consecutive timesteps and dividing by the time interval between them (i.e., $\dot{M}_{\rm gas,net} = (M_{\rm gas,in} - M_{\rm gas,out})/\Delta T$).

The H I mass of each halo is parameterized as a function of halo mass, SFR, redshift, and halo formation history. The H_2 mass is parameterized as a function of stellar mass, offset from the star formation main sequence, and redshift. The parameter space posterior distributions are determined by comparing the implied gas masses and their abundances to observations across a wide range of redshifts. The NeutralUniverseMachine model uses observational constraints from various sources, including the H I and H_2 mass functions, the molecular-to-atomic ratio, the H I–halo mass relation, and the H I/H_2–stellar mass relations at $z \sim 0$. This model also incorporates the evolution of cosmic gas densities $\rho_{\rm HI}$ and $\rho_{\rm H_2}$ across the redshift range of $0 < z < 6$. In practice, much of the gas-to-halo mass constraining power of the NeutralUniverseMachine arises because it combines observed gas mass measurements (as functions of stellar mass, SFR, and redshift) with the stellar-to-halo mass relations from the UniverseMachine.

Observations of H I in the local universe come from the ALFALFA survey (Haynes et al. 2018), while measurements of H_2 in the local universe are originate from CO observations in the xCOLD GAS survey (Saintonge et al. 2017). The cosmic density of H I is determined through 21 cm measurements in the local universe and through measurements of damped Lyman-alpha systems (DLAs) at high redshifts. The cosmic density of H_2 is inferred from CO and dust continuum measurements.

The NeutralUniverseMachine catalog provides H I and H_2 masses for every halo in the UniverseMachine catalog. We obtain the average H I and H_2 masses at a given halo mass and redshift from the NeutralUniverseMachine catalog. For

[*1] https://halos.as.arizona.edu/UniverseMachine/DR1

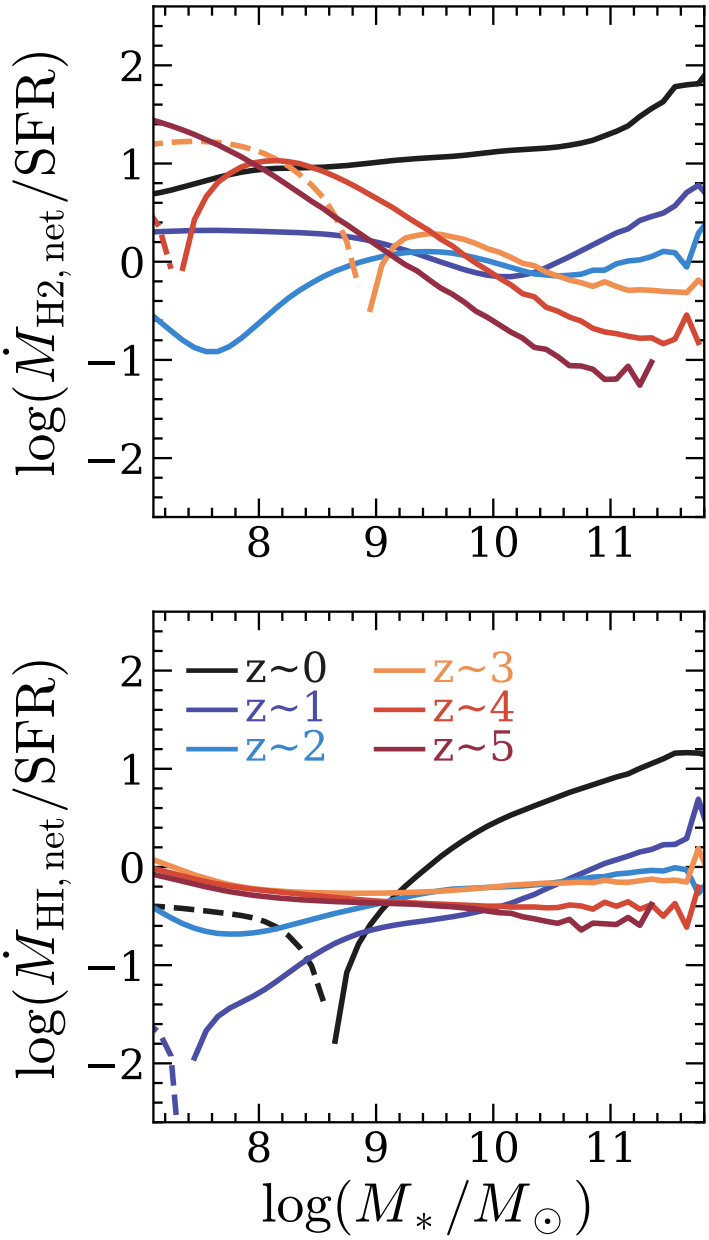

Fig. 5.2 Net gas inflow/outflow rates divided by SFR as functions of stellar mass and redshift, averaged for star-forming galaxies. Solid lines represent cases where inflow exceeds outflow, while dashed lines indicate the opposite.

both H I and H_2 masses, we consider masses of hydrogen without accounting for the contribution of helium and heavier elements, as observed metallicities are expressed relative to the hydrogen content of the gas.

5.1.6 Computing Metallicity Evolution

We constrain the average ISM and CGM metallicities in each halo mass and redshift bin for both star-forming galaxies (SFGs) and quenched galaxies (QGs). The calcu-

lations are based on the average SFRs and ISM masses, which are obtained from the UniverseMachine and NeutralUniverseMachine catalogs, respectively. Galaxies are moved between SFG and QG categories based on their histories in the UniverseMachine. While only the metallicities of star-forming galaxies are compared to observational data, metallicities of quenched galaxies are also tracked to consider populations that become star forming again after being quenched (i.e., those that rejuvenate). We define star-forming galaxies as galaxies with sSFR $> 10^{-11}$ yr^{-1}.

We propagate the average values in each halo mass and redshift bin (as in Behroozi et al. 2013a), rather than calculating for individual halos as done in the UniverseMachine. As well, the constraints for SFR, M_*, $M_{\rm HI}$, and $M_{\rm H_2}$ from the UniverseMachine and NeutralUniverseMachine are fixed, instead of re-fitting them simultaneously with the metallicity calculations. This approach is ideal for our goal of understanding the average evolution of metallicities, as it enables $\sim 10,000\times$ faster model evaluation than computing individual galaxy properties. Behroozi et al. (2013a) discuss that the impact of averaging different SFHs in bins of halo mass is minor compared to observational uncertainties. Since we only predict the average metallicities of galaxy populations using this approach, we use the observational uncertainties for the average metallicities instead of the galaxy-to-galaxy metallicity scatter when comparing to observations (see also Section 5.2). As noted in the introduction, tracking chemical evolution for individual halos will be performed in future papers in this series.

The evolution of the gas-phase metallicity in the ISM arises from several physical processes. Here, we consider the effects of new star formation, metal loss by outflow, halo assembly, and accretion from the CGM. The metal mass evolution in the ISM is expressed as follows:

$$\begin{aligned}\dot{M}_{Z,\rm ISM} = \{y f_{\rm ISM} - Z_{\rm ISM}(1-R)\}{\rm SFR} \\ + Z_{\rm inf}\dot{M}_{\rm inf\rightarrow ISM} + Z_{\rm ISM,merger}\dot{M}_{\rm merger}.\end{aligned} \tag{5.2}$$

In equation 5.2, the metal mass produced by new star formation is determined by the galaxy's SFR and stellar yield y. The stellar yield refers to the fractional mass of metals injected into the ISM per unit mass of stars formed. R denotes the return mass fraction, defined as the proportion of total mass that a stellar generation returns to the ISM. We adopt constant values of $y = 0.016$ and $R = 0.46$ for the stellar oxygen yield with the Chabrier (2003) IMF, following Vincenzo et al. (2016). We assume that the oxygen is promptly ejected into the ISM shortly after the beginning of star formation, because oxygen is mainly produced by short-lived massive stars undergoing core-collapse supernovae and the contributions from type Ia supernovae and AGB stars are negligible (Peeples et al. 2014).

Of the metal (i.e., oxygen) mass ejected by supernovae, a fraction $f_{\rm ISM}$ distributes to the ISM, while the remaining fraction $1-f_{\rm ISM}$ is ejected into the CGM, with $f_{\rm ISM}$ being a free parameter. Star formation also removes some small amount of metals from the ISM due to metals being locked up in stars. The metals that are already present in the ISM from previous enrichment episodes and removed from the ISM by a wind correspond to a negative $f_{\rm ISM}$. However, since $f_{\rm ISM}$ also accounts for metal reinfall and a more sustained distribution, it is expected to be positive. A negative

$f_{\rm ISM}$ would rapidly deplete metals from the ISM, whereas observed galaxy populations show an increasing metal mass over time. We further discuss this in Section 6.3.2. Lastly, $Z_{\rm inf}$ and $\dot{M}_{\rm inf}$ refer to the metallicity and accretion rate of infalling gas, which could come from the IGM, the CGM, or some combination of both.

As galaxies mature, the gas-phase metallicity reaches a limiting state (i.e., saturation). The saturation metallicity $Z_{\rm sat}$ is expressed as $Z_{\rm sat} = y f_{\rm ISM}/(1-R)$, assuming rapid mixing with the ISM during star formation. Once the ISM metallicity reaches the saturation level, it is difficult for it to increase much further, as the metals produced by star formation are exactly balanced by the metals locked up in new stars, and the only new sources of metals are mergers and accretion.

5.1.7 Accounting for Halo Growth and Mergers

We calculate the metal mass accretion through halo assembly using the UniverseMachine catalog, which provides information on individual halos at each timestep (i.e., redshift) of the dark matter simulation, including their mass and progenitors. In each timestep, the average metal mass and ISM mass within each halo mass bin are carried over to the next timestep as halo assembly progresses. Figure 5.3 schematically illustrates how gas and metal masses are propagated to a given $M_{\rm h}$ bin at timestep n. We average the gas and metal masses for the most massive progenitors (MMPs) for all halos in a given $M_{\rm h}$ bin.

Mergers primarily impact massive halos with $M_h \gtrsim 10^{13} M_\odot$; below this halo mass, mergers bring in only small fractions of the total stellar, gas, and metal mass (Behroozi et al. 2019). Merging galaxies may either disrupt as stellar streams (contributing stellar mass to the intrahalo light) or merge with the central galaxy. In the UniverseMachine, merger outcomes are determined by the location where the merging satellite halo falls below the disruption threshold. If this distance R is less than $0.4R_{\rm vir}$ (where $R_{\rm vir}$ is the host halo's virial radius), the satellite merges into the central galaxy; otherwise it merges into the intrahalo light. This threshold is constrained empirically from the clustering of galaxies (which determines the survival time of merging satellites) and the increase in massive galaxy number densities over time (which determines the fraciton of mergers that end up in the central galaxy vs. the intrahalo light). We adopt the same distance threshold for determining the outcome for gas and metal masses during satellite mergers. It is worth noting that not all gas in mergers may cool to the ISM of the host galaxies, even when stars are disrupted. Confining this fraction requires additional observational constraints. Instead, we prepare models with two extreme conditions: one where the ISM of all satellites with $R < 0.4\ R_{\rm vir}$ is integrated into the host galaxy's ISM, and the other where all merging ISM is integrated into the host galaxy's CGM.

5.1.8 Inflows and CGM Metallicity

We also incorporate metals present in inflowing gas from the CGM and IGM, although the metallicities in both gas reservoirs are typically small and thus have little impact on our model. Since observational constraints on the fraction of gas incoming from the IGM vs. the CGM are limited, we consider two extreme models where all inflowing

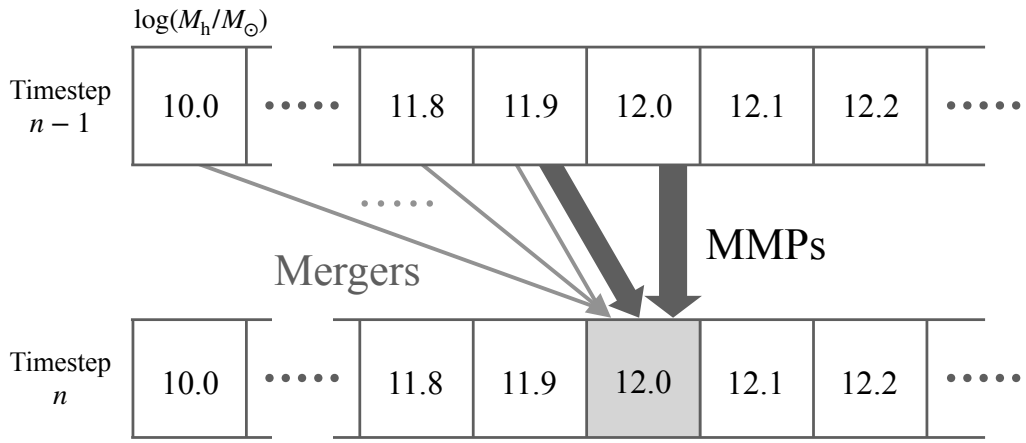

Fig. 5.3 Schematic of gas and metal masses propagation in halo assembly histories. In this thesis, we average halo assembly histories in bins of halo mass, redshift, and star-formation status (star-forming or quiescent) to accelerate model evaluation. Gas and metal masses from most-massive progenitors are propagated to descendant halos at the next redshift, and the corresponding masses for merging satellites are distributed to the descendant halo's CGM or ISM depending on how far the satellites are from their host halos at the last snapshot before disruption (see text).

gas originates from 1) the CGM, or 2) the IGM. To evaluate the amount of metals transported through inflows, we utilize the time differences in ISM mass. If the ISM mass obtained from the NeutralUniverseMachine catalog exceeds that from progenitors, we infer that the net excess gas is supplied by inflows.[*2] We assume that the IGM metallicity is 0.01 $Z_\odot$, as there is evidence that the IGM contains metals (e.g., Simcoe et al. 2004) and that therefore inflowing gas probably also contains metals. We show in A.3 that setting inflow metallicity to $Z_{\rm CGM}$, 0.01, or zero makes little difference to the best-fit values of model parameters or to the quality of the fits.

The metals in the CGM originate from progenitors (i.e., merging satellites) and outflows from galactic star formation. We define the CGM mass as the difference between total baryonic halo mass and the sum of ISM and stellar masses (i.e., $M_{\rm CGM} = f_b M_{\rm h} - M_{\rm ISM} - M_*$, where $f_b = 0.15$ is the cosmic baryon fraction, and in this case we account for helium in the mass of the ISM). We note that galactic outflows may transport both gas mass and metal mass out of the CGM. In this thesis, lacking strong observational constraints, we assume that outflows transport an equal fraction

[*2] If there are additional inflows balanced by outflows from the galaxy, we assume that they do not contribute any extra metals, as it is difficult to remove gas without removing the associated metals as well. The opposite physical situation of additional inflows mixing with ISM gas and being blown out—contributing to more efficient metal removal—is incorporated into the definition of $f_{\rm ISM}$.

of metals and gas mass out of the CGM, so that the CGM metallicity is unchanged by beyond-CGM outflows.

Metal mass evolution in the CGM is expressed as:

$$\begin{aligned}\dot{M}_{Z,\mathrm{CGM}} &= (1 - f_{\mathrm{ISM}}) y \mathrm{SFR} \\ &\quad - Z_{\mathrm{CGM}} \dot{M}_{\mathrm{HI}\rightarrow\mathrm{CGM}} + Z_{\mathrm{CGM,merger}} \dot{M}_{\mathrm{merger}}. \end{aligned} \tag{5.3}$$

The metal mass in the CGM is propagated to the next timestep. This includes CGM metals of all progenitors, as well as ISM metals of progenitors that are disrupted outside of the host galaxy's ISM (i.e., at $R > 0.4\ R_{\mathrm{vir}}$). Additionally, if the ISM mass obtained from the NeutralUniverseMachine catalog is decreasing with time in a given halo mass bin beyond the rate of star formation, the remaining excess balance of gas is assumed to be transported to the CGM, along with a metal mass corresponding to the ISM metallicity. Lastly, in one of our considered models, metals may also be lost to the CGM due to excess inflow into the ISM from mergers.

We found that rejuvenating galaxies require careful treatment due to our method of averaging galaxy properties. This is because, in the UniverseMachine, galaxies can undergo a slow quenching process where they oscillate between being star-forming and quenched for multiple Gyr before they quench completely. Since quenched galaxies have significantly lower average ISM masses than star-forming galaxies, the simplest mass binning scheme, where the inherently broad distribution of SFRs is condensed into a single average SFR (similar to a delta distribution), would require substantial fresh low-metallicity accretion from the IGM/CGM every time a quenched galaxy returned (however briefly) to forming stars, resulting in unphysically low metallicities for massive star-forming galaxies. To address this issue, we assume that quenching galaxies lose gas in the central star-forming regions (shutting off star formation) but that the ejected gas and metals remain nearby. The gas and metals then return to the central region when the galaxies temporarily resume star formation.

5.1.9 Well-Mixed ISM Model

In this thesis, we test two models for the distribution of metals in the ISM. In the first model, which we call the "Well-Mixed ISM Model," we assume that metals in the ISM, including ionized, molecular, and atomic hydrogen, are well-mixed. We calculate the ISM mass as the sum of the molecular and atomic hydrogen masses in the NeutralUniverseMachine catalog.

In this model, the only free variable is the ISM distribution fraction f_{ISM}, which we parameterize as a function of halo mass and redshift ($f_{\mathrm{ISM}}(M_{\mathrm{h}}, z)$). We have found that a double power-law function effectively reproduces the MZR at $z \sim 0$. Therefore, we adopt the following parameterization:

$$f_{\mathrm{ISM}} = \frac{1}{1 + (M_{\mathrm{h}}/C_0)^{\alpha}(1+z)^{\beta}}. \tag{5.4}$$

Here, α, C_0, and β are the three free parameters representing the low-mass slope, characteristic halo mass, and redshift evolution, respectively. The high-mass slope is set to zero to reproduce the metallicity saturation observed at $z \sim 0$, and the

normalization is set to one, as the maximum distribution fraction is 100%. We use a simple parameterization for redshift scaling that effectively shifts the characteristic halo mass as a power law with redshift, as the mass coverage of observed MZRs at high redshifts is not wide enough to constrain both the low- and high-mass slopes. Of note, only the observational constraints for metallicity measurements in the star-forming regions and the CGM are used to constrain the well-mixed ISM model, while observed metallicities in the H I regions are not used.

5.1.10 Multi-phase Metallicity Model

We also explore a more realistic and flexible approach where we separate the metallicity in molecular and atomic gas. Star formation occurs in the molecular gas, while atomic gas surrounds the star-forming regions. In this model, we define two ISM distribution fractions, $f_{\rm H2}$ and $f_{\rm HI}$, representing the fractions of metals that are distributed to the molecular and atomic regions, respectively. We track the evolution of metal mass in these two regions and derive metallicities by dividing by the respective gas masses. Gas accretes onto the H I regions from the CGM or IGM before accreting onto the H_2 regions from the H I regions.

Metallicity evolution in the two regions is expressed as follows:

$$\begin{aligned}\dot{M}_{Z,\rm H2} = &\{y f_{\rm H2} - Z_{\rm H2}(1-R)\}{\rm SFR} \\ &+ Z_{\rm HI}\dot{M}_{\rm HI\rightarrow H2} + Z_{\rm H2,merger}\dot{M}_{\rm H2,merger},\end{aligned} \tag{5.5}$$

and

$$\begin{aligned}\dot{M}_{Z,\rm HI} = &\, y f_{\rm HI} - Z_{\rm HI}\dot{M}_{\rm HI\rightarrow H2} \\ &+ Z_{\rm inf}\dot{M}_{\rm inf\rightarrow HI} + Z_{\rm HI,merger}\dot{M}_{\rm HI,merger}.\end{aligned} \tag{5.6}$$

We parameterize $f_{\rm H2}$ and $f_{\rm HI}$ as functions of halo mass and redshift:

$$f_{\rm H2} = \frac{1}{1+(M_{\rm h}/C_0)^{\alpha_0}(1+z)^{\beta_0}}, \tag{5.7}$$

$$f_{\rm HI} = \frac{1-f_{\rm H2}}{1+(M_{\rm h}/C_1)^{\alpha_1}(1+z)^{\beta_1}}, \tag{5.8}$$

We adopt the same functional form for $f_{\rm H2}$ as for $f_{\rm ISM}$ in the well-mixed ISM model, as they are fitted to the same observational data. For $f_{\rm HI}$, we use a similar functional form to $f_{\rm H2}$ because the observational data do not provide enough information to justify a different shape. The term $(1 - f_{\rm H2})$ ensures that $f_{\rm HI}$ does not exceed unity. This results in six free parameters: α_0, α_1, β_0, β_1, C_0, and C_1. The observational constraints in the star-forming regions, the CGM, and the H I regions are used in the multi-phase metallicity model.

5.1.11 Model Fitting and Posteriors

Any choice of parameters for $f_{\rm ISM}$ determines the evolution of the average ISM and CGM metallicities ($\langle Z_{\rm ISM}(M_{\rm h}, z)\rangle$ and $\langle Z_{\rm CGM}(M_{\rm h}, z)\rangle$) for the well-mixed ISM

model. Similarly, any parameter choice for $f_{\rm H2}$ and $f_{\rm HI}$, determine $\langle Z_{\rm H2}(M_{\rm h}, z)\rangle$, $\langle Z_{\rm HI}(M_{\rm h}, z)\rangle$, and $\langle Z_{\rm CGM}(M_{\rm h}, z)\rangle$, for the multi-phase metallicity model. To compare with observational data, we convert these functions to functions of stellar mass, using the halo mass function and the observed stellar mass probability distribution function (i.e., the probability of having a stellar mass at a given halo mass), both obtained from the UniverseMachine catalog. For H I metallicity, since observations use velocity width as a proxy for mass, we convert the H I metallicity (as a function of halo mass) to a function of velocity width. We assume a log-normal distribution with 0.2 dex scatter for velocity width at a given halo maximum circular velocity, based on the results of Bird et al. (2015) who have investigated the relation between DLA's velocity width and halo virial velocity using hydrodynamical simulations. We compare the predicted metallicities after conversion ($\langle Z_{\rm ISM}(M_*, z)\rangle$ and $\langle Z_{\rm CGM}(M_*, z)\rangle$ for the well-mixed ISM model, or $\langle Z_{\rm H2}(M_*, z)\rangle$, $\langle Z_{\rm HI}(\Delta V, z)\rangle$, and $\langle Z_{\rm CGM}(M_*, z)\rangle$, for the multi-phase metallicity model) with observational data to calculate the likelihood (i.e., $\exp(-0.5\chi^2)$) that the chosen parameter set matches the observed data. We use a Markov Chain Monte Carlo (MCMC) algorithm (Foreman-Mackey et al. 2013b) to explore the parameter posterior distribution that is consistent with the observations.

5.2 Observational Data

5.2.1 Metallicity Measurements in Star-forming Regions

For constraints on metallicity in the H_2 regions, we use the mass-metallicity relations (MZRs) from emission lines of galaxies at $z \sim 0$–5. These metallicities are measured using nebular emission lines that are emitted from star-forming (i.e., H II) regions. Given that the H II and H_2 regions are physically close to each other, we assume that the metallicity in the H_2 regions is the same as that in the H II regions.

Local Universe

Various studies have investigated the MZR of galaxies in the local Universe, employing different methods, including the direct $T_{\rm e}$ method and strong-line calibrations. The direct method, based on electron temperature measurements, is widely regarded as more reliable. However, its reliance on detecting faint auroral lines such as [O III]λ4363 may introduce a bias towards low-metallicity (and thus low-mass) galaxies within MZR samples. In contrast, strong-line calibrations can be applied for a broader range of galaxies, thus reducing sample bias. Despite the larger systematics in individual estimations compared to the direct method, it is crucial to rely on observational data reflecting the true average for constraining the average metallicity of galaxies. Therefore, employing the MZR with strong-line calibrations is more suitable for our purpose.

We use the MZR of Curti et al. (2020), which determined the metallicities for individual SDSS galaxies at $z \sim 0$ and reported the metallicity distribution (i.e., median and scatter) in bins of stellar mass. The metallicities are determined using the strong-line calibrations presented in Curti et al. (2017). The total number of galaxies in this sample is 151862. The slope and saturation metallicity of the Curti et al.

(2020) MZR are consistent with other MZR determinations using the T_e method (e.g., Andrews & Martini 2013), while the normalization is significantly lower compared to the results from photoionization models (e.g., Tremonti et al. 2004).

The observational samples still have some biases towards high sSFR galaxies in the low-mass regime, due to a minimum SFR selection in Curti et al. (2020). To account for the biases, we correct the Curti et al. (2020) data to estimate the true average metallicities of star-forming galaxies in the local universe. We can do this by assuming that unobserved galaxies lie on the mass-metallicity-SFR relation (i.e., FMR), in which galaxy metallicity is only a function of the parameter $\mu \equiv \log M_* - 0.66 \log \mathrm{SFR}$. Galaxies with SFRs below the cut in Curti et al. (2020) will have higher μ than galaxies above the cut, so no extrapolation of the shape of $Z(\mu)$ is needed.

In practice, we calculate the average and scatter of sSFR using the UniverseMachine catalog at $z \sim 0$ in each stellar mass bin. We assume a log-normal distribution for sSFRs. As above, we also estimate the metallicity distribution based on the sSFR distribution assuming the mass-metallicity-SFR relation defined by Curti et al. (2020) (for total SFR). By integrating the sSFR and metallicity distributions in the range of sSFR values that are not covered by Curti et al. (2020), we determine the fraction and average metallicity values for galaxies below the SFR cut in Curti et al. (2020). Finally, to derive the average metallicity of all star-forming galaxies in bins of stellar mass, we perform a weighted summation using the fraction above, by combining the estimated metallicity values for unobserved galaxies with the metallicity values for observed galaxies.

We estimate the uncertainties of the average metallicity, as Curti et al. (2020) do not provide this information, but rather scatter of the metallicity distribution. The typical systematic uncertainty of calibrations reported in Curti et al. (2020) is 0.12 dex, while the typical measurement error of SDSS galaxies is 0.02 dex (Andrews & Martini 2013). The mean sum of squares of these two is 0.12 dex. Hence, we adopt an uncertainty of 0.12 dex for every stellar mass bin.

High Redshifts

We focus our modeling of metallicity evolution within the redshift range of $0 < z < 5$ due to limited gas measurements beyond this range. We utilize the MZR reported at $z \sim 2.3$ and $z \sim 3.3$ from Sanders et al. (2021), as well as at $z \sim 5$ from Nakajima et al. (2023). We apply no correction for biases in the high-redshift samples, as the galaxies show consistency with the main sequence galaxies at their corresponding redshifts.

Sanders et al. (2021) have determined metallicities for stacked spectra of galaxies from the MOSDEF survey (Kriek et al. 2015), which were binned by redshift and stellar mass. The total number of galaxies in the sample is 265 at $z \sim 2.3$ and 130 at $z \sim 3.3$. To determine metallicity, Sanders et al. (2021) employ strong-line metallicity calibrations derived in Bian et al. (2018). We adopt the root mean square of the measurement uncertainty and the calibration scatter as the uncertainty of the metallicity measurements. The typical uncertainty is 0.11 dex.

Nakajima et al. (2023) have derived metallicities for 135 galaxies at redshifts $z = 4$–10 using JWST/NIRSpec early release data. The direct method is employed to galaxies with [O III]λ4363 detection, while strong-line metallicity calibrations of Nakajima et al. (2022) are applied for galaxies without [O III]λ4363 detection. We also include

metallicity measurements for 62 low-mass galaxies from JWST Advanced Deep Extragalactic Survey (JADES; Eisenstein et al. 2023) reported by Curti et al. (2024). The determination of gas-phase metallicity utilizes a modified version of the Curti et al. (2020) calibrations, incorporating new diagnostic methods suggested by Laseter et al. (2024). We use 104 galaxies in total within the redshift range of $z = 4$–6 (median at $z \sim 5$), 62 from Nakajima et al. (2023) and 34 from Curti et al. (2024). We adopt the median and its uncertainties within three stellar mass bins for the individual metallicity measurements. These uncertainties combine the measurement uncertainties and calibration uncertainties. The typical uncertainty is 0.23 dex.

5.2.2 Metallicity Measurements in Atomic Gas

DLAs are considered to trace dense H I gas within galaxies. We rely on metallicity measurements of DLAs across a range of redshifts ($z = 0$–5) compiled in Møller et al. (2013). These measurements are derived from the analysis of metal absorption lines, including those of zinc (Zn), sulfur (S), silicon (Si), and iron (Fe), and are reported as [X/H]. An offset of 0.3 dex has been applied to [Fe/H] measurements to correct for α-enhancement (Rafelski et al. 2012). We compare the oxygen abundance predicted by the model with these observational data, assuming solar abundance of Asplund et al. (2009).

The velocity widths of low-ionization lines are used as a proxy for the halo mass. Bird et al. (2015) have demonstrated using cosmological hydrodynamic simulations that the relation between the velocity width of DLAs and their virial velocity is order of unity, despite large scatter. Based on their results, we convert virial velocities into observed velocity widths assuming a log-normal distribution with a scatter of 0.2 dex and no offset (see also Section 5.1.11).

We divide the individual DLA measurements into bins of redshift ($z \sim 1$, 2, 3, and 4) and velocity widths. Then, we estimate the average metallicities and their uncertainties in each bin, using the bootstrap estimation method. More specifically, we derive the average metal number density $\langle n_{\rm X} \rangle$ and hydrogen density $\langle n_{\rm H} \rangle$, which are then combined to yield the average metallicity $\langle {\rm X/H} \rangle = \langle n_{\rm X} \rangle / \langle n_{\rm H} \rangle$.

It should be noticed that these data do not account for the effects of dust depletion. Of the DLA measurements used in this study, ∼70% (76 out of 110) rely on Zn and S, which are considered non-refractory elements with minimal dust depletion. The remaining measurements involve Si and Fe, which exhibit significant depletion, especially at high metallicity. De Cia et al. (2018) provide the amount of dust corrections for some DLAs used in this thesis. We confirm that the average metallicities in each velocity width and redshift bin, before and after the dust correction, agree within the uncertainties, except for the highest velocity and redshift bin. Since the sample size for dust depletion correction is small in the highest redshift bin (7), we use the uncorrected values reported in Møller et al. (2013).

5.2.3 Metallicity Measurements in the CGM

We utilize metallicity measurements in the CGM at $z \sim 0.2$, obtained from quasar spectra in the COS-Halos survey (Tumlinson et al. 2011, 2013), as reported in

Prochaska et al. (2017). We divide the sample into three stellar mass (from Werk et al. 2012) bins and apply a similar methodology to that used for DLAs (as described in Section 5.2.2) to estimate the average metallicity and its uncertainty within each bin.

5.3 Results

5.3.1 Fitting Results

Our model is combined with four models, each making different assumptions about inflow sources (either IGM or CGM; see Section 5.1.8) and mergers (either 100% of mergers bringing metals to the host galaxy's ISM or no contribution from mergers; see Section 5.1.7), for both the well-mixed ISM model and the multi-phase metallicity model. For each model, we performed an MCMC sampling with 32 walkers and 1000 steps per walker, exceeding 10 times the autocorrelation time to effectively explore the posterior distribution. For reviews of the dependencies of the results on the assumptions about inflow sources and mergers, refer to A.3. Subsequently, we resampled the ISM distribution fractions and metallicity evolution 1000 times by randomly selecting models from the four posterior distributions without weighting. The best-fitting models are determined based on the median values obtained from the resampled models. The best-fitting parameters and their 68th percentiles of the posterior distributions for the well-mixed ISM model are $[\log C,\ \alpha,\ \beta] = [14.5^{+0.3}_{-4.1},\ -0.11^{+0.22}_{-0.03},\ -1.8^{+0.6}_{-0.6}]$, and for the multi-phase metallicity model are $[\log C_0,\ \alpha_0,\ \beta_0,\ \log C_1,\ \alpha_1,\ \beta_1] = [13.3^{+0.4}_{-0.3},\ -0.26^{+0.05}_{-0.05},\ 0.5^{+0.1}_{-0.2},\ 13.2^{+0.9}_{-0.8},\ -0.6^{+0.2}_{-0.2},\ -1.6^{+0.8}_{-0.9}]$, respectively.

We show the mass-metallicity relations of the well-mixed ISM model in left panel of Figure 5.4. The observational data from star-forming regions (top panel) and the CGM (bottom panel) are incorporated as observational constraints, while data from H I regions (middle panel) are not included in the fitting procedure. The naive reduced χ^2 of the best-fitting model is 4.6 for the number of observational data points including the H I metallicities (54). The largest discrepancy occurs at $z \sim 5$ in star-forming regions, where the model shows higher metallicity compared to the observed metallicity at higher-mass points, and even surpasses the predicted metallicity of the model at $z \sim 3$ for the given mass. The elevated metallicities at $z \sim 5$ probably arise from the slight decrease of ISM (specifically H I) masses from $z \sim 4$ to $z \sim 5$ in the NeutralUniverseMachine model. Note that this decreasing trend is unclear in the observed H I cosmic density, due to large uncertainties and scatter in the observational data (see Figure 4 of Guo et al. 2023).

In addition, this model fails to reproduce the observed MZR in H I regions at $z \lesssim 3$ (noting that the observational data are not used in the model fitting process), as indicated by the left panel of Figure 5.4. Since the observed metallicity in H I regions is lower than that in star-forming regions at a given mass, the model predicts higher metallicity in H I regions. These discrepancies imply that the assumption of metals being homogeneously mixed within the ISM is too simplistic. This is not just an artifact of our modeling, but can be readily validated from the data directly. For

example, the ISM metallicity of a $10^9 M_\odot$ galaxy at $z = 2$ is equivalent to $[X/H] \sim -0.5$, but this galaxy mass corresponds to a $\sim 10^{11} M_\odot$ host halo with $\Delta V \sim 100$ km/s, for which typical DLA metallicities are $[X/H] \sim -1.5$.

We show the mass-metallicity relations of the multi-phase metallicity model in right panel of Figure 5.4. This model shows substantially better agreement with the MZRs for all gas phases, including star-forming regions, H I regions, and the CGM. For the number of observational data points we use (54), the naive reduced χ^2 of the best-fitting model is 0.60, suggesting a reasonable fit.

5.3.2 Distribution of Metals between the ISM and CGM

In Figure 5.5, we show the fraction of metals distributed to the ISM for the multi-phase metallicity model. The top left panel illustrates the total metal distribution fractions to all phases of the ISM, i.e., $f_{\rm H2} + f_{\rm HI}$. The top right panel shows the distribution fraction to the CGM, $1 - (f_{\rm H2} + f_{\rm HI})$. The bottom left panel displays the metal distribution fraction to H_2, i.e., $f_{\rm H2}$. The bottom right panel shows the fraction of metals distributed to H I compared to those not distributed to the H_2 regions, i.e., $f_{\rm HI}/(1 - f_{\rm H2})$. In these panels, the mass range covered by the observational data are shown in thick lines. The total ISM distribution fraction ($f_{\rm H2} + f_{\rm HI}$) gradually increases with halo mass from $\sim 20\%$ to $\sim 80\%$ ($\sim 30\%$ to $\sim 100\%$) in $10^{10} - 10^{13} M_\odot$ halos at $z = 0$ ($z = 5$).

At $z \sim 0$, the H I distribution fractions remain low ($f_{\rm HI} \lesssim 0.2$) across mass, indicating that a higher fraction of metals are distributed to H_2 rather than H I. This suggests poor metal mixing between the H_2 and H I phases, highlighting the importance of modeling the multi-phase gas contents of molecular and atomic gas separately.

The total distribution fractions exhibit a gradual increase towards high redshift, with the combination of a slight decrease of $f_{\rm H2}$ and significant increase of $f_{\rm HI}$ with increasing redshift. For the increasing trend of the H I distribution fraction as well as the decreasing trend of the H_2 distribution fraction, we discuss physical interpretations in Section 6.3.2. For interested readers, we show $f_{\rm ISM}$ for the well-mixed model in Appendix A.4.

5.3.3 Systematic Uncertainties

Before interpreting the redshift evolution of the ISM distribution fractions ($f_{\rm HI}$ and $f_{\rm H2}$), we should note that our model has many systematic uncertainties beyond the posterior distributions, relating to both measurements and modeling.

First, the normalization of the MZR is sensitive to metallicity calibrations (Kewley & Ellison 2008; Hirschmann et al. 2023). Notably, significant differences arise at low metallicities (12+log(O/H) < 8). When comparing the metallicity estimated from the R23 index of the Nakajima et al. (2022) calibration (utilized in the $z \sim 5$ MZR) to that estimated from the Curti et al. (2020) calibration (employed at $z \sim 0$), a discrepancy of 0.1 dex emerges at low metallicity (12+log(O/H) < 8), while the two are comparable at higher metallicity (12+log(O/H) > 8). On the other hand, utilizing the R23 calibrations of (Sanders et al. 2024), calibrated with JWST high redshift galaxies, results in a potentially lower metallicity by 0.1–0.2

dex at low metallicity (12+log(O/H) < 8) compared to the original Nakajima et al. (2022) metallicity. Further investigation is required to confirm how this impacts the median values and $f_{\rm ISM}$. If the observed metallicity values at $z \sim 5$ would be lower (corresponding to the Sanders et al. (2024) normalization), the estimated $f_{\rm H2}$ would correspondingly decrease, strengthening the evolution of $f_{\rm H2}$.

There also exist systematic uncertainties in molecular gas measurements, which are utilized to constrain the molecular gas mass in the NeutralUniverseMachine model. The main systematic uncertainty is the CO-to-H_2 conversion factor, $\alpha_{\rm CO}$. All CO measurements (listed in table 3 of Walter et al. (2020)) except for $z \sim 0$ measurements of xCOLD GASS, employ a Galactic conversion. However, high-redshift galaxies may possess a higher conversion factor than local galaxies due to their lower metallicity on average, with theoretical and observational support indicating an expected increase in $\alpha_{\rm CO}$ in low metallicity gas (Bolatto et al. 2013). If $\alpha_{\rm CO}$ is indeed higher at high redshifts compared to $z \sim 0$, the molecular gas estimates at high redshifts would increase compared to the original values. This would require a higher H_2 distribution fraction in our model, resulting in a smaller redshift evolution for $f_{\rm H2}$.

Atomic gas masses at high redshifts ($z > 1.5$) are derived from DLA measurements. However, DLA measurements typically do not capture low H I column density ($\log N({\rm HI}) < 20.3$ cm^{-2}), potentially resulting in an underestimation of the H I density by 10—20% (Zafar et al. 2013; Berg et al. 2019; Péroux & Howk 2020). Conversely, the contribution of H I gas outside the ISM increases towards high redshifts (Péroux & Howk 2020), accounting for approximately 10% of cosmic H I densities at $z \sim 5$ (Villaescusa-Navarro et al. 2018). While these effects may partially offset each other, the possibility of significant systematic errors cannot be discounted. Additionally, DLA measurements constrain cosmic H I density, which is most effective for constraining the H I mass in 10^{11}-10^{12} $M_\odot$ halos, and the H I content in low-mass halos may require further refinement.

The stellar yield is heavily dependent on the adopted IMF and stellar evolution model. While we assume a fixed value for the oxygen yield of 0.016, it varies from $y = 0.005$ to 0.033 depending on the IMF and model (Vincenzo et al. 2016). As discussed in Sanders et al. (2021), it is conceivable that the oxygen yield increases with redshift due to potential changes in the IMF, including a higher upper mass cutoff and a shallower high-mass slope at high redshifts. If the yield does indeed increase with redshift, the estimated $f_{\rm H2}$ would decrease at higher redshifts to match the observed metallicity, leading to a larger evolution in $f_{\rm H2}$. We show the impact of the adopted stellar yield on our modeling in A.5.

The UniverseMachine model carries typical systematic uncertainties of 0.2 dex in stellar mass, arising from assumptions in fitting SEDs (see Conroy 2013, for a review). Typical star formation histories, metallicities, and dust all change toward higher redshifts. For example, higher-redshift galaxies tend to have rising, burstier star formation histories, which may lead to inflated estimates of stellar masses if the same star formation history priors are used as for low-redshift galaxies. Lower actual stellar masses at high redshifts would lead to higher inferred $f_{\rm H2}$, and hence lower redshift evolution in $f_{\rm H2}$.

Table 5.1 Summary of theoretical models.

Name	Model Outputs	Major Observational Constraints	Redshift Range
Bolshoi–Planck	DM halo properties	Cosmology measurements	0–10
UniverseMachine	SM & SFR	SMFs, 2PCFs, & SFRs	0–10
NeutralUniverseMachine	H I & H_2 masses	H I & H_2 measurements	0–6
ChemicalUniverseMachine	ISM & CGM metallicities	Metallicity in star-forming regions	0–6

Notes. DM represents dark matter, SM refers to stellar mass, SMF stands for the stellar mass function, and 2PCF is the abbreviation for the two-point correlation function.

Galaxy Component	Reference	Redshift	Corrected
Star-forming regions	Curti et al. (2020)	$z \sim 0$	Yes
	Sanders et al. (2021)	$z \sim 2.3$ & 3.3	No
	Nakajima et al. (2023)	$4 < z < 6$	No
	Curti et al. (2024)	$4 < z < 6$	No
H I gas	Møller et al. (2013)	$0.1 < z < 5$	No
CGM	Prochaska et al. (2017)	$z \sim 0$	No

Table 5.2 Summary of observational constrains on the mass metallicity relation.

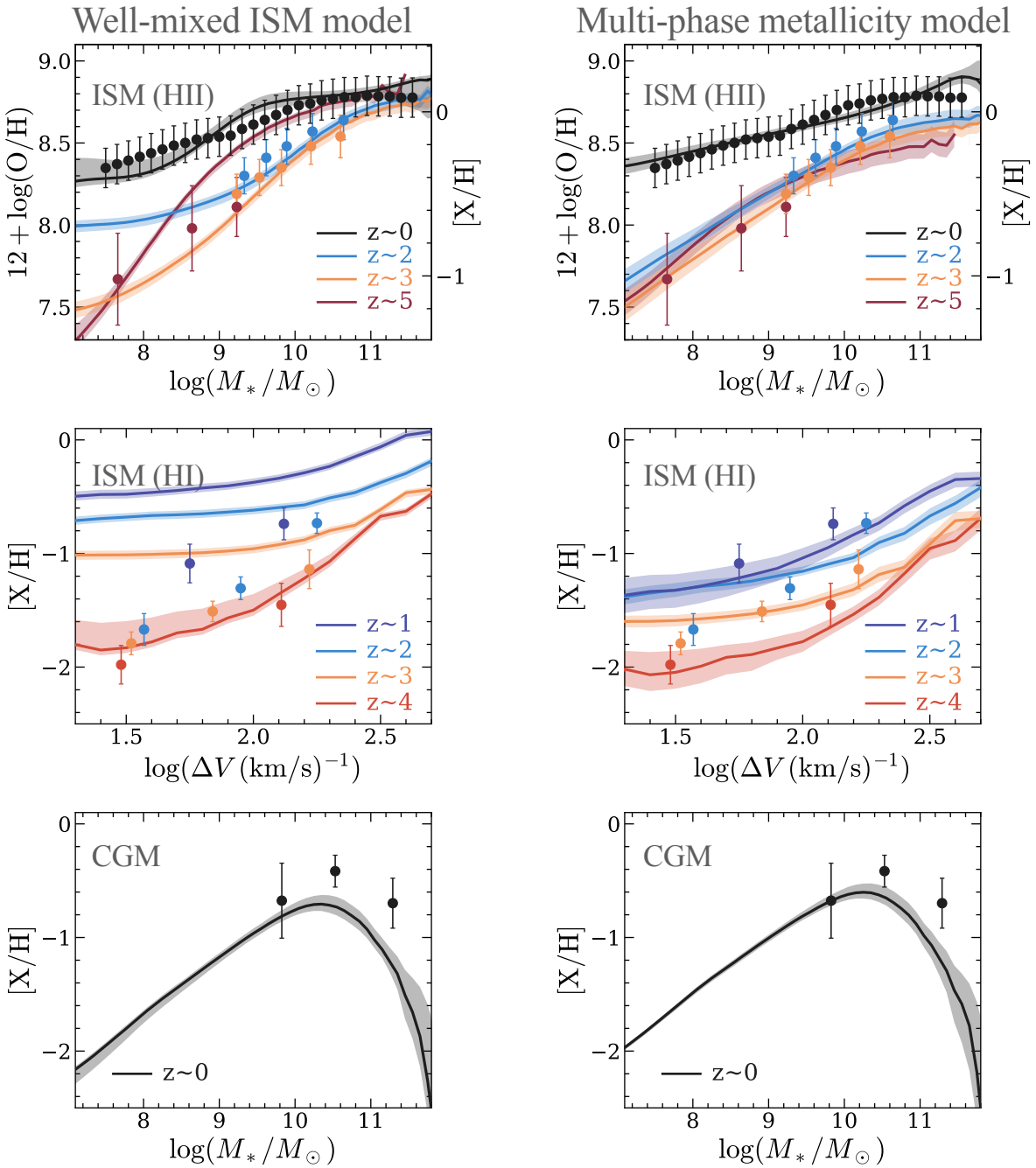

Fig. 5.4 Comparison between the models (curves) and the observed mass-metallicity relations (circles). Left: The well-mixed ISM model in star-forming regions (top), H I regions (middle), and the CGM (bottom). The shaded regions represent 68th percentile of the posterior distributions. Colored circles represent observational data at $z \sim 0$ (Curti et al. 2020), $z \sim 2$, $z \sim 3$ (Sanders et al. 2021), and $z \sim 5$ (Nakajima et al. 2023; Curti et al. 2024) in the top panel; $z \sim 1-4$ (Møller et al. 2013) in the middle panel; and $z \sim 0$ (Prochaska et al. 2017) in the bottom panel. Note that the observational data of H I metallicity is not used for the model fitting. Right: Same as the left panel, but for the multi-phase metallicity model, with additional observational constraints of H I metallicity. This model shows better fits to the observational data.

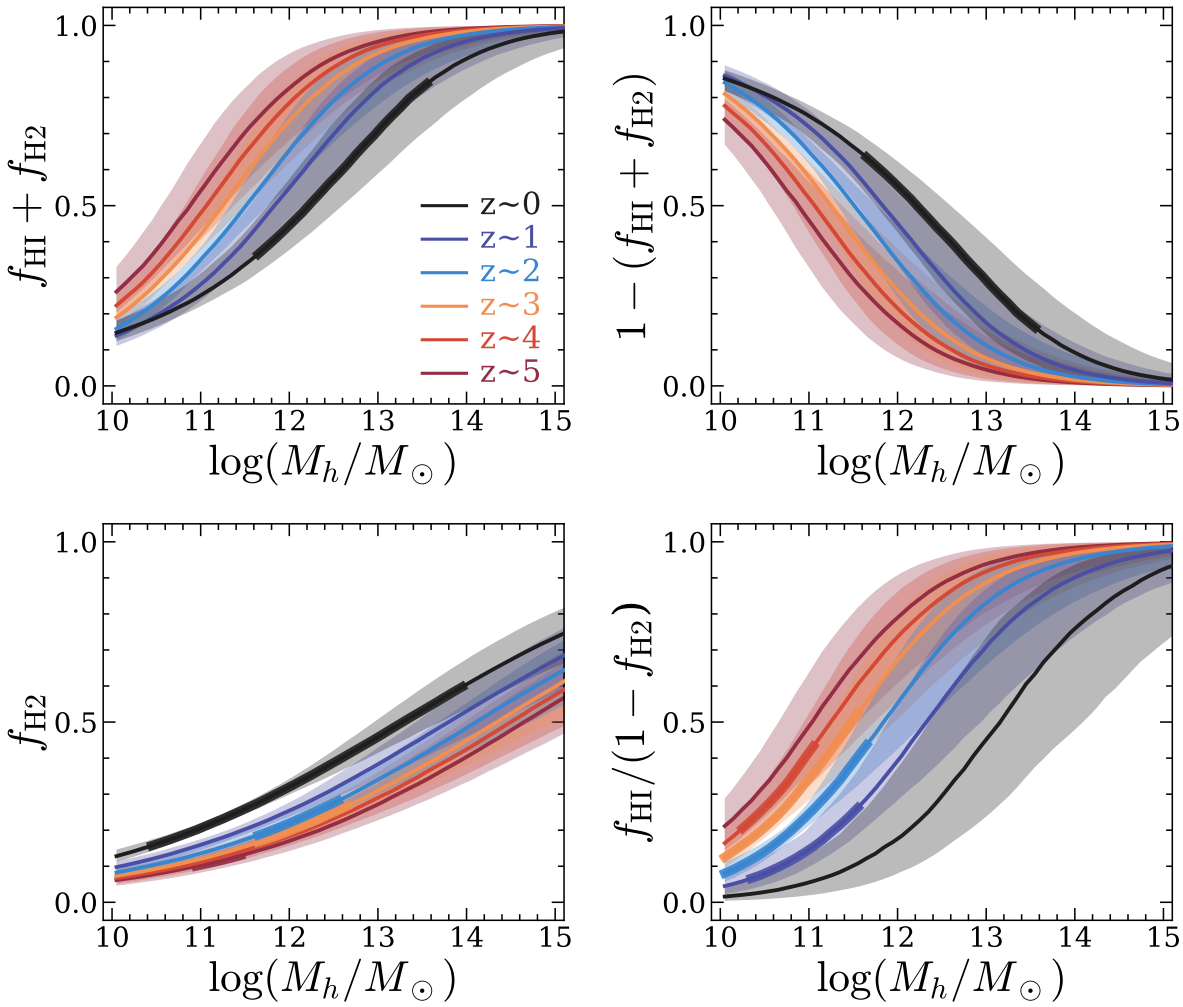

Fig. 5.5 The ISM distribution fractions of the multi-phase metallicity model. Each panel shows the distribution of metals within the multi-phase gas of the galaxy, including the total distribution fraction to the ISM (i.e., either H_2 or H I regions; upper left), the CGM (upper right), H_2 regions (lower left), and the fraction to H I among metals that are not distributed to H_2 (lower left). In all panels, thick lines show the mass ranges covered by the observational data. The shaded regions represent 68th percentile of the posterior distributions.

Chapter 6

Discussion

In this final chapter, we synthesize the individual findings of this thesis to construct a comprehensive discussion from the perspective of the cosmic evolution of the baryon cycle in galaxies. We connect the insights gained from our studies of the local universe (Chapter 2), the high-redshift universe (Chapters 3 and 4), and our theoretical framework (Chapter 5). By weaving these threads together, we present a unified picture of the evolution of the baryon cycle as revealed by this thesis.

6.1 Baryon Cycle in the Local Universe: An Equilibrium State

Our investigation begins in the local universe, where the processes of galaxy evolution are most accessible to detailed study. The unique properties of EMPGs provide a critical test for our understanding of the baryon cycle in its current state. As mentioned in Section 1.3.1, there are two possible scenarios that explain the physical origins of local EMPGs: the episodic and first star-formation scenarios. In the episodic star formation scenario, EMPGs have intermittent star-formation histories, where stochastic accretion events of metal-poor gas occur repeatedly on cosmological timescales, causing star formation. As a result, EMPGs have high sSFRs and low gas-phase metallicities. In contrast, in the first star-formation scenario, EMPGs are experiencing their first starburst with very metal-poor gas similar to the primordial gas. Here, we discuss how our findings distinguish between these models.

In Section 2.6, we show that the SFR-MZ relation of EMPGs is comparable to that of the model which is fitted to the local SDSS galaxies, which satisfies the balance between inflow and outflow. Although this result does not rule out the first star-formation scenario, the episodic star-formation scenario is more likely because our EMPGs can be explained in the same way as for the local SDSS galaxies with the equilibrium model. The episodic star-formation scenario is also supported by previous studies from the aspects of morphologies and metal's spatial inhomogeneities (e.g., Sánchez Almeida et al. 2015; Isobe et al. 2021; Nakajima et al. 2024).

The conclusion that local EMPGs are best explained by the episodic star-formation scenario within the framework of the equilibrium model has profound implications. It suggests that the baryon cycle in the present-day universe is, on the whole, in an equilibrium state, where even extreme phenomena can be explained by stochastic gas accretion rather than requiring a fundamentally different mode of galaxy formation.

This establishes a crucial baseline for comparison with the early universe.

6.2 Baryon Cycle in the Early Universe: Signs of a Non-Equilibrium State

Having established the equilibrium nature of the local baryon cycle, we now confront this picture with new evidence from the early universe. As detailed in Chapter 4, our JWST observations revealed a systematic deviation from the local FMR at high redshifts. This section explores the physical implications of this finding.

Our findings indicate that high-redshift galaxies systematically deviate from the local FMR, with a significant deviation at $z \sim 8$, where galaxies exhibit lower gas-phase metallicities than predicted from their stellar mass and star formation rate. This implies changes in the physical processes governing the baryon cycle at these early epochs. Here we discuss two plausible scenarios.

6.2.1 Enhanced Gas Cycling via Inflows and Outflows

The FMR is often interpreted as a reflection of an equilibrium state, in which galaxies regulate their star formation and chemical enrichment through a balance of gas inflows, star formation, and outflows (e.g., Davé et al. 2012; Lilly et al. 2013). The observed drop in metallicity at high redshifts implies that galaxies at this epoch may not have reached such a self-regulated state. In such a scenario, galaxies at high redshifts are likely to undergo repeated cycles of intense star formation bursts followed by strong outflows. The enriched gas can be expelled into the IGM, and subsequent accretion of metal-poor gas from the IGM can lower the metallicity of the ISM. A key reason for this could be that the supply of pristine gas via inflows overwhelms the ejection of enriched gas via outflows, thereby breaking the equilibrium. The cosmological hydrodynamic simulations by Harada et al. (2023) illustrate this exact mechanism, showing that the outflow rate remains nearly constant at $z > 5$, whereas the inflow rate increases with redshift as approximately $(1 + z)^{2.5}$. For halos with $\log M_{\rm h}/M_\odot \approx 11$, which correspond to galaxies in our sample with $\log M_*/M_\odot \approx 9$, the inflow rate starts to exceed the outflow rate at $z \sim 5 - 6$. This transition directly enhances the dilution of metals in the ISM through the accretion of low-metallicity gas. We further explore this idea in section 6.3.1.

6.2.2 Transition Driven by Cosmic Reionization

An alternative, yet complementary, explanation involves the impact of cosmic reionization on gas accretion processes. This scenario posits that the transition from the metal-poor phase at $z > 8$ to the equilibrium phase at $z < 6$ is regulated by the suppression of pristine gas supply due to reionization.

At $z > 8$, before reionization is complete, low-mass ($M_{\rm h} \lesssim 10^8 M_\odot$) dwarf galaxies can retain their gas and actively form stars. These abundant low-mass halos likely serve as a rich source of pristine gas, fueling more massive galaxies through minor mergers and smooth accretion. This continuous supply of metal-poor gas would effi-

ciently dilute the ISM of galaxies, maintaining them in the observed low-metallicity state.

However, as reionization proceeds and the cosmic UV background is established ($z \lesssim 6$), the photoheating of the IGM suppresses gas accretion onto low-mass halos (e.g., Foreman & Vollmer 2013; Hasegawa & Semelin 2013). Consequently, the supply of pristine gas via minor mergers is cut off. With the dilution channel suppressed, galaxies would then evolve toward higher metallicities through secular star formation, eventually settling onto the Fundamental Metallicity Relation observed at lower redshifts.

In summary, the break of the FMR at high redshifts provides powerful observational evidence for a fundamental shift in the mode of galaxy growth, suggesting that the baryon cycle is in a non-equilibrium state in the early universe. What physical processes, then, drove this observed evolution? To address this question, we now turn to a theoretical interpretation using the empirical model developed in this thesis.

6.3 Physical Interpretation of the Baryon Cycle Evolution

6.3.1 Insights from the ChemicalUniverseMachine on the FMR evolution

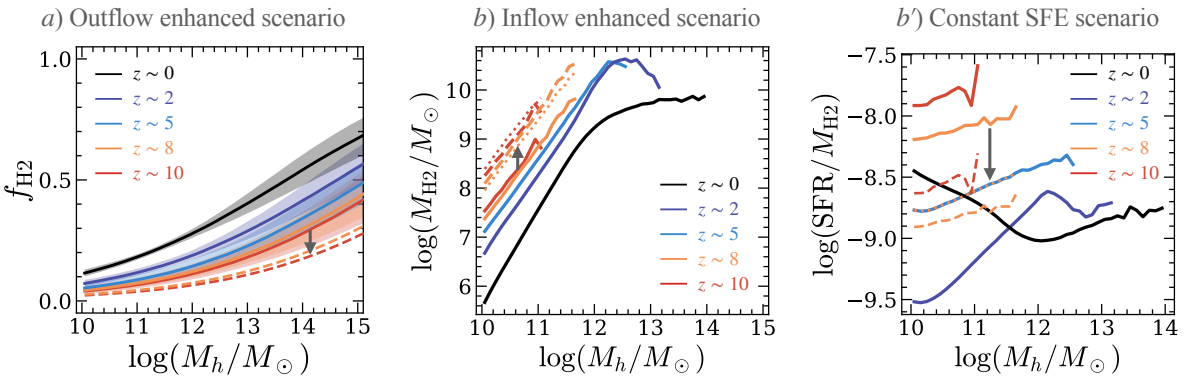

Fig. 6.1 **Left**: The H_2 distribution fraction (f_{H2}) predicted by the ChemicalUniverseMachine as a function of halo mass. Different colors correspond to redshifts $z \sim 0, 3, 4, 6$, and 8. Solid lines show the standard extrapolation and shaded regions represent the 68th percentile distributions. Dashed lines represent Scenario a, which assumes a lower f_{H2} at the fixed halo mass for $z > 5$. **Middle**: The molecular gas mass (M_{H2}) as a function of halo mass. Colors are the same as in the left panel. Solid lines show the standard extrapolation. Dashed lines represent Scenario b, which assumes a higher M_{H2} at the same halo mass. Dotted lines indicate Scenario b' for reference, which shows gas masses nearly identical to Scenario b. **Right**: The star formation efficiency (SFE = SFR/M_{H2}) as a function of halo mass. Colors and line styles are the same as in the middle panel. Dotted lines represent Scenario b', where the SFE is fixed at the $z = 5$ value for $z > 5$. Dashed lines show Scenario b for reference, which yields SFE values comparable to those at $z \sim 0$–5.

To explore the physical drivers of the observed FMR breakdown, we utilize the

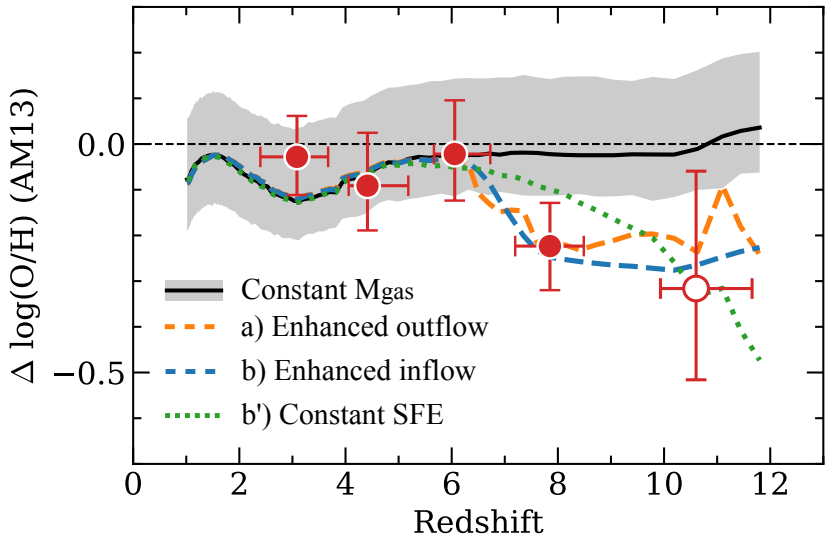

Fig. 6.2 FMR evolution shown with predictions from the ChemicalUniverseMachine model (Nishigaki et al. 2025). Filled and open red circles are the same as in Figure 4.11: filled circles represent our results derived from stacked spectra using the direct method at $z \sim 3-8$, while the open circle at $z \sim 10$ shows the average metallicity of individual galaxies. The curves and shaded regions represent the median predictions and the 68th percentile ranges, respectively, for halos with SFR $= 10\,M_\odot\,\mathrm{yr}^{-1}$. The black line shows standard extrapolation where the molecular gas mass remains roughly constant at fixed halo mass. This scenario predicts no significant deviation from the FMR, failing to reproduce the observed drop at $z > 8$. The colored lines represent different physical scenarios at $z > 5$. The orange dashed line shows the Enhanced outflow Scenario (Scenario a), where the metal retention fraction (f_{H2}) is reduced to model efficient metal removal. The blue dashed line shows the Enhanced inflow Scenario (Scenario b), where the molecular gas mass (M_{H2}) is explicitly increased to model strong dilution. The green dotted line shows the Constant SFE Scenario (Scenario b'), where the star formation efficiency is kept constant, implicitly leading to larger gas reservoirs similar to the Scenario b.

ChemicalUniverseMachine model presented in Chapter 5. As described therein, this framework models metallicity evolution by parameterizing the metal distribution fractions, f_{H2} and f_{HI}. Here, we focus on f_{H2}, as it directly regulates the metallicity of the star-forming gas. Applying this framework to the high-redshift regime ($z > 5$), however, requires extrapolation, as the gas mass fitting functions in the underlying NeutralUniverseMachine (NUM) are calibrated only at $z \lesssim 5$. In the framework of this model, the evolution of the FMR is governed by three primary factors: the star formation rate, the gas inflow rate (traced by the molecular gas mass, M_{H2}), and the outflow efficiency (parameterized by f_{H2}). Since the SFR evolution is already constrained by the UniverseMachine, we explore potential evolutionary pathways for the remaining two components, gas mass and metal retention. Note that since the model constrains average relations between stellar mass, SFR, and metallicity for a given halo mass and redshift, rather than tracing individual halo histories, we assume

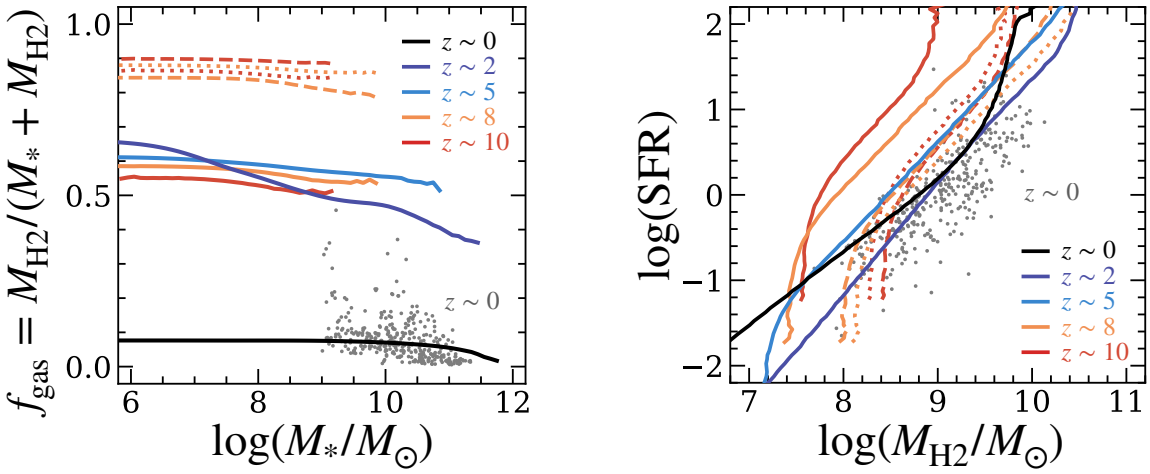

Fig. 6.3 Predictions for the evolution of gas properties in the ChemicalUniverseMachine. **Left:** The gas fraction, defined as $f_{\rm gas} = M_{\rm H2}/(M_{\rm H2} + M_*)$, as a function of stellar mass. Gray dots represent local galaxies from the xCOLD GASS survey (Saintonge et al. 2017), which are well reproduced by the model at $z \sim 0$ (purple curve). While the standard extrapolation (solid lines) predicts $f_{\rm gas} \sim 0.5$–0.7 at $z > 5$, the scenarios that reproduce the FMR break (Scenarios b and b'; dashed and dotted lines) imply extremely high gas fractions of $f_{\rm gas} \approx 0.8$–0.9. **Right:** The relationship between star formation rate and molecular gas mass ($M_{\rm H2}$). Colors and symbols are the same as in the left panel. The standard extrapolation implies an increase in star formation efficiency (SFR/$M_{\rm H2}$) with redshift. In contrast, Scenarios b and b' remain consistent with the relation observed at $z \sim 0$–5, suggesting little to no evolution in the star formation law.

for the purpose of FMR comparison that $f_{\rm H2}$ is a function solely of halo mass and redshift, independent of the specific star formation rate.

We define a baseline using the NUM-based extrapolation, which extends the scaling relations calibrated in the NeutralUniverseMachine at $z < 5$ directly out to $z > 5$. In this standard model, the star formation efficiency (SFE; defined as SFR/$M_{\rm H2}$) increases with redshift, while the gas mass ($M_{\rm H2}$) remains roughly constant at fixed halo mass. To investigate the origin of the observed FMR deviation, we compare this baseline with three alternative scenarios for $z > 5$.

The first is the Outflow Enhanced scenario (Scenario a), where we assume that the metal distribution fraction ($f_{\rm H2}$) decreases from $z \sim 6$ to $z \sim 8$, representing stronger metal loss (Figure 6.1, Left). To reproduce the observed FMR deviation at $z \sim 8$, $f_{\rm H2}$ is required to be $\sim$0.6 times the value predicted by the standard extrapolation, which corresponds to roughly a factor of $\sim$1.7 increase in outflow efficiency relative to the baseline.

The second is the Inflow Enhanced scenario (Scenario b), where we modify the evolution of $M_{\rm H2}$ such that it increases beyond the standard extrapolation between $z \sim 6$ and $z \sim 8$, also adjusted to match the observations (Figure 6.1, Middle). In this scenario, the gas mass at $z \sim 8$ is $\sim$5 times higher than the value predicted by the standard extrapolation, implying an increase by a factor of five in inflow efficiency.

The third is the Constant SFE scenario (Scenario b'), which serves as a physically

motivated variation for increasing the gas mass. This model assumes that the star formation efficiency remains constant at $z > 5$ (Figure 6.1, Right). Because the SFR increases with redshift at fixed halo mass, this assumption inherently implies that $M_{\rm H2}$ also increases, resulting in a gas-rich state. Specifically, this leads to a gas mass approximately 3 times higher than the standard extrapolation, corresponding to an increase by a factor of three in inflow efficiency.

We compare the predictions of these scenarios with our observational results in Figure 6.2. The standard NUM-based extrapolation (black curve) predicts no significant deviation from the FMR even at $z > 8$, failing to reproduce the observed decline in metallicity. In contrast, Scenarios a, b, and b' all predict lower metallicities at higher redshifts, showing good agreement with the observed trend. Here, we examine the behavior of the model parameters in these successful scenarios to understand their physical implications.

Scenario a assumes gas masses and star formation efficiencies unchanged from the standard model but reproduces the observations by assuming a decrease in the metal distribution fraction ($f_{\rm H2}$). Physically, this implies that metal loss is the dominant mechanism. A lower $f_{\rm H2}$ signifies that a smaller fraction of the metals produced by supernovae is retained within the star-forming regions. This suggests that metals are being more efficiently ejected into the CGM or IGM, pointing to enhanced outflow efficiency in the early universe compared to lower redshifts.

In contrast, Scenarios b and b' are characterized by a significant increase in gas mass ($M_{\rm H2}$) at fixed halo mass compared to the standard extrapolation, while the metal distribution fraction remains unchanged. To quantify the gas properties implied by these scenarios, we present the evolution of the gas fraction and the star formation law in Figure 6.3. The left panel shows the gas fraction, $f_{\rm gas} = M_{\rm H2}/(M_{\rm H2}+M_*)$. At high redshifts ($z > 5$), the standard extrapolation predicts gas fractions of $f_{\rm gas} \approx 0.5$–0.7. However, Scenarios b and b', which successfully reproduce the FMR break, require significantly higher values of $f_{\rm gas} \approx 0.8$–0.9. This indicates that galaxies in these epochs are likely extremely gas-dominated systems. The right panel shows the relation between SFR and $M_{\rm H2}$. The standard extrapolation (solid lines) implies a shift toward higher SFR at fixed gas mass, indicating an increase in SFE. Conversely, Scenarios b and b' align closely with the relation at $z \sim 0$ even out to $z \sim 10$. This constancy implies that the local physical processes governing the conversion of molecular clouds into stars have not fundamentally changed from lower redshifts. Consequently, the primary driver of the observed state in these scenarios is not a change in the mode of star formation, but rather the availability of fuel. The presence of a larger reservoir of gas relative to the stellar mass dilutes the concentration of metals produced by star formation. Such a gas-rich state corresponds to a scenario where the inflow of pristine gas is significantly enhanced in the early universe, leading to an accumulation of gas that drives high star formation rates while simultaneously lowering the metallicity. This picture is consistent with ALMA [C II] observations of $z \sim 5$ galaxies, which indicate that gas fractions increase with redshift, while the star formation law remains similar to that in the local universe (Dessauges-Zavadsky et al. 2020; Schaerer et al. 2020).

These comparisons demonstrate that to explain the breakdown of the FMR at high redshifts, the physical conditions of galaxies should deviate from the standard ex-

trapolation in two ways. One possibility is enhanced outflows, which lead to a more efficient removal of metals from the galaxy, as represented by Scenario *a*. Alternatively, the deviation can be explained by enhanced inflows, resulting in a stronger dilution of the ISM driven by larger gas reservoirs (Scenarios *b* and *b*′). While it is difficult to definitively distinguish between these mechanisms with current data, our analysis with the ChemicalUniverseMachine suggests that the early universe is in a non-equilibrium state driven by one, or a combination, of these evolving physical processes.

6.3.2 Evolution of Metal Distribution and the Physical Drivers

Evolution of ISM Distribution Fractions

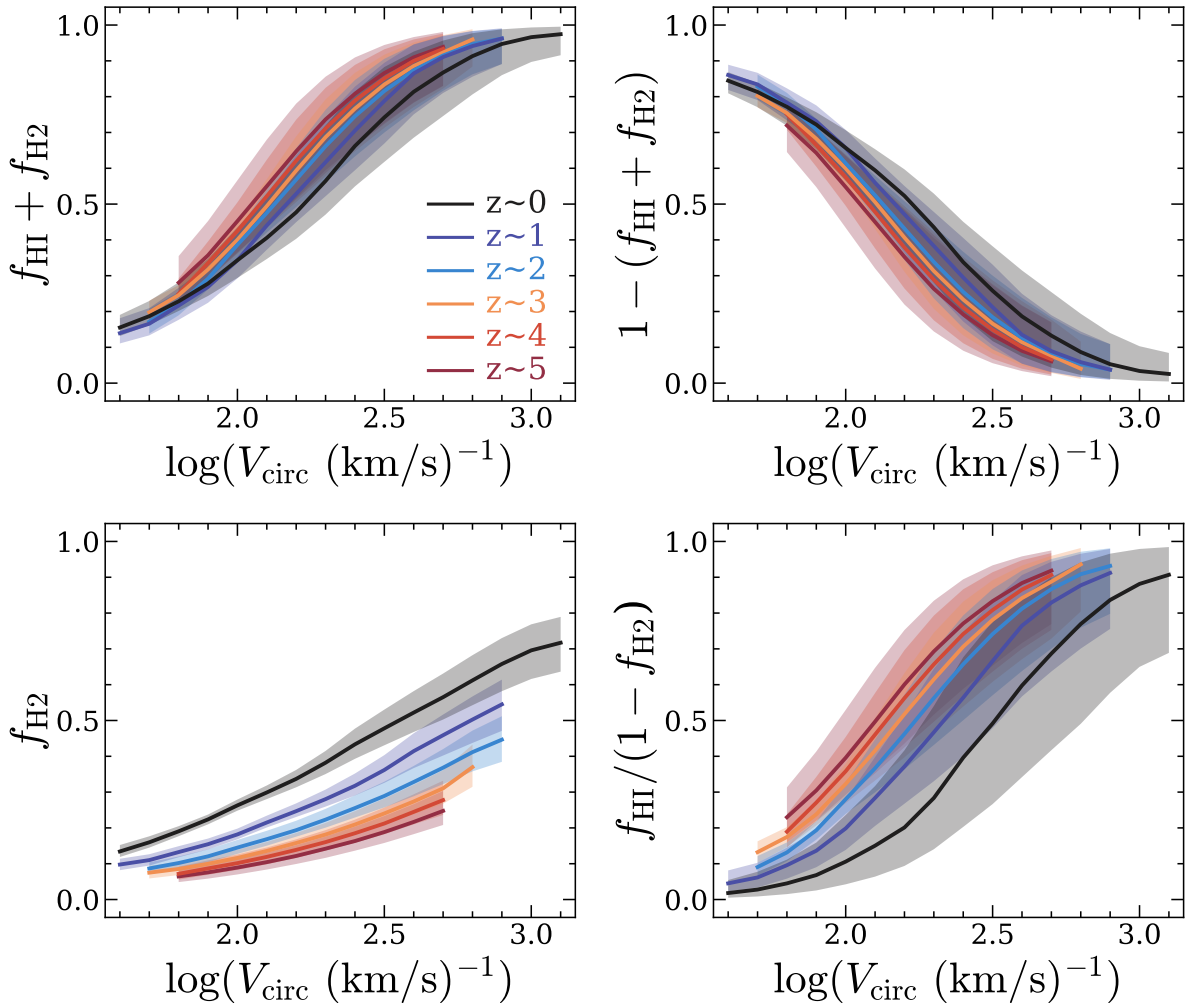

Fig. 6.4 Same as Figure 5.5, but as a function of halo circular velocity.

The multi-phase metallicity model indicates an upward trend of the total ISM distribution fraction ($f_{\rm HI}$ + $f_{\rm H2}$) with redshift (top left panel of Figure 5.5). We also evaluate the ISM distribution fractions as a function of halo circular velocity, which represents the gravitational potential well depth of halos (Figure 6.4). Interestingly,

the total ISM distribution fraction ($f_{\rm HI}$ + $f_{\rm H2}$) does not evolve with redshift at fixed circular velocity. This indicates a universal trend where the fraction of outflowing gas that reaches the CGM is determined by halo potential well depth only.

In contrast, there is redshift evolution in $f_{\rm HI}$ and $f_{\rm H2}$ both at fixed halo mass and at fixed circular velocity. The H_2 distribution fraction decreases with redshift by 0.3–0.5 dex from $z = 0$ to $z = 5$ at $M_{\rm h} \sim 10^{12}\ M_{\odot}$, while the H I distribution fraction increases by 0.5–1.2 dex.

We cannot ascribe a cause to this behavior with our current modeling approach. Nonetheless, we view it as reasonable given at least two physical possibilities:

- Higher-redshift galaxies have stronger outflows, but also shorter cooling times. For example, Sugahara et al. (2019) show that $z \sim 6$ galaxies have higher outflow velocities at a fixed circular velocity. Strong outflows can remove gas and metals from the ISM, which can then quickly recool into the H I phase, leading to a higher fraction of metals distributed to H I vs. H_2.
- Galaxies in the local universe typically exhibit thin stellar disks surrounded by more extended H I gas reservoirs. In contrast, high-redshift galaxies likely possess disordered thick disks because these galaxies are more likely to experience perturbations from increased rates of mergers or inflow from the IGM. Consequently, high-redshift galaxies may have a higher covering fraction of H I, facilitating easier mixing of gas and metals between H I and star-forming regions.

At the same time, given the observational uncertainties in measuring both gas masses and metallicities, as well as their evolution with redshift (Section 5.3.3), it may also be plausible that improved future measurements show different quantitative trends for the evolution of $f_{\rm H2}$ vs. $f_{\rm HI}$ with redshift.

Evolution of CGM Metallicity

Figure 6.5 shows the evolution of the CGM metallicity at $0 < z < 5$ predicted by the multi-phase metallicity model. It indicates slow enrichment at $z > 3$, following rapid enrichment below $z \sim 2$. This is attributable to the evolution of the stellar-to-halo mass ratio (see Figure 18 of Behroozi et al. 2019) combined with increasing the distribution fraction to the CGM towards low redshifts ($1 - (f_{\rm HI} + f_{\rm H2)}$; right top panel in Figure 5.5) at fixed halo mass.

Observations of CGM metallicity with Lyman-limit systems indicate a rapid evolution of around 1 dex from $2.3 < z < 3.3$ to $z < 1.0$ (Tumlinson et al. 2017; see also Lehner et al. 2013 and Wotta et al. 2016). The magnitude of this observed evolution is consistent with the predictions of our model if the typical host halo mass of Lyman-limit systems changes from $M_{\rm h} \sim 10^{11}\ M_{\odot}$ to $M_{\rm h} \sim 10^{12}\ M_{\odot}$.

This rapid enrichment of the CGM at high redshifts establishes a reservoir of metals surrounding galaxies. As we will discuss in Section 6.4, this pre-enriched environment likely sets the boundary condition—the metallicity floor—observed in the local universe.

Comparison with Mass-Loading Factor

Our results suggest that mass-dependent outflows play a key role in shaping the MZR, consistent with previous chemical evolution models (e.g., Finlator & Davé 2008; Peeples & Shankar 2011; Sanders et al. 2021; Tortora et al. 2022). These studies typically incorporate the mass-loading factor or metal-loading factor, both of which exhibit an inverse correlation with stellar mass and circular velocity. This behavior aligns qualitatively with our findings for $1 - (f_{\rm HI} + f_{\rm H2})$, as shown in the upper-right panels of Figures 5.5 and 6.4. While $f_{\rm ISM}$ (or $f_{\rm HI} + f_{\rm H2}$) and the mass-loading (or metal-loading) factor are conceptually similar, a direct quantitative comparison between these parameters is limited due to differences in their definitions.

Specifically, $f_{\rm ISM}$ accounts for the direct loss of supernova yields, while the mass-loading factor includes the loss of metals from the ISM along with galactic winds. Additionally, the mass-loading factor encompasses only the instantaneous ejection of gas. Depending on the feedback model, this ejected gas may return to the ISM at a later stage, especially for low-mass galaxies with short cooling times. In contrast, $f_{\rm ISM}$ focuses on the total fraction of metals, including those that are directly deposited and those that eventually return to the ISM.

The metals already present in the ISM from previous enrichment episodes and removed by a galactic wind would contribute negatively to $f_{\rm ISM}$. Although we initially explored an alternative parametrization (i.e., additional normalization parameters to Equations 5.4, 5.7, and 5.8) that allowed for negative $f_{\rm ISM}$ values, observational data strongly constrain $f_{\rm ISM}$ to have positive values. This aligns with the physical picture that a negative $f_{\rm ISM}$ would rapidly deplete metals from the ISM, whereas observed galaxy populations show an increasing metal mass over time.

6.4 The Metallicity Floor: A Legacy of Cosmic Chemical History

Here we revisit the problem of the metallicity floor in the local universe. The non-detection of Hyper Metal-Poor Galaxies (HMPGs) with $Z < 0.01 Z_\odot$ in the local universe search can be interpreted as an inevitable consequence of this chemically enriched universe, where the IGM/CGM has already been significantly enriched with metals after billions of years of cosmic star formation and outflows. Indeed, the mean metallicity of the Universe predicted from cosmic star formation density measurements increases with decreasing redshift and reaches $\sim 0.1\ Z_\odot$ at $z = 0$ (Madau & Dickinson 2014). This general picture of cosmic enrichment is further supported on the local scale in the results from our ChemicalUniverseMachine model. As shown in Section 6.3.2, our model predicts that the CGM of galaxies in halos of $\sim 10^{11}\ M_\odot$ at $z \sim 0$ is enriched to approximately $0.01\ Z_\odot$. Thus, both the cosmic average and the conditions local to galaxies indicate that the gas available for accretion in the present-day universe is no longer pristine. Consequently, this gas has a natural lower limit in its metallicity, which in turn prevents the ISM metallicity from dropping to extremely low values and explains the observed metallicity floor.

To empirically verify the existence of this floor, we revisit our observational results

from Chapter 2. Although we identified three new EMPGs, no HMPGs were found in our spectroscopic sample. Based on this null result, we place an upper limit on the number density of star-forming HMPGs in the local universe. We consider a hypothetical HMPG model with a star formation rate of SFR $= 0.1\,M_{\odot}\mathrm{yr}^{-1}$. Assuming typical physical conditions for metal-poor galaxies ($\log U = -1.5$), the expected [O III]λ4363 luminosity is $2.0 \times 10^{38}\,\mathrm{erg\ s^{-1}}$. Given the 3σ flux limit of our MagE observations ($4.2 \times 10^{-16}\,\mathrm{erg\ s^{-1}cm^{-2}}$), such a galaxy would be detectable via the auroral line only up to $z_{\mathrm{max}} \approx 0.014$. Note that at this redshift ($z \approx 0.014$), the target would be sufficiently bright to be detected in the SDSS imaging data ($r_{\mathrm{lim}} \sim 22$). Indeed, we observe a galaxy in our sample with similar properties ($z = 0.016$, $\log(\mathrm{SFR}) = -1.04$) that has an apparent magnitude of $r = 18.7$, nearly 3 magnitudes brighter than the photometric detection limit. Thus, the survey volume is strictly limited by the spectroscopic sensitivity for the [O III]λ4363 line rather than the photometric depth.

The effective survey volume V_{eff} is estimated by scaling the comoving volume within z_{max} by the sky coverage and the spectroscopic sampling rate:

$$V_{\mathrm{eff}} = V(z < 0.014) \times \frac{A_{\mathrm{SDSS}}}{4\pi \mathrm{sr}} \times \frac{N_{\mathrm{obs}}}{N_{\mathrm{cand}}} \tag{6.1}$$

where $A_{\mathrm{SDSS}} = 14555\ \mathrm{deg}^2$ is the SDSS survey area, and $N_{\mathrm{obs}}/N_{\mathrm{cand}} = 10/104$ represents the follow-up fraction. This yields an effective survey volume of $V_{\mathrm{eff}} \approx 3.2 \times 10^4\,\mathrm{Mpc}^3$.

Since no HMPGs were detected in this volume, we derive the 95% confidence level upper limit on their number density using Poisson statistics (rule of three):

$$n_{\mathrm{HMPG}} < \frac{3}{V_{\mathrm{eff}}} \approx 9.5 \times 10^{-5}\,\mathrm{Mpc}^{-3}. \tag{6.2}$$

This limit suggests that HMPGs with moderate star formation activity (SFR $\sim 0.1\,M_{\odot}\mathrm{yr}^{-1}$) are extremely rare in the local universe, consistent with the scenario of a metallicity floor. This statistical constraint confirms that galaxies with primordial-like compositions are virtually absent in the local universe.

This naturally leads to the understanding that HMPGs can indeed be found at higher redshifts, where the IGM was significantly less chemically enriched. In such pristine environments, accretion could fuel star formation with gas much closer to primordial composition, allowing galaxies to exist below the locally observed floor. The high-redshift EMPG candidates identified in Chapter 3 of this thesis, characterized by their extremely weak [O III] emission, represent potential examples of such a population. Furthermore, recent discoveries of individual galaxies with less than 1% solar metallicity at $z > 4$ by other JWST studies (e.g., Nakajima et al. 2025) lend strong support to this picture, suggesting the metallicity floor is an evolutionary feature rather than a fundamental limit.

6.5 Further Considerations on Statistical Averages and Individual Dynamical States in Local EMPGs

The results in Chapter 2 show that the average relationship between stellar mass, metallicity, and SFR for EMPGs is broadly consistent with the low-mass extrapolation of the equilibrium model (Lilly et al. 2013) derived from SDSS galaxies. This suggests that, when viewed statistically as a whole population, EMPGs possess the property of converging toward a quasi-steady state over long timescales, where gas inflow, star formation, and outflow reach a balance.

However, it is important to note that the FMR defined by large-scale surveys like SDSS, and the equilibrium model used to explain it physically, are essentially statistical and averaged models that treat a galaxy as a self-contained, isolated system. In contrast, many EMPGs are suggested to be dynamic rather than isolated systems. As demonstrated by Isobe et al. (2021), a significant fraction ($\sim$85%) of EMPGs are associated with "tail" structures and exhibit large velocity offsets (e.g., ~ 100 km s^{-1}), indicating that these galaxies can be in a dynamic state such as undergoing gas or satellite infall.

This thesis focuses on the observed average, which is considered to represent the statistical destination where individual objects, initially in temporary non-equilibrium states due to intense gas infall, eventually settle after time-integration or as a population. In existing FMR studies using SDSS galaxies, it is common to stack or average objects with similar stellar masses and SFRs to ensure a sufficient signal-to-noise ratio; consequently, the scatter of individual galaxies from the FMR has not been extensively discussed. In dynamical systems such as EMPGs, individual galaxies are expected to deviate from the equilibrium model, depending on the specific timing and intensity of gas inflow events.

Therefore, it is essential to establish a robust baseline by investigating the intrinsic scatter of the FMR within the SDSS galaxy population. Against this baseline, we can examine whether EMPGs exhibit a characteristically larger scatter and investigate the correlation between morphological classifications and their offsets from the FMR. By moving beyond general averages and focusing on the link between structure and chemistry in individual systems, we can gain deeper insights into the physical processes that govern the chemical properties of galaxies in their early stages of formation.

6.6 Impact of Selection Biases

The sample analyzed in Chapter 4 is constructed by integrating data from multiple surveys with distinct science goals, namely DREAMS, JADES, CEERS, and ERO. These high-redshift catalogs are primarily based on rest-frame UV selection. In addition to this intrinsic survey bias, our specific requirement for robust Hβ line detections ($S/N > 3$) likely reinforces the preference for UV-bright and less obscured star-forming galaxies. This is because heavy dust extinction can suppress nebular emission lines below our detection threshold. Since dust extinction correlates with

metallicity (Heckman et al. 1998; Garn & Best 2010; Zahid et al. 2014), the exclusion of heavily dust-obscured populations suggests that our derived metallicity values could lead to underestimating the true cosmic average.

At the same time, the depth and observational strategy of the JADES survey (Eisenstein et al. 2023) provide a means to assess and partially mitigate this bias. Rather than relying solely on standard UV selection, JADES targets sources from deep NIRCam imaging extending to $\sim 5\mu$m, enabling the identification of redder, optically faint sources (D'Eugenio et al. 2025). Indeed, for the bulk of our sample at $z \lesssim 4$, we observe galaxies with relatively high color excess (median $E(B-V) \sim 0.2$), suggesting that we capture at least a portion of the moderately obscured population which contributes significantly ($\gtrsim 50\%$; e.g., Madau & Dickinson 2014; Zavala et al. 2021) to the cosmic star formation rate density at these redshifts.

The potential impact of this bias may also naturally diminish toward higher redshifts. The contribution of dust-obscured star formation to the cosmic SFRD is predicted to decrease rapidly at $z > 5$ (e.g., Bouwens et al. 2020; Zavala et al. 2021; Algera et al. 2023). This expected trend aligns with our finding of the FMR break at $z \gtrsim 8$, as the statistical dominance of heavily obscured galaxies is likely less pronounced at these early epochs compared to cosmic noon.

It is also important to consider that our investigation of redshift evolution incorporates a controlled study within a fixed M_*-SFR subsample. This approach is intended to isolate evolutionary trends from mass- and SFR-dependent selection effects. The key consideration is therefore whether a significant population of dusty galaxies exists specifically within this defined M_*-SFR window. As shown in Figure 6.6, comparison with known ALMA-detected DSFGs at $1 < z < 3$ (Dunlop et al. 2017; Aravena et al. 2020) indicates that these extreme dusty populations predominantly occupy the high-mass ($M_* \gtrsim 10^{10} M_\odot$) and high-SFR (SFR $\gtrsim 100\, M_\odot \mathrm{yr}^{-1}$) regime, falling largely outside the parameter range of our subsample. This implies that the influence of currently known DSFGs on the evolutionary trends observed within the subsample is likely limited.

In summary, while the observed redshift evolution appears robust within the parameter space accessible to current JWST observations, we acknowledge that our sample may not be fully representative of the global galaxy demographics. Our findings characterize the chemical enrichment of the relatively luminous, UV-detectable universe; whether the missing populations, including both obscured and intrinsically faint systems, follow a similar evolutionary path remains an open question that future, deeper multi-wavelength surveys must address.

6.7 Future Prospects

On the observational front, future work should include conducting high-z EMPG searches, similar to the one presented in Chapter 3, but with even deeper imaging data. Subsequently, spectroscopic follow-up observations with JWST will be essential to confirm the nature of these candidates and measure their metallicities. Furthermore, the breakdown of the FMR at redshifts $z > 8$ should be verified with higher statistical precision using larger JWST spectroscopic surveys. Increasing the sample size will

allow for finer binning in stellar mass and redshift, which will in turn make it possible to investigate whether the observed offset from the MZR and FMR has any mass dependence.

On the theoretical front, the ChemicalUniverseMachine can be further advanced in several ways. One key improvement would be to incorporate the diversity of individual galaxy star formation histories, moving beyond the population-averaged approach. The model could also be expanded to include the evolution of stellar metallicities, such as iron abundances, which would enable multifaceted tests against upcoming observations from facilities like the Prime Focus Spectrograph (PFS). Incorporating explicit parameterizations for inflow and outflow rates would allow for a more detailed analysis of the physical origins of the FMR's evolution. Moreover, gas mass measurements of high-redshift galaxies from the ALMA can be used to extend the NeutralUniverseMachine, and by extension the ChemicalUniverseMachine, to redshifts of approximately $z \sim 7$, enabling a more robust discussion of the FMR evolution at these early epochs.

The observational techniques and theoretical models presented in this thesis are poised to play a significant role in future galaxy formation studies. The high-redshift EMPG search pioneered in this thesis charts a course toward the eventual discovery of the very first galaxies. Furthermore, the developed ChemicalUniverseMachine will serve as a powerful theoretical tool for physically interpreting the increasingly large datasets from future observations. This empirical model holds a complementary relationship with physical models such as hydrodynamical simulations; the evolution of the metal distribution fraction, empirically constrained by this thesis, can provide new observational constraints for these simulations. This is expected to improve the modeling of feedback processes that have historically been subject to large uncertainties. In this way, this research, through its dual approach of observation and theory, contributes to a deeper understanding of galaxy evolution.

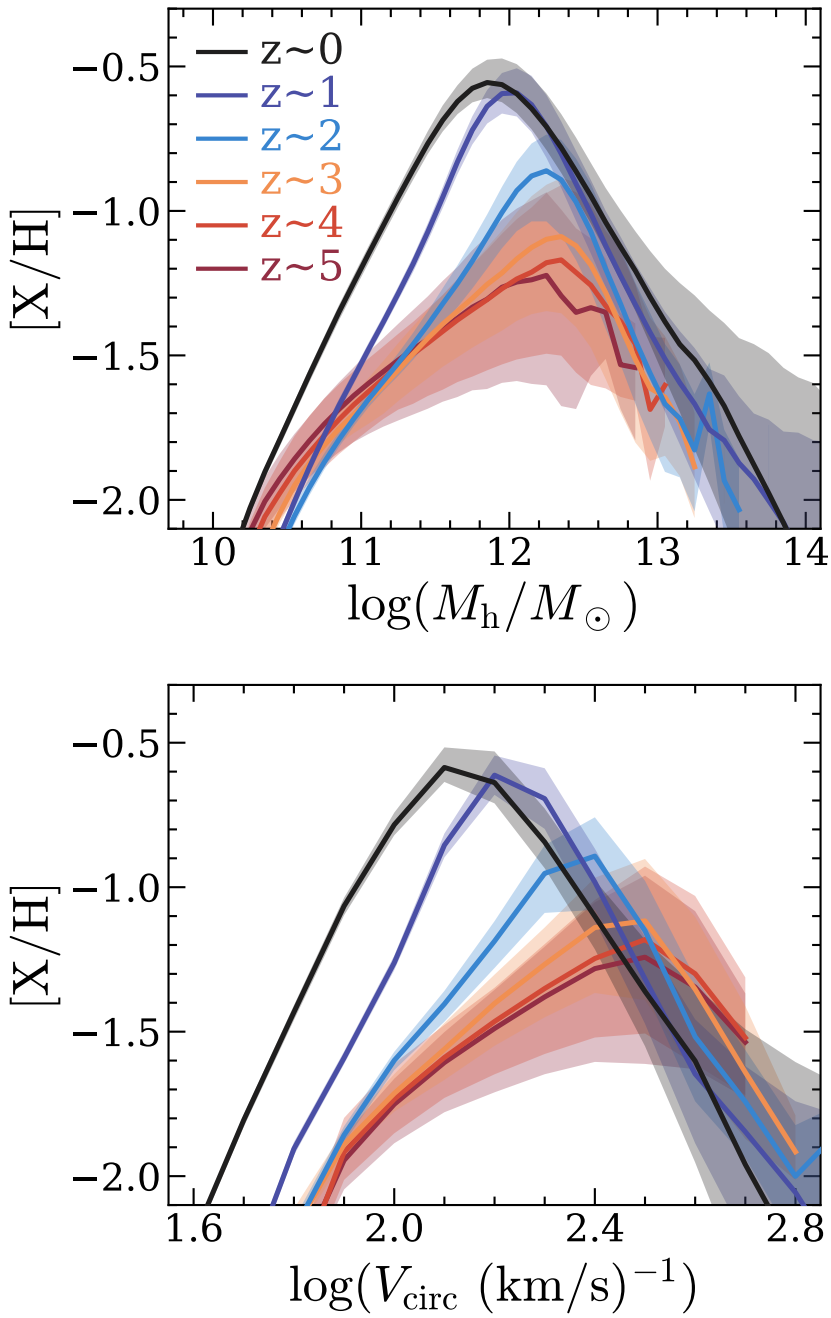

Fig. 6.5 The CGM metallicity evolution of the multi-phase metallicity model as a function of halo mass (top) and those as a function of circular velocity (bottom). We expect CGM metallicities to be higher for group and cluster-scale masses than shown here, as we do not include metals ejected from satellites in the host CGM.

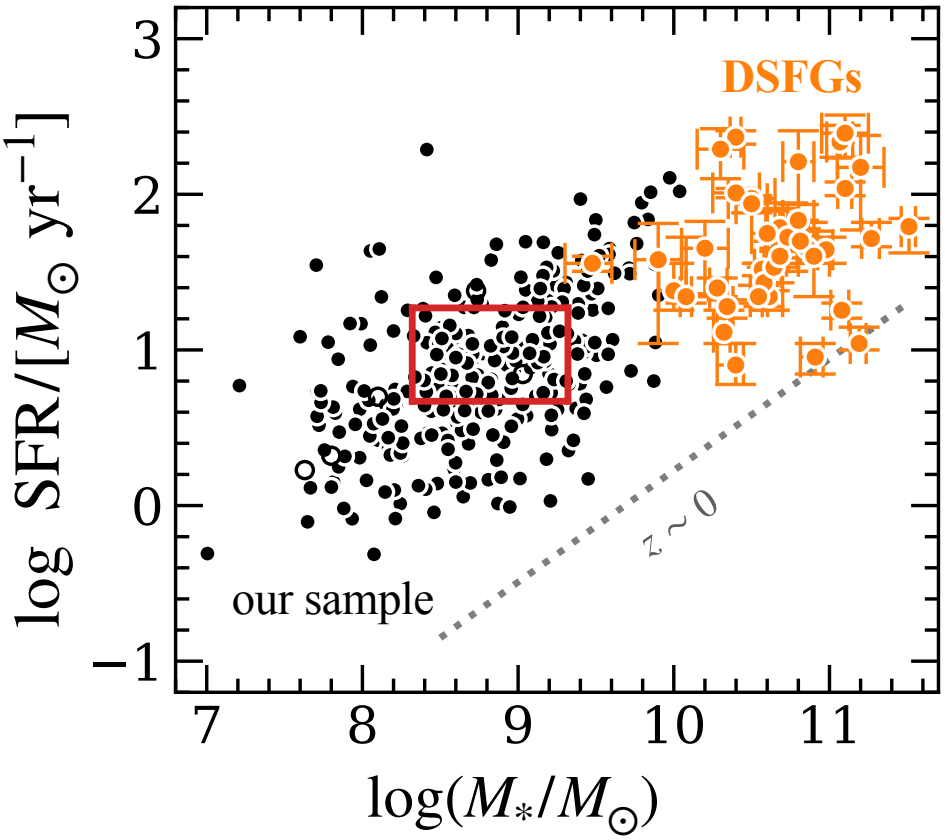

Fig. 6.6 The distribution of galaxies in the star formation rate (SFR) versus stellar mass (M_*) plane. Black dots represent the individual galaxies from our sample used in the analysis of Chapter 4. Orange circles denote dusty star-forming galaxies (DSFGs) detected by ALMA at $1 < z < 3$ (Dunlop et al. 2017; Aravena et al. 2020). The red rectangle highlights the fixed M_*-SFR subsample window defined for our analysis (Chapter 4). The known dusty populations predominantly occupy the high-mass and high-SFR regime ($M_* \gtrsim 10^{10.5} M_\odot$), falling significantly outside the defined subsample window.

Chapter 7

Summary

In this thesis, we have investigated the cosmic evolution of the baryon cycle, the fundamental engine of galaxy formation, by utilizing gas-phase metallicity as a primary tracer. Through a combination of statistical analyses of local extremely metal-poor galaxies (EMPGs), deep spectroscopic studies of high-redshift galaxies using JWST, and novel empirical modeling, we have established a unified picture of chemical evolution from the early universe to the present day. The major findings and conclusions of this thesis are summarized below.

The Equilibrium State of the Local Baryon Cycle

- Our investigation began with a comprehensive search for EMPGs in the local universe, using machine learning techniques applied to SDSS imaging data, followed by spectroscopic follow-up.
- We find that the average properties of local EMPGs, particularly their position on the SFR-mass-metallicity plane, can be well explained by equilibrium models that assume a balance between gas inflows, star formation, and outflows. This suggests that the baryon cycle in the present-day universe operates, on average, in an equilibrium state.
- Despite our extensive search, we found no galaxies with metallicities below 1% solar ($Z < 0.01 Z_{\odot}$) and placed a strict upper limit on their number density. We conclude that a metallicity floor likely exists in the local universe. This is interpreted as a consequence of the cumulative enrichment of the intergalactic and circumgalactic medium over cosmic time, which prevents the formation of pristine, metal-free galaxies at the current epoch.

Breakdown of the Fundamental Metallicity Relation in the Early Universe

- To probe the early evolutionary stages of galaxies, we performed a stacking analysis of JWST spectra to derive robust metallicity measurements for galaxies across a wide redshift range ($z = 2 - 10$).
- We investigate the validity of the Fundamental Metallicity Relation (FMR) in the early universe. While the results at $z \sim 4 - 6$ depend on sample selection, we find robust evidence for a breakdown of the FMR at $z \gtrsim 8$. Galaxies at these high redshifts show systematically lower metallicities than predicted by their mass and star-formation rates. This deviation suggests that the baryon cycle in

the early universe is in a non-equilibrium state, distinct from the self-regulated state observed at later epochs.

- Despite this change in the baryon cycle regulation, our results do not necessarily imply fundamental differences in the local physics of star formation (e.g., ionization-metallicity relation and star formation law) compared to the local universe. Instead, the non-equilibrium state is characterized by a shift in the balance of gas flows.

Theoretical Constraints on the Baryon Cycle Evolution

- To interpret these observational results, we developed the ChemicalUniverseMachine, a new empirical model integrating dark matter halo growth with chemical enrichment.
- Our modeling concludes that the observed metallicity deficit at $z > 8$ cannot be explained by standard extrapolations of lower-redshift scaling relations. Reproducing the observations requires an evolution in the baryon cycle parameters, specifically either an enhancement in outflow efficiency or an increase in gas supply (inflow) relative to the star formation rate.

Collectively, this thesis delineates the cosmic evolution of galaxies from a non-equilibrium phase in the early universe, driven by efficient gas flows, to the stable equilibrium phase observed today.

Appendix

A Supplement to Chapter 5

A.1 Corner Plot

Figure A1 shows the posterior distributions of the MCMC chains for the multi-phase metallicity model.

A.2 Other Fitting Functions

We investigate alternative fitting forms for $f_{\rm H2}$ and $f_{\rm HI}$ to test the impact of different functional forms on our results, as shown in equations 1–4:

$$f_{\rm H2} + f_{\rm HI} = \frac{1}{1 + (M_h/C)^{\alpha} \times (1+z)^{\beta}} \tag{1}$$

$$\frac{f_{\rm HI}}{f_{\rm H2} + f_{\rm HI}} = \gamma \log M_h + \delta \tag{2}$$

$$\gamma = \gamma_0 + \gamma_1(1-a) + \gamma_2 \ln(1+z) \tag{3}$$

$$\delta = \delta_0 + \delta_2 \ln(1+z) \tag{4}$$

In this section, we consider the total distribution fraction (equation 1), which increases with halo mass, and use a linear relation to describe the relative distribution fraction of H I (equation 2). The figures A2 shows the best fitting function on the metallicity relations and figure A3 presents the best fitting functions of the distribution fractions. We find that the functional form does not qualitatively affect our main results: the total distribution fraction increases with halo mass, and its shape is largely independent of redshift.

A.3 Effects of Inflow Metallicity and Mergers

Figures A4–A7 show results of models with different assumptions on inflow metallicity and mergers. The best-fitting parameters of each model are shown in Table A1. The metallicity of the inflowing gas influences the metallicity of the H I at low masses at $z \sim 4$, while the occurrence of mergers plays a more significant role at higher masses. However, we do not observe any significant differences in the metallicity of H_2 or the CGM. We find that the functional form does not qualitatively affect the shape of the

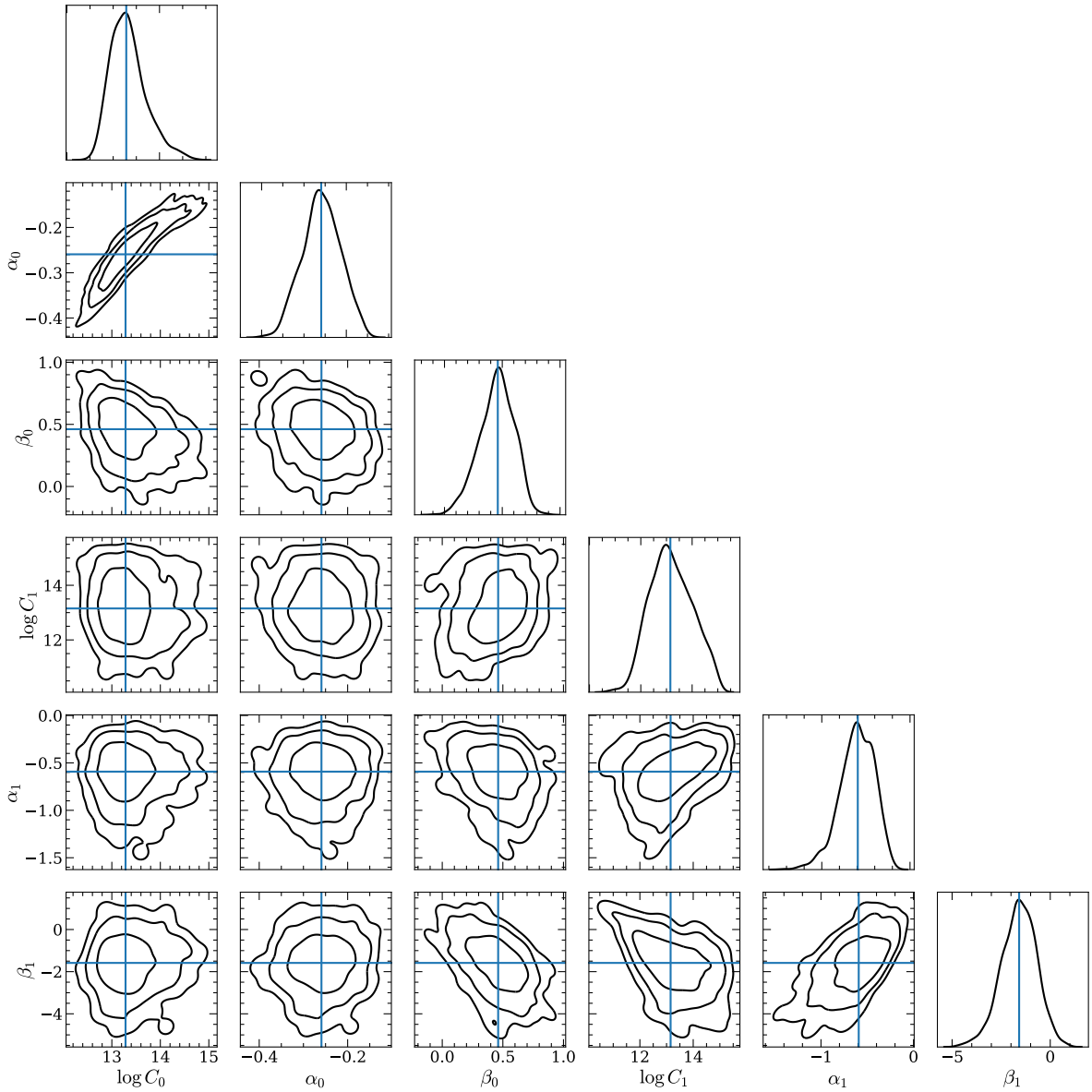

Fig. A1 Corner plot of the multi-phase ISM model presented in the main text. The blue lines indicate the best-fitting parameters. The contours represent the 1, 2, and 3 sigma levels.

ISM distribution fractions (Figures A6 and A7). As described in Section 5.3.1, the results presented in the main text are based on a model that randomly samples with equal weighting from four posterior distributions of the models: $Z_{\rm inf} = 0.01 Z_{\odot}$ with mergers, $Z_{\rm inf} = Z_{\rm CGM}$ with mergers, $Z_{\rm inf} = 0.01 Z_{\odot}$ without mergers, and $Z_{\rm inf} = Z_{\rm CGM}$ without mergers.

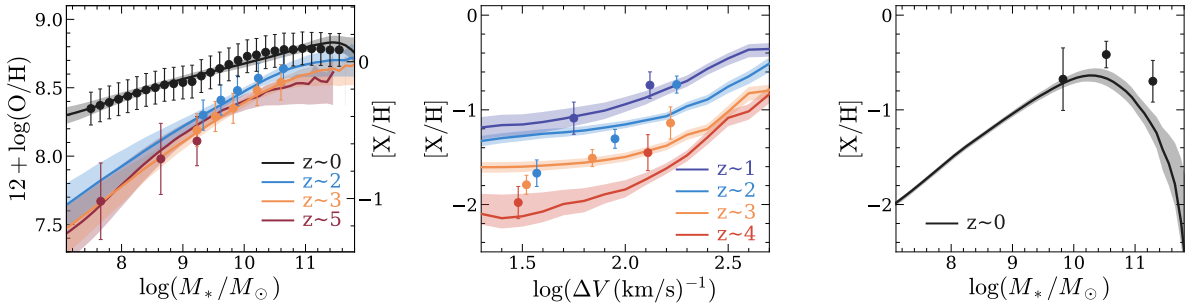

Fig. A2 Same as the right panel of Figure 5.4, but for the other functional form (Equations 1–4).

Model $Z_{\rm inf}$	mergers	$\log C_0$	α_0	β_0	$\log C_1$	α_1	β_1
0	yes	$13.3^{+0.4}_{-0.3}$	$-0.25^{+0.05}_{-0.04}$	$0.5^{+0.1}_{-0.1}$	$13.1^{+0.6}_{-0.4}$	$-0.8^{+0.2}_{-0.2}$	$-2.7^{+0.6}_{-0.7}$
0.01 $Z_\odot$	yes	$13.4^{+0.4}_{-0.3}$	$-0.25^{+0.04}_{-0.04}$	$0.4^{+0.1}_{-0.2}$	$12.5^{+0.5}_{-0.4}$	$-0.7^{+0.2}_{-0.2}$	$-1.4^{+0.7}_{-0.9}$
$Z_{\rm CGM}$	yes	$13.3^{+0.4}_{-0.3}$	$-0.25^{+0.05}_{-0.04}$	$0.5^{+0.1}_{-0.1}$	$13.4^{+0.7}_{-0.6}$	$-0.6^{+0.2}_{-0.2}$	$-1.9^{+0.7}_{-0.9}$
0	no	$13.2^{+0.4}_{-0.3}$	$-0.27^{+0.05}_{-0.04}$	$0.5^{+0.1}_{-0.1}$	$13.9^{+0.6}_{-0.5}$	$-0.6^{+0.1}_{-0.1}$	$-2.4^{+0.5}_{-0.6}$
0.01 $Z_\odot$	no	$13.2^{+0.4}_{-0.3}$	$-0.27^{+0.05}_{-0.05}$	$0.4^{+0.1}_{-0.2}$	$12.9^{+0.7}_{-0.6}$	$-0.6^{+0.1}_{-0.2}$	$-1.2^{+0.7}_{-0.8}$
$Z_{\rm CGM}$	no	$13.2^{+0.4}_{-0.3}$	$-0.27^{+0.04}_{-0.04}$	$0.5^{+0.1}_{-0.1}$	$14.0^{+0.6}_{-0.6}$	$-0.5^{+0.1}_{-0.1}$	$-1.7^{+0.7}_{-0.8}$

Table A1 Best-fitting parameters of each model.

A.4 ISM distribution fractions for the well-mixed ISM model

Figure A8 shows the ISM distribution fractions for the well-mixed ISM model. At $z \sim 0$, the ISM distribution fractions increase with halo mass, albeit with the shallow slope ($\alpha \simeq -0.05$). For redshift evolution, the ISM distribution fractions increase with redshift. Remarkably, at $z \gtrsim 1$, the ISM distribution fractions are higher compared to those at $z \sim 0$, irrespective of the halo mass, across the mass range of $10 < \log M_{\rm h}/M_\odot < 15$. This suggests that there are some physical mechanism beyond the potential well depth to retain metals in the ISM.

A.5 Effects of Stellar Yield

The stellar yield is highly sensitive to the adopted IMF and stellar evolution model, as mentioned in section 5.3.3. While our fiducial value of $y_O = 0.016$ is widely used in the literature on the mass–metallicity relation, yields lower than this value are suggested by empirical calibration studies for Milky Way stars (e.g., Weinberg et al. 2024).

To evaluate the impact of the adopted yield on model parameters and fits, we explore a lower yield value of $y_O = 0.007$, which is 2–3 times below the fiducial value. We assume that the yields of other elements are also low, along with y_O, when calculating the CGM metallicity. The best-fit models and associated parameters are

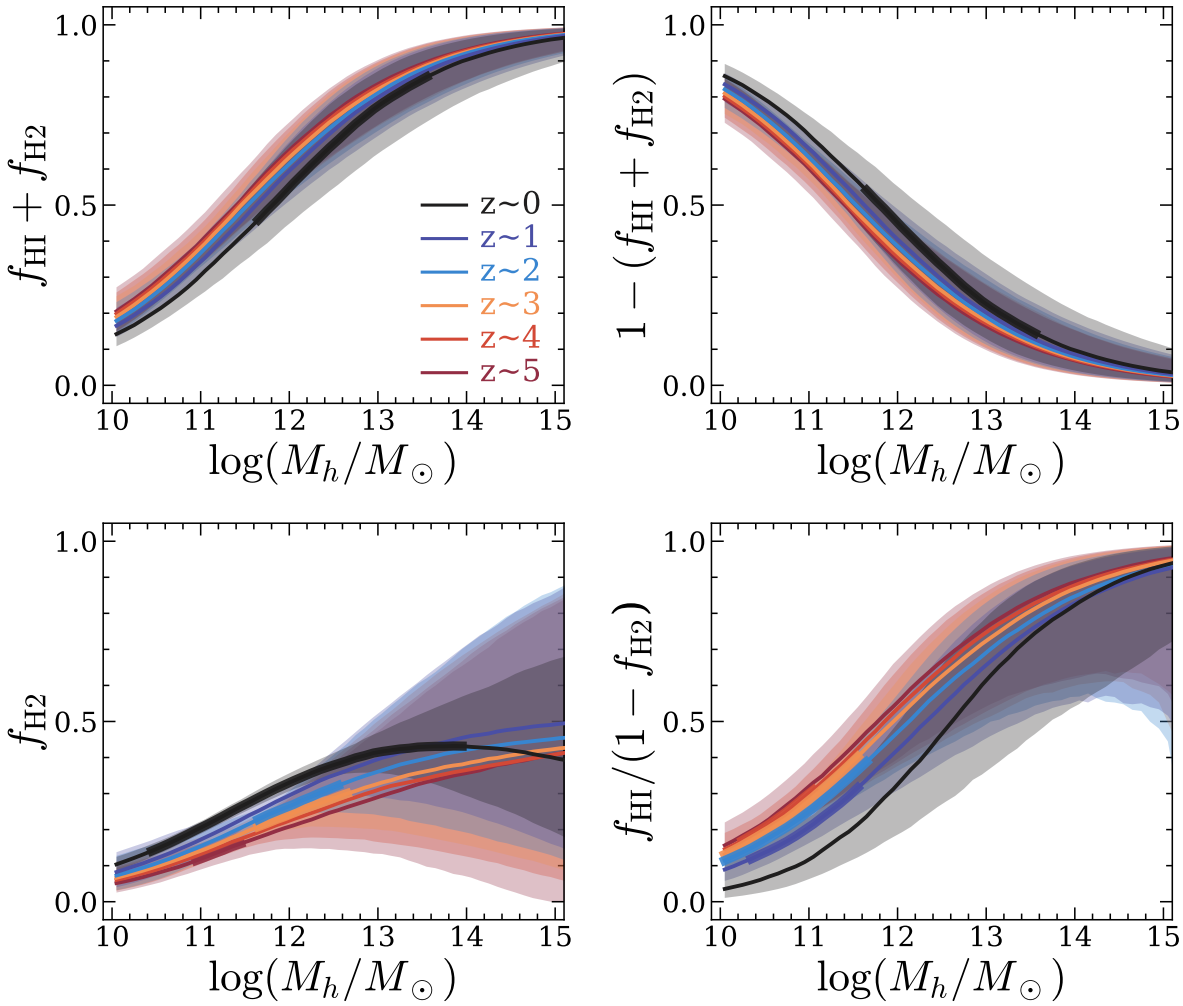

Fig. A3 Same as the Figure 5.5, but for the other functional form (Equations 1–4).

shown in Figures A9 and A10, respectively.

With this lower yield, the model provides a poorer fit to observational data, particularly for the CGM metallicity at $z \sim 0$. Since our fiducial model already predicts relatively high f-values at large $M_{\rm h}$ (see Figure 5.5), the lower yield necessitates even higher f-values, which are insufficient to enrich the CGM to observed levels. It is important to note that the low y values adopted here are those predicted for oxygen in the Milky Way stars, and it remains uncertain whether other elements exhibit similarly low values. The metallicity measurements in the CGM (Prochaska et al. 2017) are more strongly constrained by elements such as C, N, and Si than by O. The discrepancy between our model fits and observations may be attributed to non-solar relative abundances within the gas from nucleosynthesis and/or differential dust depletion.

The model also exhibits a worse fit at $z \sim 2 - 3$, which could potentially be addressed by adopting a more complex parameterization of the fitting function, where we employ a simple evolutionary formalism to reproduce the metallicity evolution with the fiducial yield (Equations 5.7 and 5.8). While a detailed discussion of stellar

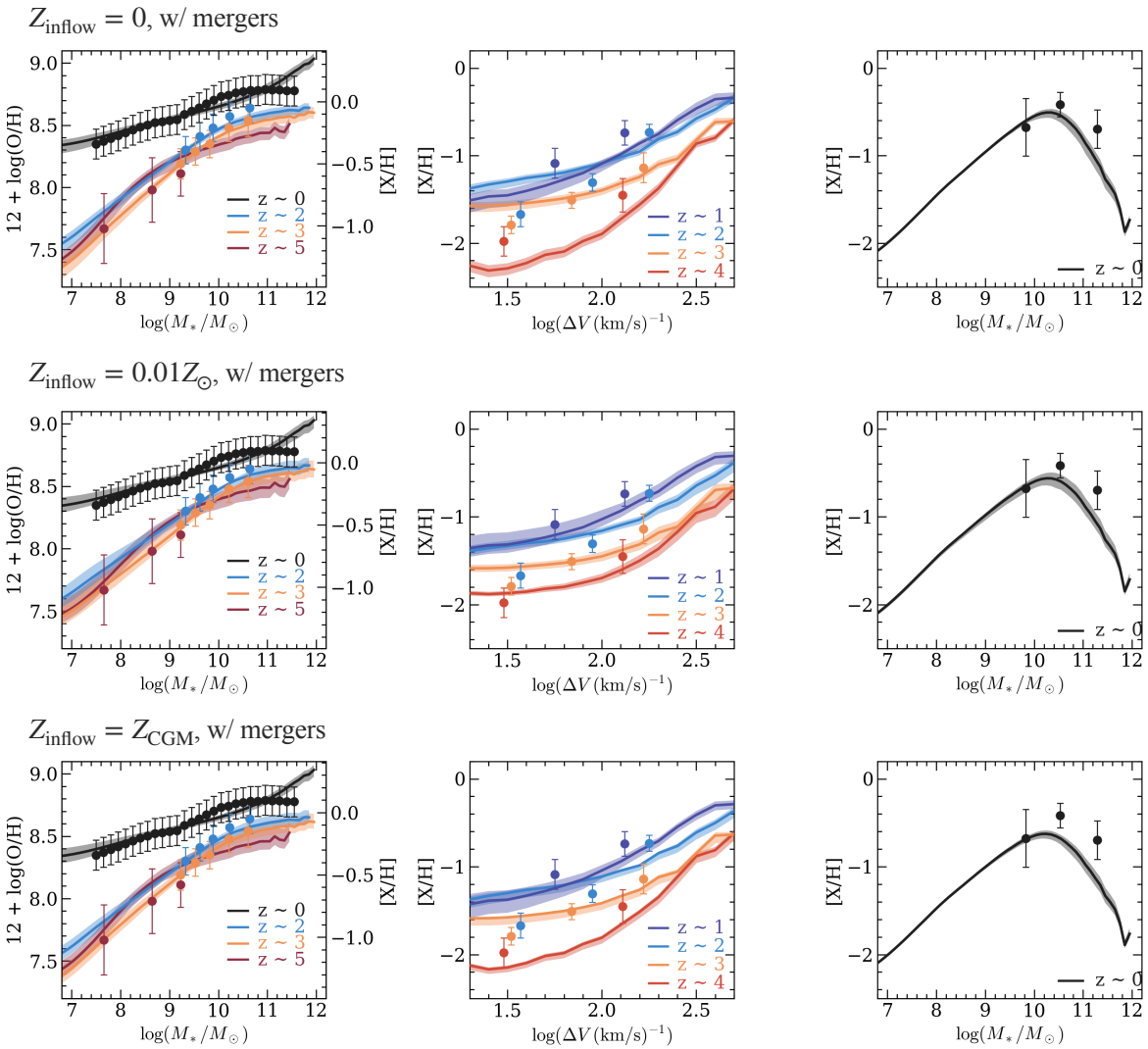

Fig. A4 Similar to Figure 5.4, but for the models with mergers, with varying the assumptions of inflow metallicity: $Z_{\rm inf} = 0$ (top), $Z_{\rm inf} = 0.01Z_{\odot}$ (middle), and $Z_{\rm inf} = Z_{\rm CGM}$ (bottom).

yields is beyond the scope of this paper, within the framework of this chemical evolution model, assumptions, and formalism, lower stellar yields do not fully account for the observed metallicity relations.

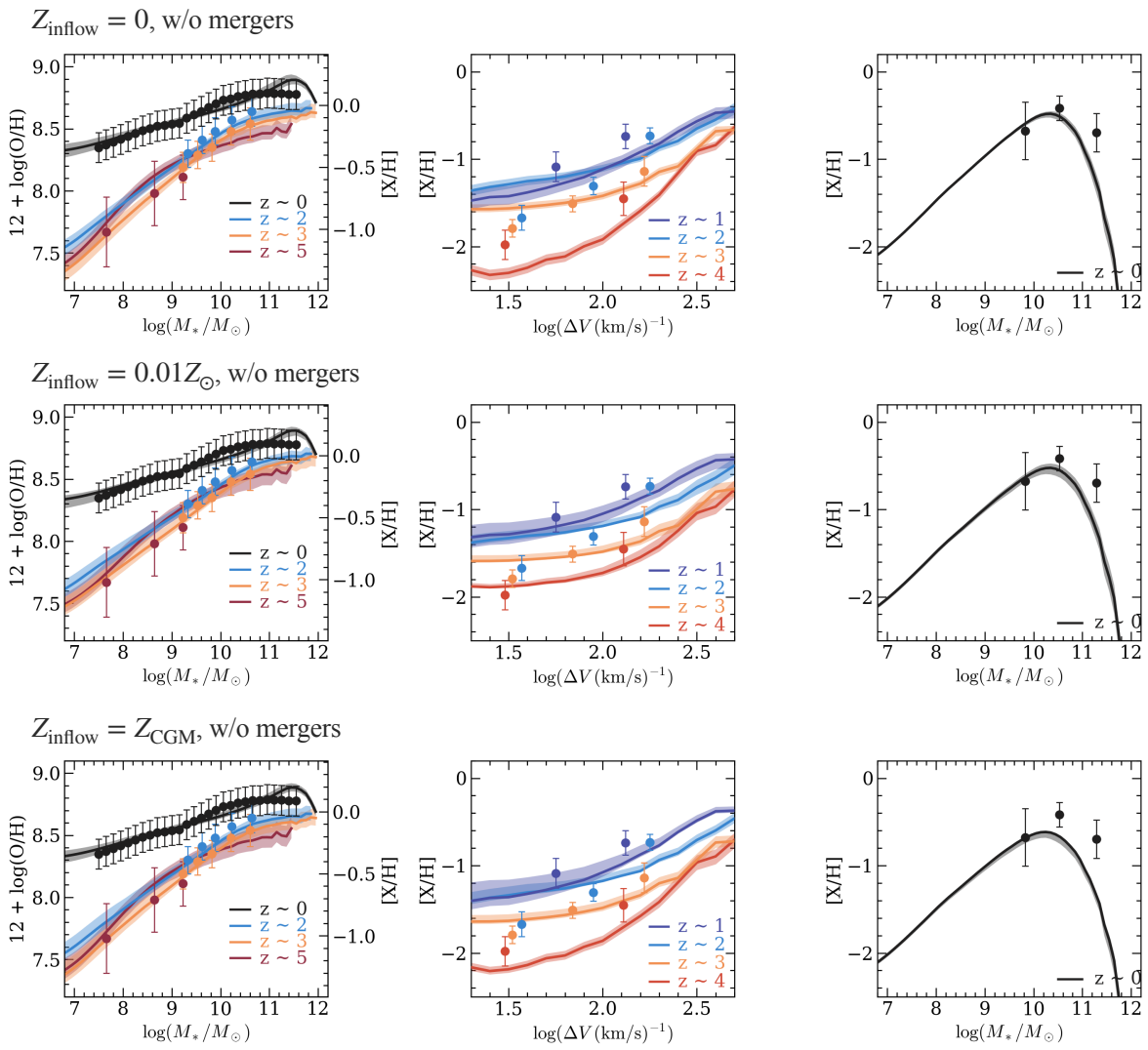

Fig. A5 Same as Figure A4, but for the models without mergers.

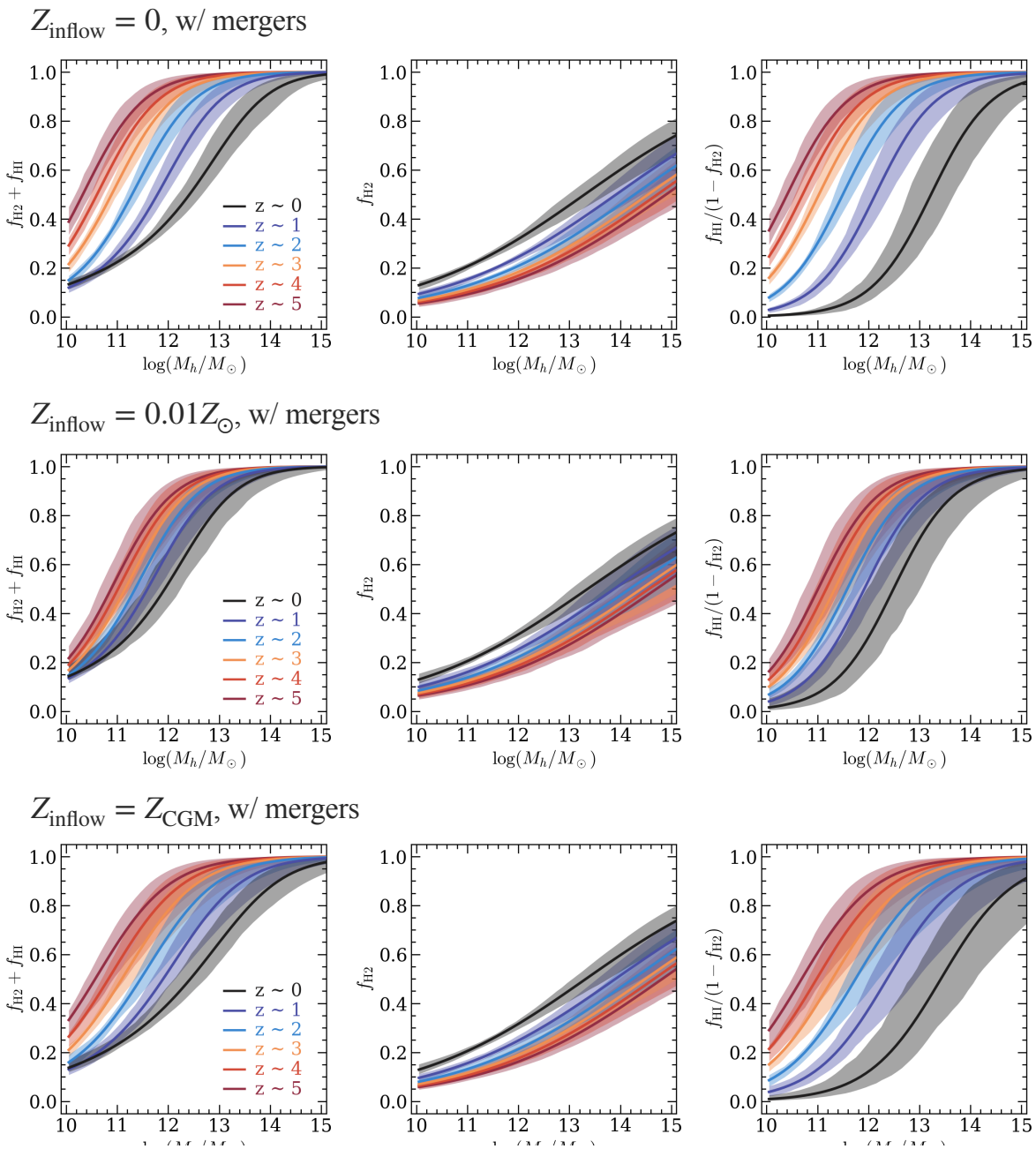

Fig. A6 Similar to Figure 5.5, but for the models with mergers, with varying the assumptions of inflow metallicity: $Z_{\rm inf} = 0$ (top), $Z_{\rm inf} = 0.01Z_{\odot}$ (middle), and $Z_{\rm inf} = Z_{\rm CGM}$ (bottom).

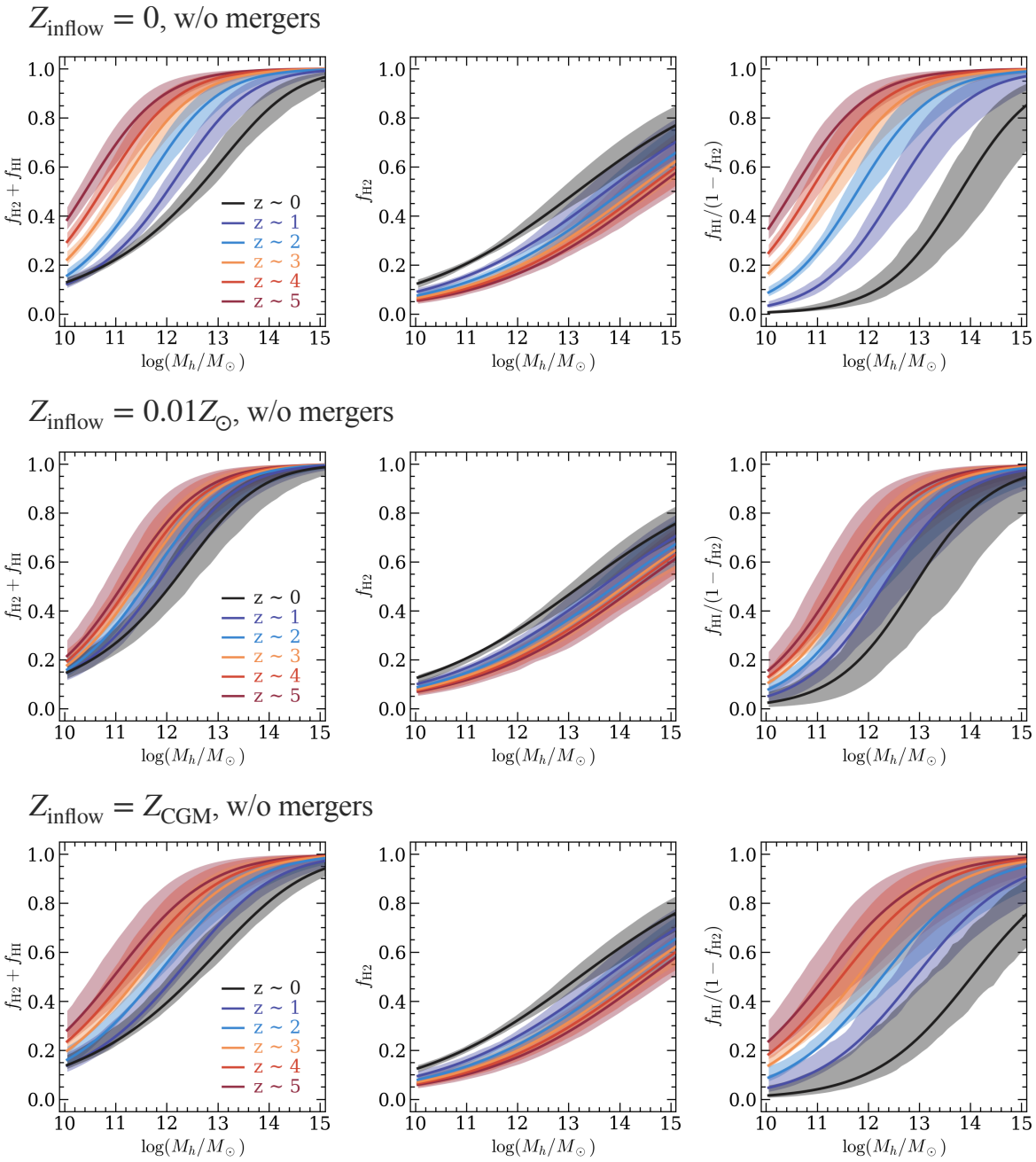

Fig. A7 Same as Figure A6, but for the models without mergers.

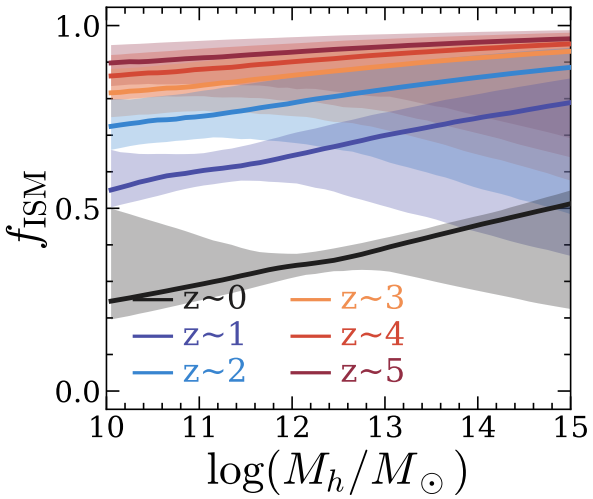

Fig. A8 The ISM distribution fraction of the well-mixed ISM model. The shaded regions represent 68th percentile of the posterior distributions.

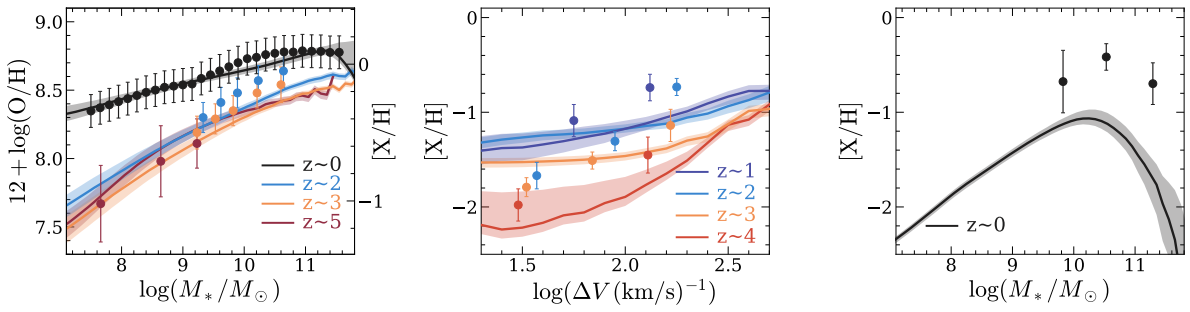

Fig. A9 Same as the right panel of Figure 5.4, but for the lower yield value of $y_O = 0.007$.

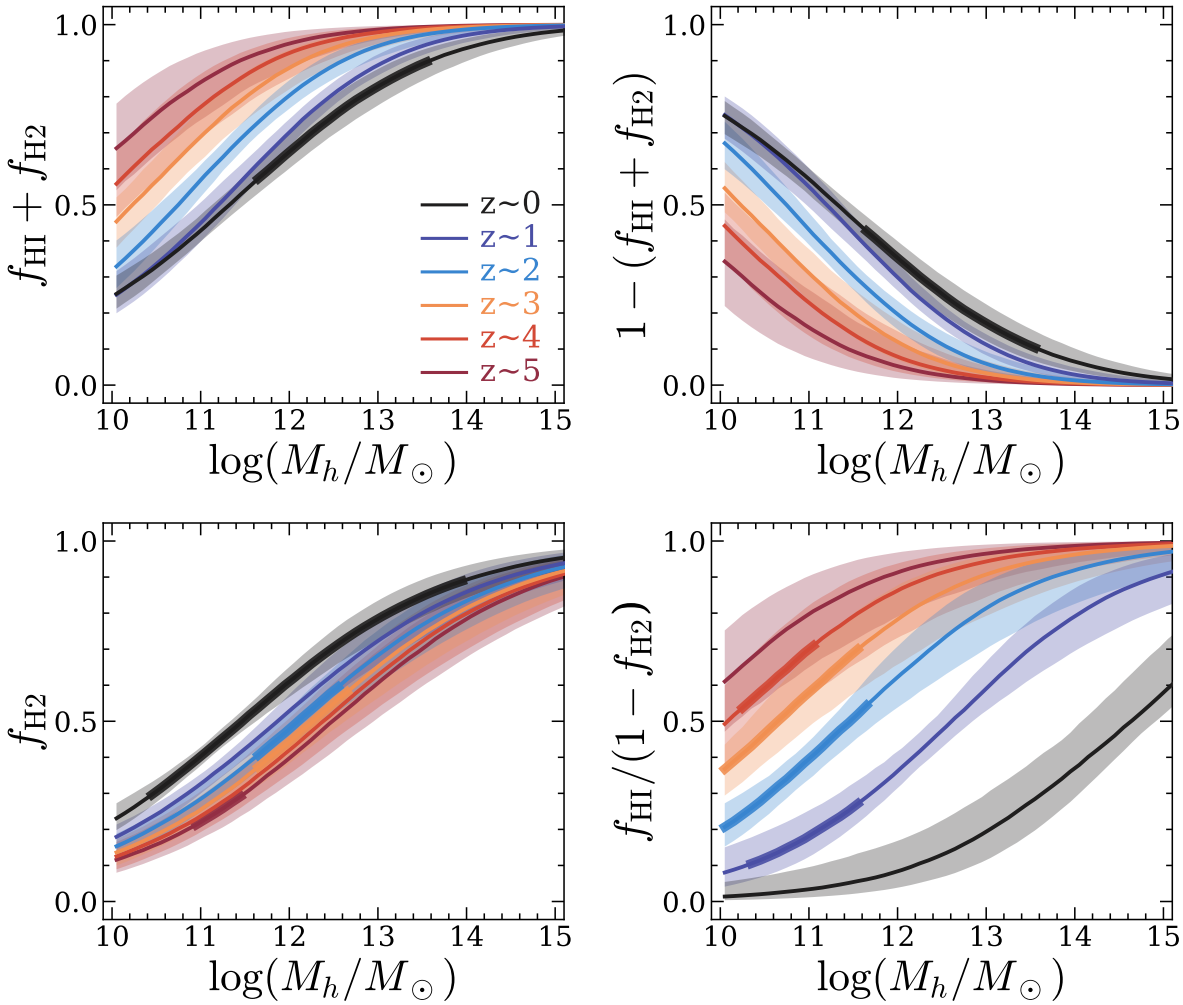

Fig. A10 Same as the Figure 5.5, but for the lower yield value of $y_O = 0.007$.

Reference

Ahumada, R., Prieto, C. A., Almeida, A., et al. 2020, ApJS, 249, 3
Algera, H. S. B., Inami, H., Oesch, P. A., et al. 2023, MNRAS, 518, 6142
Aller, L. H. 1984, Physics of thermal gaseous nebulae, doi:10.1007/978-94-010-9639-3
Álvarez-Márquez, J., Crespo Gómez, A., Colina, L., et al. 2025, A&A, 695, A250
Andrews, B. H., & Martini, P. 2013, ApJ, 765, 140
Annibali, F., La Torre, V., Tosi, M., et al. 2019, MNRAS, 482, 3892
Applebaum, E., Brooks, A. M., Christensen, C. R., et al. 2021, ApJ, 906, 96
Aravena, M., Boogaard, L., Gónzalez-López, J., et al. 2020, ApJ, 901, 79
Asplund, M., Amarsi, A. M., & Grevesse, N. 2021, A&A, 653, A141
Asplund, M., Grevesse, N., Sauval, A. J., & Scott, P. 2009, ARA&A, 47, 481
Bañados, E., Rauch, M., Decarli, R., et al. 2019, ApJ, 885, 59
Baker, W. M., Maiolino, R., Belfiore, F., et al. 2023, MNRAS, 519, 1149
Baldwin, J. A., Phillips, M. M., & Terlevich, R. 1981, PASP, 93, 5
Begley, R., Cullen, F., McLure, R. J., et al. 2024, MNRAS, 527, 4040
Behroozi, P., Wechsler, R. H., Hearin, A. P., & Conroy, C. 2019, MNRAS, 488, 3143
Behroozi, P. S., Wechsler, R. H., & Conroy, C. 2013a, ApJ, 770, 57
Behroozi, P. S., Wechsler, R. H., & Wu, H.-Y. 2013b, ApJ, 762, 109
Behroozi, P. S., Wechsler, R. H., Wu, H.-Y., et al. 2013c, ApJ, 763, 18
Berg, D. A., Erb, D. K., Henry, R. B. C., Skillman, E. D., & McQuinn, K. B. W. 2019, ApJ, 874, 93
Berg, T. A. M., Fumagalli, M., D'Odorico, V., et al. 2021, MNRAS, 502, 4009
Bian, F., Kewley, L. J., & Dopita, M. A. 2018, ApJ, 859, 175
Bird, S., Haehnelt, M., Neeleman, M., et al. 2015, MNRAS, 447, 1834
Bird, S., Vogelsberger, M., Haehnelt, M., et al. 2014, MNRAS, 445, 2313
Bolatto, A. D., Wolfire, M., & Leroy, A. K. 2013, ARA&A, 51, 207
Bouwens, R., González-López, J., Aravena, M., et al. 2020, ApJ, 902, 112
Brinchmann, J. 2023, MNRAS, 525, 2087
Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000
Bryan, G. L., & Norman, M. L. 1998, ApJ, 495, 80
Bunker, A. J., Cameron, A. J., Curtis-Lake, E., et al. 2024, A&A, 690, A288
Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, ApJ, 533, 682
Cameron, A. J., Saxena, A., Bunker, A. J., et al. 2023, A&A, 677, A115
Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, ApJ, 345, 245
Castelli, F., & Hubrig, S. 2004, A&A, 425, 263
Chabrier, G. 2003, PASP, 115, 763
Chevallard, J., & Charlot, S. 2016, MNRAS, 462, 1415
Choi, J., Dotter, A., Conroy, C., et al. 2016, ApJ, 823, 102

Coe, D., Salmon, B., Bradač, M., et al. 2019, ApJ, 884, 85
Conroy, C. 2013, ARA&A, 51, 393
Conroy, C., & Gunn, J. E. 2010, ApJ, 712, 833
Conroy, C., Gunn, J. E., & White, M. 2009, ApJ, 699, 486
Curti, M., Cresci, G., Mannucci, F., et al. 2017, MNRAS, 465, 1384
Curti, M., Mannucci, F., Cresci, G., & Maiolino, R. 2020, MNRAS, 491, 944
Curti, M., D'Eugenio, F., Carniani, S., et al. 2023, MNRAS, 518, 425
Curti, M., Maiolino, R., Curtis-Lake, E., et al. 2024, A&A, 684, A75
Davé, R., Finlator, K., & Oppenheimer, B. D. 2012, MNRAS, 421, 98
De Cia, A., Ledoux, C., Petitjean, P., & Savaglio, S. 2018, A&A, 611, A76
Dessauges-Zavadsky, M., Ginolfi, M., Pozzi, F., et al. 2020, A&A, 643, A5
D'Eugenio, F., Cameron, A. J., Scholtz, J., et al. 2025, ApJS, 277, 4
Diego, J. M., Adams, N. J., Willner, S. P., et al. 2024, A&A, 690, A114
Dunlop, J. S., McLure, R. J., Biggs, A. D., et al. 2017, MNRAS, 466, 861
Eisenstein, D. J., Willott, C., Alberts, S., et al. 2023, arXiv e-prints, arXiv:2306.02465
Eldridge, J. J., Stanway, E. R., Xiao, L., et al. 2017, **??**jnlPASA, 34, e058
Ellison, S. L., Patton, D. R., Simard, L., & McConnachie, A. W. 2008, ApJ, 672, L107
Erb, D. K., Shapley, A. E., Pettini, M., et al. 2006, ApJ, 644, 813
Esteban, C., García-Rojas, J., Carigi, L., et al. 2014, MNRAS, 443, 624
Ferland, G. J., Porter, R. L., van Hoof, P. A. M., et al. 2013, Rev. Mex. Astron. Astrofis., 49, 137
Ferland, G. J., Chatzikos, M., Guzmán, F., et al. 2017, Rev. Mex. Astron. Astrofis., 53, 385
Finkelstein, S. L., Dickinson, M., Ferguson, H. C., et al. 2017, The Cosmic Evolution Early Release Science (CEERS) Survey, JWST Proposal ID 1345. Cycle 0 Early Release Science
Finkelstein, S. L., Bagley, M. B., Arrabal Haro, P., et al. 2022a, arXiv e-prints, arXiv:2207.12474
Finkelstein, S. L., Bagley, M. B., Ferguson, H. C., et al. 2022b, arXiv e-prints, arXiv:2211.05792
Finlator, K., & Davé, R. 2008, MNRAS, 385, 2181
Foreman, M. R., & Vollmer, F. 2013, New Journal of Physics, 15, 083006
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013a, PASP, 125, 306
—. 2013b, PASP, 125, 306
Fu, S., Sun, F., Jiang, L., et al. 2025, ApJ, 987, 186
Fukugita, M., Ichikawa, T., Gunn, J. E., et al. 1996, AJ, 111, 1748
Garn, T., & Best, P. N. 2010, MNRAS, 409, 421
Garnett, D. R. 1992, AJ, 103, 1330
Gordon, K. D., Clayton, G. C., Misselt, K. A., Landolt, A. U., & Wolff, M. J. 2003, ApJ, 594, 279
Guo, H., Wang, J., Jones, M. G., & Behroozi, P. 2023, ApJ, 955, 57
Guseva, N. G., Izotov, Y. I., Papaderos, P., & Fricke, K. J. 2007, A&A, 464, 885
Gutkin, J., Charlot, S., & Bruzual, G. 2016, MNRAS, 462, 1757
Harada, N., Yajima, H., & Abe, M. 2023, MNRAS, 525, 5868
Harikane, Y., Ouchi, M., Shibuya, T., et al. 2018, ApJ, 859, 84

Harikane, Y., Ouchi, M., Oguri, M., et al. 2022, arXiv e-prints, arXiv:2208.01612
—. 2023a, ApJS, 265, 5
Harikane, Y., Zhang, Y., Nakajima, K., et al. 2023b, ApJ, 959, 39
Hasegawa, K., & Semelin, B. 2013, MNRAS, 428, 154
Hayashi, K., Hirai, Y., Chiba, M., & Ishiyama, T. 2022, arXiv e-prints, arXiv:2206.02821
Hayes, M., Schaerer, D., Östlin, G., et al. 2011, ApJ, 730, 8
Haynes, M. P., Giovanelli, R., Kent, B. R., et al. 2018, ApJ, 861, 49
Heckman, T. M., Robert, C., Leitherer, C., Garnett, D. R., & van der Rydt, F. 1998, ApJ, 503, 646
Hirschmann, M., Charlot, S., & Somerville, R. S. 2023, MNRAS, 526, 3504
Hsiao, T. Y.-Y., Abdurro'uf, Coe, D., et al. 2024, ApJ, 973, 8
Inoue, A. K. 2011, MNRAS, 415, 2920
Isobe, Y., Ouchi, M., Kojima, T., et al. 2021, ApJ, 918, 54
Isobe, Y., Ouchi, M., Suzuki, A., et al. 2022a, ApJ, 925, 111
Isobe, Y., Ouchi, M., Nakajima, K., et al. 2022b, arXiv e-prints, arXiv:2206.04709
Izotov, Y. I., Stasińska, G., Meynet, G., Guseva, N. G., & Thuan, T. X. 2006, A&A, 448, 955
Izotov, Y. I., & Thuan, T. X. 1998, ApJ, 500, 188
Izotov, Y. I., Thuan, T. X., & Guseva, N. G. 2012, A&A, 546, A122
—. 2019, MNRAS, 483, 5491
Jenkins, E. B. 1987, in Interstellar Processes, ed. D. J. Hollenbach & H. A. Thronson, Jr., Vol. 134, 533
Johnson, B. D., Leja, J., Conroy, C., & Speagle, J. S. 2021, ApJS, 254, 22
Kashikawa, N., Aoki, K., Asai, R., et al. 2002, PASJ, 54, 819
Kauffmann, G., Heckman, T. M., Tremonti, C., et al. 2003, MNRAS, 346, 1055
Kawamata, R., Ishigaki, M., Shimasaku, K., et al. 2018, ApJ, 855, 4
Ke, G., Meng, Q., Finley, T., et al. 2017, in Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17 (Red Hook, NY, USA: Curran Associates Inc.), 3149–3157
Kennicutt, Jr., R. C. 1998, ARA&A, 36, 189
Kewley, L. J., Dopita, M. A., Sutherland, R. S., Heisler, C. A., & Trevena, J. 2001, ApJ, 556, 121
Kewley, L. J., & Ellison, S. L. 2008, ApJ, 681, 1183
Kewley, L. J., Maier, C., Yabe, K., et al. 2013, ApJ, 774, L10
Kirby, E. N., Cohen, J. G., Guhathakurta, P., et al. 2013, ApJ, 779, 102
Klypin, A., Yepes, G., Gottlöber, S., Prada, F., & Heß, S. 2016, MNRAS, 457, 4340
Kniazev, A. Y., Grebel, E. K., Hao, L., et al. 2003, ApJ, 593, L73
Kojima, T., Ouchi, M., Rauch, M., et al. 2020, ApJ, 898, 142
Kravtsov, A. V., & Klypin, A. A. 1999, ApJ, 520, 437
Kravtsov, A. V., Klypin, A. A., & Khokhlov, A. M. 1997, ApJS, 111, 73
Kriek, M., Shapley, A. E., Reddy, N. A., et al. 2015, ApJS, 218, 15
Kroupa, P. 2001, MNRAS, 322, 231
Kurita, M., Kino, M., Iwamuro, F., et al. 2020, PASJ, 72, 48
Laseter, I. H., Maseda, M. V., Curti, M., et al. 2024, A&A, 681, A70
Ledoux, C., Petitjean, P., Fynbo, J. P. U., Møller, P., & Srianand, R. 2006, A&A,

457, 71
Lehner, N., Howk, J. C., Tripp, T. M., et al. 2013, ApJ, 770, 138
Lequeux, J., Peimbert, M., Rayo, J. F., Serrano, A., & Torres-Peimbert, S. 1979, A&A, 80, 155
Lilly, S. J., Carollo, C. M., Pipino, A., Renzini, A., & Peng, Y. 2013, ApJ, 772, 119
Luridiana, V., Morisset, C., & Shaw, R. A. 2015, A&A, 573, A42
Ma, X., Hopkins, P. F., Faucher-Giguère, C.-A., et al. 2016, MNRAS, 456, 2140
Ma, Z., Sun, B., Cheng, C., et al. 2024, ApJ, 975, 87
Madau, P. 1995, ApJ, 441, 18
Madau, P., & Dickinson, M. 2014, ARA&A, 52, 415
Maiolino, R., & Mannucci, F. 2019, A&A Rev., 27, 3
Maiolino, R., Scholtz, J., Curtis-Lake, E., et al. 2024, A&A, 691, A145
Mannucci, F., Cresci, G., Maiolino, R., Marconi, A., & Gnerucci, A. 2010, MNRAS, 408, 2115
Marshall, J. L., Burles, S., Thompson, I. B., et al. 2008, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 7014, Ground-based and Airborne Instrumentation for Astronomy II, ed. I. S. McLean & M. M. Casali, 701454
Matsubayashi, K., Ohta, K., Iwamuro, F., et al. 2019, PASJ, 71, 102
Matthee, J., Mackenzie, R., Simcoe, R. A., et al. 2022, arXiv e-prints, arXiv:2211.08255
McKee, C. F., & Ostriker, J. P. 1977, ApJ, 218, 148
McQuinn, K. B. W., Berg, D. A., Skillman, E. D., et al. 2020, ApJ, 891, 181
Møller, P., Fynbo, J. P. U., Ledoux, C., & Nilsson, K. K. 2013, MNRAS, 430, 2680
Nakajima, K., & Maiolino, R. 2022, MNRAS, 513, 5134
Nakajima, K., Ouchi, M., Isobe, Y., et al. 2023, arXiv e-prints, arXiv:2301.12825
Nakajima, K., Ouchi, M., Shimasaku, K., et al. 2013, ApJ, 769, 3
Nakajima, K., Ouchi, M., Xu, Y., et al. 2022, ApJS, 262, 3
Nakajima, K., Ouchi, M., Isobe, Y., et al. 2024, arXiv e-prints, arXiv:2412.04541
Nakajima, K., Ouchi, M., Harikane, Y., et al. 2025, arXiv e-prints, arXiv:2506.11846
Neeleman, M., Kanekar, N., Prochaska, J. X., et al. 2017, Science, 355, 1285
Nishigaki, M., Behroozi, P., Ouchi, M., et al. 2025, ApJ, 984, 135
Oguri, M. 2010, PASJ, 62, 1017
Oguri, M., Miyazaki, S., Li, X., et al. 2021, PASJ, 73, 817
Oke, J. B., & Gunn, J. E. 1983, ApJ, 266, 713
Ono, Y., Ouchi, M., Shimasaku, K., et al. 2010, ApJ, 724, 1524
Onodera, M., Carollo, C. M., Lilly, S., et al. 2016, ApJ, 822, 42
Osterbrock, D. E., & Ferland, G. J. 2006, Astrophysics of gaseous nebulae and active galactic nuclei
Peeples, M. S., & Shankar, F. 2011, MNRAS, 417, 2962
Peeples, M. S., Werk, J. K., Tumlinson, J., et al. 2014, ApJ, 786, 54
Pérez-Montero, E. 2014, MNRAS, 441, 2663
Péroux, C., & Howk, J. C. 2020, ARA&A, 58, 363
Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, A&A, 594, A13
Pontoppidan, K. M., Barrientes, J., Blome, C., et al. 2022, ApJ, 936, L14
Popesso, P., Concas, A., Cresci, G., et al. 2023, MNRAS, 519, 1526

Prochaska, J., Hennawi, J., Westfall, K., et al. 2020, The Journal of Open Source Software, 5, 2308
Prochaska, J. X., Werk, J. K., Worseck, G., et al. 2017, ApJ, 837, 169
Rafelski, M., Wolfe, A. M., Prochaska, J. X., Neeleman, M., & Mendez, A. J. 2012, ApJ, 755, 89
Rahmati, A., & Schaye, J. 2014, MNRAS, 438, 529
Recchi, S., Matteucci, F., & D'Ercole, A. 2001, MNRAS, 322, 800
Rieke, M. J., Robertson, B., Tacchella, S., et al. 2023, ApJS, 269, 16
Rodríguez-Puebla, A., Behroozi, P., Primack, J., et al. 2016, MNRAS, 462, 893
Saintonge, A., Catinella, B., Tacconi, L. J., et al. 2017, ApJS, 233, 22
Salpeter, E. E. 1955, ApJ, 121, 161
Sánchez Almeida, J., Morales-Luis, A. B., Muñoz-Tuñón, C., et al. 2014, ApJ, 783, 45
Sánchez Almeida, J., Pérez-Montero, E., Morales-Luis, A. B., et al. 2016, ApJ, 819, 110
Sánchez Almeida, J., Elmegreen, B. G., Muñoz-Tuñón, C., et al. 2015, ApJ, 810, L15
Sanders, R. L., Shapley, A. E., Topping, M. W., Reddy, N. A., & Brammer, G. B. 2024, ApJ, 962, 24
Sanders, R. L., Shapley, A. E., Reddy, N. A., et al. 2020, MNRAS, 491, 1427
Sanders, R. L., Shapley, A. E., Jones, T., et al. 2021, ApJ, 914, 19
—. 2023, ApJ, 942, 24
Schaerer, D. 2003, A&A, 397, 527
Schaerer, D., Ginolfi, M., Béthermin, M., et al. 2020, A&A, 643, A3
Schaye, J., Crain, R. A., Bower, R. G., et al. 2015, MNRAS, 446, 521
Schlafly, E. F., & Finkbeiner, D. P. 2011, ApJ, 737, 103
Scholtz, J., Silcock, M. S., Curtis-Lake, E., et al. 2025, arXiv e-prints, arXiv:2507.17809
Selsing, J., Fynbo, J. P. U., Christensen, L., & Krogager, J. K. 2016, A&A, 585, A87
Simcoe, R. A., Sargent, W. L. W., & Rauch, M. 2004, ApJ, 606, 92
Somerville, R. S., & Davé, R. 2015, ARA&A, 53, 51
Somerville, R. S., Hopkins, P. F., Cox, T. J., Robertson, B. E., & Hernquist, L. 2008, MNRAS, 391, 481
Stanway, E. R., & Eldridge, J. J. 2018, MNRAS, 479, 75
Stark, D. P., Ellis, R. S., Charlot, S., et al. 2017, MNRAS, 464, 469
Sugahara, Y., Ouchi, M., Harikane, Y., et al. 2019, ApJ, 886, 29
Thuan, T. X., & Izotov, Y. I. 2005, ApJS, 161, 240
Topping, M. W., Stark, D. P., Senchyna, P., et al. 2024, MNRAS, 529, 3301
Torrey, P., Vogelsberger, M., Marinacci, F., et al. 2019, MNRAS, 484, 5587
Tortora, C., Hunt, L. K., & Ginolfi, M. 2022, A&A, 657, A19
Tremonti, C. A., Heckman, T. M., Kauffmann, G., et al. 2004, ApJ, 613, 898
Treu, T., Roberts-Borsani, G., Bradac, M., et al. 2022, ApJ, 935, 110
Treu, T. L., Abramson, L. E., Bradac, M., et al. 2017, Through the Looking GLASS: A JWST Exploration of Galaxy Formation and Evolution from Cosmic Dawn to Present Day, JWST Proposal ID 1324. Cycle 0 Early Release Science
Tumlinson, J., Peeples, M. S., & Werk, J. K. 2017, ARA&A, 55, 389
Tumlinson, J., Thom, C., Werk, J. K., et al. 2011, Science, 334, 948

—. 2013, ApJ, 777, 59
Villaescusa-Navarro, F., Genel, S., Castorina, E., et al. 2018, ApJ, 866, 135
Vincenzo, F., Matteucci, F., Belfiore, F., & Maiolino, R. 2016, MNRAS, 455, 4183
Vogelsberger, M., Genel, S., Springel, V., et al. 2014, MNRAS, 444, 1518
Walter, F., Brinks, E., de Blok, W. J. G., et al. 2008, AJ, 136, 2563
Walter, F., Carilli, C., Neeleman, M., et al. 2020, ApJ, 902, 111
Weinberg, D. H., Griffith, E. J., Johnson, J. W., & Thompson, T. A. 2024, ApJ, 973, 122
Werk, J. K., Prochaska, J. X., Thom, C., et al. 2012, ApJS, 198, 3
Wheeler, C., Hopkins, P. F., Pace, A. B., et al. 2019, MNRAS, 490, 4447
Williams, C. C., Tacchella, S., Maseda, M. V., et al. 2023, ApJS, 268, 64
Willott, C. J., Desprez, G., Asada, Y., et al. 2024, ApJ, 966, 74
Wise, J. H., Turk, M. J., Norman, M. L., & Abel, T. 2012, ApJ, 745, 50
Wolfe, A. M., Gawiser, E., & Prochaska, J. X. 2005, ARA&A, 43, 861
Wotta, C. B., Lehner, N., Howk, J. C., O'Meara, J. M., & Prochaska, J. X. 2016, ApJ, 831, 95
Xu, Y., Ouchi, M., Rauch, M., et al. 2022, ApJ, 929, 134
Zafar, T., Péroux, C., Popping, A., et al. 2013, A&A, 556, A141
Zahid, H. J., Kashino, D., Silverman, J. D., et al. 2014, ApJ, 792, 75
Zavala, J. A., Casey, C. M., Manning, S. M., et al. 2021, ApJ, 909, 165
Zavala, J. A., Bakx, T., Mitsuhashi, I., et al. 2024, ApJ, 977, L9