



# NGC 147 Corroborates the Break in the Stellar Mass–Stellar Metallicity Relation for Galaxies

Zhuyun Zhuang<sup>1</sup> , Evan N. Kirby<sup>1</sup> , Nicha Leethochawalit<sup>2,3,4</sup> , and Mithi A. C. de los Reyes<sup>1</sup>

<sup>1</sup> Department of Astronomy, California Institute of Technology, 1200 E. California Blvd., CA 91125, Pasadena, CA 91125, USA; [zzhuang@astro.caltech.edu](mailto:zzhuang@astro.caltech.edu)

<sup>2</sup> School of Physics, The University of Melbourne, Parkville, VIC 3010, Australia

<sup>3</sup> ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Australia

<sup>4</sup> National Astronomical Research Institute of Thailand (NARIT), MaeRim, Chiang Mai, 50180, Thailand

Received 2021 May 28; revised 2021 July 2; accepted 2021 July 8; published 2021 October 13

## Abstract

The stellar mass–stellar metallicity relation (MZR) is an essential approach to probing the chemical evolution of galaxies. It reflects the balance between galactic feedback and gravitational potential as a function of stellar mass. However, the current MZR of local dwarf satellite galaxies ( $M_* \lesssim 10^8 M_\odot$ ; measured from resolved stellar spectroscopy) may not be reconcilable with that of more massive galaxies ( $M_* \gtrsim 10^{9.5} M_\odot$ ; measured from integrated-light spectroscopy). Such a discrepancy may result from a systematic difference between the two methods, or it may indicate a break in the MZR around  $10^9 M_\odot$ . To address this question, we measured the stellar metallicity of NGC 147 from integrated light using the Palomar Cosmic Web Imager. We compared the stellar metallicity estimates from integrated light with measurements from resolved stellar spectroscopy and found them to be consistent within 0.1 dex. On the other hand, the high-mass MZR overpredicts the metallicity by 0.6 dex at the mass of NGC 147. Therefore, our results tentatively suggest that the discrepancy between the low-mass MZR and high-mass MZR should not be attributed to a systematic difference in techniques. Instead, real physical processes cause the transition in the MZR. In addition, we discovered a positive age gradient in the innermost region and a negative metallicity gradient from the resolved stars at larger radii, suggesting a possible outside-in formation of NGC 147.

*Unified Astronomy Thesaurus concepts:* Local Group (929); Galaxy abundances (574); Metallicity (1031); Chemical abundances (224); Stellar abundances (1577); Dwarf elliptical galaxies (415)

*Supporting material:* machine-readable table

## 1. Introduction

In the past few decades, numerous works have revealed the tight correlation between the stellar masses (luminosities) and the metallicities of galaxies, i.e., the stellar mass–metallicity relation (MZR) (e.g., McClure & Bergh 1968; Lequeux et al. 1979; Tremonti et al. 2004; Lee et al. 2006; Kirby et al. 2013; Leethochawalit et al. 2018). The observed MZR indicates that more massive galaxies are more metal-rich than less massive ones. The trend can be interpreted as a result of metal retention, wherein high-mass galaxies with deep gravitational potential wells can resist galactic winds and retain more metals (Dekel & Silk 1986). Recent works have argued metallicity may also be regulated by star formation efficiency (e.g., Calura et al. 2009; Magrini et al. 2012) as well as by the interplay between inflow, outflow, and enrichment (Finlator & Davé 2008). Generally, galactic metallicities are measured in two forms: gas-phase and stellar-phase. While gas-phase metallicity reflects the metals in the interstellar medium (ISM) at the time of observation, stellar metallicity indicates the amount of metals incorporated into the stars at the time they formed, which is less susceptible to instantaneous fluctuations and represents the metal abundance averaged over the star formation history (SFH). Therefore, the stellar MZR is a more stable indicator of the chemical evolution of galaxies than the gas-phase MZR.

The MZR<sup>5</sup> at  $z \sim 0$  can be measured from two independent techniques. Local dwarf galaxies can be resolved into stars, so

their stellar metallicities are derived from spectroscopy of individual stars (e.g., Kirby et al. 2013). On the other hand, more massive galaxies are too far for us to resolve the stars. Therefore, metallicities are obtained from the integrated-light spectra of whole galaxies (e.g., Gallazzi et al. 2005; Zahid et al. 2013; Leethochawalit et al. 2018, 2019). Although the low-mass and high-mass MZRs both indicate more massive galaxies are more metal-rich, they follow different forms of expression. Kirby et al. (2013, hereafter K13) used Geha et al.’s (2010) Keck/DEIMOS spectra to derive the low-mass MZR from resolved stellar spectroscopy in Local Group dwarf galaxies ( $M_* \lesssim 10^8 M_\odot$ )<sup>6</sup> as  $[\text{Fe}/\text{H}] \propto M_*^{0.30 \pm 0.02}$ , while Leethochawalit et al. (2019, hereafter L19) measured the high-mass MZR from integrated-light spectra of local massive quiescent galaxies ( $M_* \gtrsim 10^{9.5} M_\odot$ ) as  $[\text{Fe}/\text{H}] \propto M_*^{0.11 \pm 0.02}$ . Choi et al. (2014) similarly observed a flat slope in the high-mass MZR, where the metallicities for massive quiescent galaxies ( $M_* \geq 10^{10.6} M_\odot$ ) in the local universe were measured from stacked spectra in bins of stellar mass. Sybilska et al. (2017) estimated the stellar metallicities for 258 nearby massive quiescent galaxies ( $M_* \gtrsim 10^{9.8} M_\odot$ ) in the ATLAS<sup>3D</sup> survey (Cappellari et al. 2011) using spectrophotometric indices. They also recovered a high-mass MZR similar to that of L19. All these studies indicate a shallower slope and higher normalization in the high-mass MZR than in the low-mass MZR, such that an extrapolation of the high-mass MZR to low

<sup>5</sup> In this paper, MZR refers to the stellar-phase MZR, if not otherwise specified.

<sup>6</sup> K13 measured the stellar metallicities of both quiescent and star-forming dwarf galaxies, but they found no difference in the MZR between these two groups.

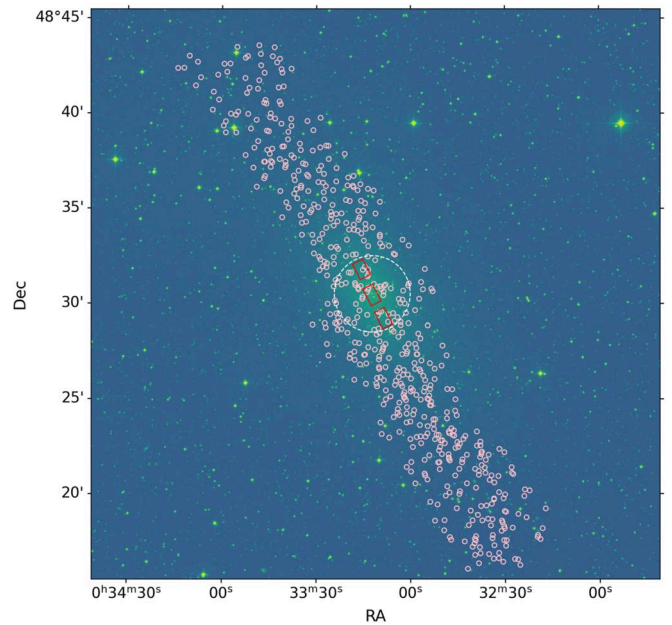
mass would predict much higher metallicities than are observed in low-mass galaxies.

One possible explanation for the discrepancy is that there is a physical break in the MZR around  $10^8$ – $10^{10} M_\odot$ . This behavior would mimic the gas-phase MZR, which follows a single power law at low to intermediate masses ( $10^6 M_\odot \lesssim M_* \lesssim 10^{9.5} M_\odot$ ), and flattens at the high-mass end (e.g., Blanc et al. 2019). The possible transition may result from different feedback mechanisms of the two mass ranges. Theoretically, it has been suggested that energy-driven winds dominate in low-mass galaxies (Murray et al. 2005), while momentum-driven winds dominate in high-mass galaxies (Hopkins et al. 2012). Finlator & Davé (2008) also suggested that the observed gas-phase MZR at  $z \sim 2$  for massive galaxies indicates momentum-driven winds are dominant in the outflows of high-mass galaxies. Such a transition in the slope of the stellar MZR was seen by Sybilka et al. (2017), who found that the slope around  $10^{9.5} M_\odot$  steepens significantly as compared to that at the lowest and highest masses. However, the stellar MZR derived by Sybilka et al. (2017) does not go below  $10^9 M_\odot$ , so it cannot fill the mass gap in the stellar MZR and capture the behavior of the MZR across the full range of galaxy masses. The existing measurements in the mass gap are too scarce to constrain the intermediate-mass stellar MZR.

However, the discrepancy in the MZRs may also result from a systematic difference between the two techniques used to measure stellar metallicities. Previous works that tested the consistency between the stellar population parameters—including the metallicities—among various measurement methods have mainly focused on simple systems like globular clusters (e.g., González Delgado & Cid Fernandes 2010; Conroy et al. 2014, 2018). So far, only Ruiz-Lara et al. (2018) have compared the metallicity of each SFH age bin obtained from the integrated-light spectrum of Leo A, a star-forming dwarf irregular (dIrr) galaxy with  $M_* \sim 3 \times 10^6 M_\odot$ , with that measured by Kirby et al. (2017) from individual stars. They found that the integrated-light [Fe/H] at all age bins is higher than the resolved-star [Fe/H]. The differences in [Fe/H] range from  $\sim 0.25$  to 1.5 dex. However, the differences here can be explained by other reasons, such as the inaccurate modeling of Balmer nebular emission lines, the incompleteness of metal-poor stars in current stellar libraries, or a negative metallicity gradient, as suggested by Kirby et al. (2017).

In this work, we focus on NGC 147, a dwarf elliptical (dE) satellite of M31 with  $M_* \sim 10^8 M_\odot$  (K13). This galaxy is a good target to investigate possible systematic differences between different techniques of measuring stellar metallicities. First, the galaxy is near enough (712 kpc; Conn et al. 2012) to obtain individual stellar metallicities but also far enough to be observed in integrated light with an integral field unit (IFU). Second, no H I gas has been detected in NGC 147, so the spectrum is free from nebular emission line contamination. Third, NGC 147 is more metal-rich than Leo A, so it will suffer less from the incompleteness of metal-poor stars in current stellar libraries.

The main goals of this paper are (1) to determine whether there is a systematic offset between integrated-light and resolved-star spectroscopic measurements of NGC 147’s stellar metallicity, (2) to use these measurements to bolster or diminish evidence of a break in the stellar MZR, and (3) to leverage IFU data cubes to study the spatial distribution of the stellar population in the innermost region. We use both new IFU data cubes and archival resolved-star spectra (Section 2). We use a variety of full spectral



**Figure 1.** Digitized Sky Survey (DSS) image of NGC 147. The red rectangles show the three PCWI fields observed. The dashed circle indicates the circular effective radius. The pink circles show the locations of the RGB star samples used in this study.

fitting algorithms (Section 3) to estimate the global stellar metallicity and age from integrated light (Section 4). Section 4.2 specifically addresses goal (1) in determining the consistency between the two different metallicity techniques. We also present some interesting spatial trends of the stellar population that suggest that NGC 147 formed outside in (Section 5). We compare our work with the previous literature on stellar metallicity determinations and discuss the subsequent implications for the MZR (goal 2) in Section 6. Finally, we summarize our findings in Section 7.

## 2. Observations

### 2.1. Palomar Cosmic Web Imager Sample

#### 2.1.1. Observations and Data Reduction

Observations for NGC 147 were performed with the Palomar Cosmic Web Imager (PCWI) on 2019 September 23 and 24. PCWI is a medium-resolution integral field spectrograph (IFS) mounted on the Cassegrain focus of the Palomar 200 inch Hale Telescope with a field of view of  $60'' \times 40''$ . It consists of 24 spatial slices with a width of  $\sim 2''.5$  and an in-slice pixel scale of  $\sim 0''.5$ . The seeing was around  $2''$  for two nights, so both the spatial axis along each slice and the spectral  $z$ -axis were seeing-limited while the spatial axis across the slices was slit-limited. NGC 147 has an effective radius ( $R_{\text{eff}}$ ) of  $\sim 6''.7$ , which means that a single PCWI pointing covers only a small portion of the galaxy. We observed the central regions of the galaxy within  $2'$  with three pointings (Figure 1), as described in Table 2.

The MEDREZ grating was used to provide the largest spectral range ( $\sim 900 \text{ \AA}$ ) that PCWI offers and a resolution of  $R \simeq 1500$ . With two configurations centered at  $4300 \text{ \AA}$  and  $5140 \text{ \AA}$ , the spectra span  $3750$  to  $5550 \text{ \AA}$ , which is sufficient to measure age, metallicity, and Mg enhancement via full-spectrum fitting (Leethochawalit et al. 2018, 2019).

Because the galaxy is larger than the field of view of PCWI, additional off-field blank-sky observations were taken adjacent

to each science observation with the same exposure time for sky subtraction. The weather was cloudy with the Moon up in the late half of the second night, leading to significant variance in the sky spectra. Therefore, we had to abandon all science frames of the blue side of the north pointing due to inaccurate sky subtraction. Table 2 lists all the observations used in this study.

The raw data were reduced using the standard PCWI Data Reduction Pipeline (DRP).<sup>7</sup> This pipeline converts raw, 2D science frames into flux-calibrated, 3D cubes with spatial positions and wavelengths. PCWI is subject to gravitational flexure because it is mounted at the Cassegrain focus. Flexure induces minor changes in the spectrum, such as alteration of the wavelength solution. Flexure correction in the PCWI DRP relies on cross-correlation of a time series of data cubes. However, the spectral offsets were not completely eliminated after flexure correction. We recalibrated the wavelength solution by cross-correlating the spatially collapsed spectra of each image with the sky spectrum taken nearest to the night calibration set. In this way, the wavelength solution was standardized to a common reference. We applied only a constant offset and ignored higher-order terms.

The world coordinates of each cube were corrected with `CWITools` (O’Sullivan & Chen 2020) via comparing the location of known foreground point sources with accurate R.A. and decl. The sky frames were subtracted from the adjacent science frames directly to obtain sky-subtracted cubes. The sky-subtracted cubes were coadded using `CWITools`, which calculates the footprint of each input pixel on the coadd frame and distributes flux onto the coadd grid.

The instrumental resolution was measured by fitting the widths of arc lines as a function of wavelength. PCWI uses a ThAr lamp, which gives a dense spectrum of emission lines. As a result, some arc lines were blended. We restricted the arc line list to isolated arc lines in order to eliminate the bias introduced by blended lines. We identified isolated arc lines by setting thresholds on the first and second derivatives of the arc flux. First, we required that the first derivative be close to zero and that the second derivative be negative, in order to find the locations of the line peaks. Second, we required that the peak be sharp rather than having a flat or broadened feature. This means that the first derivative must abruptly change from positive to negative, i.e., that the second derivative must be smaller than some threshold. We experimented with different combinations of thresholds and visually inspected the result. We found that setting the first derivative of the spectrum to smaller than 0.2 and the second derivative to smaller than  $-0.01$  provides relatively clean isolated lines. The FWHMs of the isolated lines were then fitted into a quadratic polynomial to obtain the resolution ( $\simeq 3.27 \text{ \AA}$  or  $90 \text{ km s}^{-1}$ ).

However, there was still a slight offset between the blue and red spectra, as we only applied first-order correction. Instead of fitting for the higher-order terms, we transformed the observed wavelengths into the rest frame separately for both wavelength settings. The spatially collapsed spectra of each cube were extracted from the sky-subtracted cubes. We then cross-correlated them with a template galaxy spectrum that was smoothed to the PCWI resolution and resampled to the wavelength grid of our data using `SpectRes` (Carnall 2017). A template with a simple stellar population (SSP) was

generated from Flexible Stellar Population Synthesis (FSPS; Conroy et al. 2009, 2010), based on the physical properties of NGC 147 listed in Table 1. Because NGC 147 has an extended SFH that ended  $\sim 4$  Gyr ago, we adopted a stellar population age of 8 Gyr, which approximately corresponds to the average age of the stellar population (Weisz et al. 2014). We conducted further analyses only on the rest-frame spectra.

### 2.1.2. Foreground Star Masking

As shown in Figure 1, some bright foreground stars are located in the field of PCWI. We masked them before the spectrum extraction to avoid contamination of the galaxy spectrum. The masking regions were determined from the white-light image. For NGC 147, the tip of the red giant branch has an  $i$ -band magnitude of  $\sim 20.82$  (Crnojević et al. 2014), so even the brightest star members would be too faint to be detected individually by PCWI. Therefore, any resolved stars in the field were considered foreground stars. We used the DSS image to identify the foreground stars. For each pointlike source on the DSS image within the PCWI field, we modeled its point-spread function on the white-light image assuming a 2D elliptical Gaussian profile. The point-source mask was determined as a circle centered at the fitted peak, with a radius three times the fitted semimajor axis. Nevertheless, some bright residuals, which may be related to scattered light, were still present in the field. We masked all regions above 1.35 times the median. We also experimented with several thresholds, and it turned out that 1.35 times the median removed most of the bright extended features, which might have contaminated the spectra, while leaving most of the galaxy contribution.

### 2.1.3. Spectrum Extraction

IFS data are known to underestimate the variance due to the covariance between adjacent spaxels. This effect originates from the redistribution of the flux from the same pixel onto a new spatial grid when coadding or smoothing the data cubes. When we coadded the single-exposure cubes, the flux within the same slice with a width of  $\sim 2''.5$  was redistributed again into the final coadded image with a pixel scale of  $0''.58$ . Therefore, the signal-to-noise ratio (S/N) derived from the PCWI DRP was significantly overestimated.

To estimate the true noise, we followed the procedures in O’Sullivan et al. (2020) for PCWI data. First, the variance for each cube was first rescaled by 1.5 to account for the covariance introduced by coadding. For the covariance added by spatially binning the neighboring spaxels, O’Sullivan et al. (2020) measured the true noise ( $\sigma_{\text{adjusted}}$ ) in the binned data as

$$\sigma_{\text{adjusted}} = [1 + 0.79 \log(N_b)] \sigma_{\text{nocov}}, \quad (1)$$

where  $\sigma_{\text{nocov}}$  is the noise derived from error propagation and  $N_b$  is the number of spaxels in the binned spectrum. For large  $N_b \gg 120$ , the signals in different spaxels are uncorrelated, so the scaling becomes  $\sigma_{\text{adjusted}} \simeq 2.6 \sigma_{\text{nocov}}$ .

We stacked the extracted spectra from all the coadded cubes into a single spectrum, which resembles the long-slit integrated-light spectra of more massive galaxies. To remove sensitivity to possibly imperfect flux calibration, we normalized the spectra by the median of the flux between 4400 and 4800  $\text{\AA}$ . We then scaled them according to the surface brightness profile measured by Crnojević et al. (2014) (Table 2) and summed them to obtain the global spectrum. In this case, the number of

<sup>7</sup> PCWI DRP: <https://github.com/scizen9/pdrp>.



**Table 1**  
Basic Properties of NGC 147

R.A. (J2000)	Decl. (J2000)	$M_V$ (mag)	$m_{i,TRGB}$ (mag)	$d_{MW}$ (kpc)	$R_{eff}$ (arcmin (kpc))	$\sigma_v$ (km s <sup>-1</sup> )	$\log(M_*/M_\odot)$ (dex)	[Fe/H] (dex)
00 <sup>h</sup> 33 <sup>m</sup> 12 <sup>s</sup> .1	+48°30′31″	−15.33	20.82 <sup>+0.08</sup> <sub>−0.08</sub>	712 <sup>+21</sup> <sub>−19</sub>	6.70 (1.41)	16 ± 1	8.00 ± 0.05	−0.83 ± 0.25

**Note.** The columns show (1–2) the equatorial coordinates from NED for the galaxy center; (3–6) the absolute total  $V$  magnitude, the apparent TRGB  $i$ -band magnitude, the deprojected distance from the Milky Way (MW), and the effective radius taken from Crnojević et al. (2014); (7) the velocity dispersion measured from the resolved stars (Geha et al. 2010); and (8–9) the stellar mass and stellar metallicity obtained from the resolved RGB stars taken from K13.

**Table 2**  
PCWI Observations of NGC 147

Pointing	R.A. (J2000)	Decl. (J2000)	Surface Brightness <sup>a</sup> (g-band mag arcsec <sup>-2</sup> )	Central Wavelength (Å)	Date	Exposure Time (s)
Central	00 <sup>h</sup> 33 <sup>m</sup> 12 <sup>s</sup> .223	+48°30′27″.41	21.32	4300	2019 Sep 23	6 × 900
				5140	2019 Sep 23	6 × 900
South	00 <sup>h</sup> 33 <sup>m</sup> 08 <sup>s</sup> .561	+48°29′14″.526	21.59	4300	2019 Sep 23	4 × 900
				5140	2019 Sep 24	5 × 900
North	00 <sup>h</sup> 33 <sup>m</sup> 15 <sup>s</sup> .570	+48°31′48″.328	21.58	5140	2019 Sep 24	6 × 900

**Notes.** Some observations lacked quality sufficient for this study. This table does not include those observations.

<sup>a</sup> Derived from the best-fit surface brightness profile for diffuse light in Crnojević et al. (2014).

spaxels used for spectrum extraction is much larger than 120. Therefore, we scaled the weighted variance by  $2.6^2 = 6.76$ .

In addition, we extracted the spatially resolved spectra using the Voronoi binning algorithm of Cappellari & Copin (2003), which is one of the most commonly used binning methods for IFS data. It can adaptively combine the spectra in adjacent pixels to reach a given minimum S/N, and it allows us to adjust the S/N using Equation (1) when optimizing the binning strategy. We only included the central and south pointings for the spatially resolved studies because the north pointing lacks reliable spectral coverage below 4700 Å, which is important to determine the stellar population age. For the central and south pointings, the S/N was calculated from the wavelength range between 4400 and 4600 Å. We required each bin to have  $S/N > 20 \text{ pix}^{-1}$  ( $26 \text{ Å}^{-1}$ ) for the central pointing and  $S/N > 15 \text{ pix}^{-1}$  ( $20 \text{ Å}^{-1}$ ) for the south pointing. We made this choice because the south pointing has much fewer bins in the field due to its lower surface brightness. In the end, we had 17 and 4 bins for the central and south pointings, respectively (Figure 5).

We further corrected the variance of the stacked spectrum and the spatially resolved spectra by comparing with the best-fit model in order to obtain accurate uncertainties of the stellar population parameters (see Section 3).

### 2.2. Keck/DEIMOS Sample

We used archival Keck/DEIMOS (Faber et al. 2003) observations of four slitmasks along the major axis of NGC 147. These observations were originally obtained by Geha et al. (2010). We downloaded the raw data from the Keck Observatory Archive and reduced them with the *spec2d* software (Cooper et al. 2012; Newman et al. 2013).

K13 already measured stellar metallicities from coadded spectra of red giants in NGC 147. However, one of the purposes of this paper is to determine whether individual stellar metallicities are consistent with integrated light. Spectral coaddition is a form of integrated light, which means that the existing measurements in K13 are not appropriate for our

purpose. Therefore, we remeasured the stellar metallicities for individual stars.

We measured metallicities by  $\chi^2$  fitting to a grid of synthetic spectra, as described by Kirby et al. (2008, 2009, 2010). This method uses photometric estimates of effective temperature and surface gravity to guide the fitting. We used photometry from the Pan-Andromeda Archaeological Survey (PAndAS; Martin et al. 2013), which took place at the Canada–France–Hawaii 3.6 m Telescope. This photometry is different from that used by K13, which possibly had errors in its zero-points. Only stars with uncertainties of [Fe/H] smaller than 0.5 dex were included in the further analysis. This cut gives us 317 stars in the sample. The updated measurements are listed in Table 3.

### 3. Methods

We used full-spectrum fitting via stellar population synthesis to measure the stellar population of NGC 147. Compared to spectrophotometric indices, full-spectrum fitting derives the stellar population properties using information from the whole spectrum instead of portions with strong absorption lines, so high-precision measurements can be achieved.

We adopted the fitting method of L19, from which the high-mass MZR was derived, and we refer the reader to that paper for a full description. We chose the same fitting algorithms to minimize the systematic difference caused by different full-spectrum fitting codes and different templates. In summary, L19 used the SSP models from FSPS (Conroy et al. 2009, 2010) version 3.0 with the Kroupa (2001) initial mass function (IMF), Padova isochrones (Marigo & Girardi 2007), and the MILES spectral library (Sánchez-Blázquez et al. 2006). The metallicities and ages of the templates range within  $-1.98 \leq \log Z \leq 0.20$  and 0.3 Myr–14 Gyr, respectively. We interpreted [Z/H] in the base models as [Fe/H]. In addition to the stellar metallicity [Fe/H], enhancements of Mg and N were also included using the theoretical response functions from Conroy et al. (2018), which depend on age and metallicity. The templates were smoothed to match the resolution of the observed data. We fit for the velocity dispersion, stellar population age, [Fe/H], and [Mg/Fe] by interpolating the SSP

**Table 3**  
Metallicity Catalog for RGB Stars Observed with Keck/DEIMOS

ID	R.A.	Decl.	Filter 1	Magnitude 1 <sup>a</sup> (mag)	Filter 2	Magnitude 2 <sup>a</sup> (mag)	$T_{\text{eff}}$ (K)	$\log g$ ( $\text{cm s}^{-2}$ )	$\xi$ ( $\text{km s}^{-1}$ )	[Fe/H] (dex)
4456	00 <sup>h</sup> 32 <sup>m</sup> 23 <sup>s</sup> .9	+48°19′53″.8	<i>g</i>	23.065 ± 0.027	<i>i</i>	20.921 ± 0.012	3744 ± 44	0.26 ± 0.1	2.08	−1.12 ± 0.12
4608	00 <sup>h</sup> 32 <sup>m</sup> 25 <sup>s</sup> .8	+48°18′57″.2	<i>g</i>	23.106 ± 0.028	<i>i</i>	21.007 ± 0.013	3770 ± 42	0.32 ± 0.1	2.07	−0.49 ± 0.11
4778	00 <sup>h</sup> 32 <sup>m</sup> 27 <sup>s</sup> .6	+48°18′13″.8	<i>g</i>	22.715 ± 0.020	<i>i</i>	21.117 ± 0.014	4126 ± 45	0.62 ± 0.1	2.00	−1.10 ± 0.12
4885	00 <sup>h</sup> 32 <sup>m</sup> 29 <sup>s</sup> .5	+48°17′34″.2	<i>g</i>	23.297 ± 0.032	<i>i</i>	20.808 ± 0.011	3592 ± 56	0.04 ± 0.1	2.13	−3.03 ± 0.17
4894	00 <sup>h</sup> 32 <sup>m</sup> 29 <sup>s</sup> .7	+48°19′03″.8	<i>g</i>	23.239 ± 0.031	<i>i</i>	21.424 ± 0.018	3929 ± 41	0.64 ± 0.1	1.99	−0.81 ± 0.12
4961	00 <sup>h</sup> 32 <sup>m</sup> 30 <sup>s</sup> .5	+48°19′46″.6	<i>g</i>	24.125 ± 0.066	<i>i</i>	21.426 ± 0.018	3528 ± 51	0.23 ± 0.1	2.09	−3.21 ± 0.23
5026	00 <sup>h</sup> 32 <sup>m</sup> 31 <sup>s</sup> .1	+48°18′04″.2	<i>g</i>	22.977 ± 0.025	<i>i</i>	21.163 ± 0.014	3936 ± 39	0.53 ± 0.1	2.02	−0.85 ± 0.11
5129	00 <sup>h</sup> 32 <sup>m</sup> 32 <sup>s</sup> .0	+48°16′37″.4	<i>g</i>	23.416 ± 0.035	<i>i</i>	21.145 ± 0.014	3748 ± 52	0.31 ± 0.1	2.07	−2.23 ± 0.16
5324	00 <sup>h</sup> 32 <sup>m</sup> 33 <sup>s</sup> .6	+48°16′36″.3	<i>g</i>	23.099 ± 0.027	<i>i</i>	21.238 ± 0.015	3923 ± 49	0.54 ± 0.1	2.02	−1.08 ± 0.13
5368	00 <sup>h</sup> 32 <sup>m</sup> 33 <sup>s</sup> .9	+48°17′50″.8	<i>g</i>	22.779 ± 0.021	<i>i</i>	21.256 ± 0.016	4232 ± 42	0.72 ± 0.1	1.97	−1.36 ± 0.12
...	...	...	...	...	...	...	...	...	...	...

**Notes.** This table is available in its entirety in machine-readable format in the online journal. A portion is shown here for guidance regarding its form and content.

<sup>a</sup> PAndAS photometry corrected for extinction.

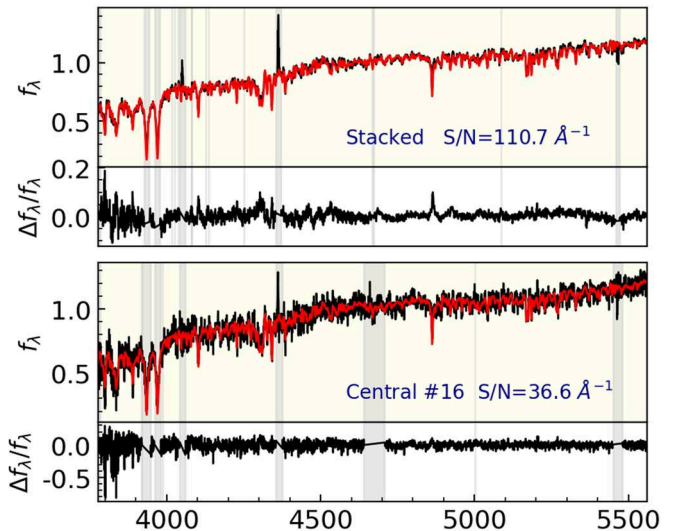
(This table is available in its entirety in machine-readable form.)

models. Therefore, the obtained ages and metallicities are SSP-equivalent. Leethochawalit et al. (2018) tested the SSP assumption by fitting mock composite stellar population (CSP) model spectra, which have exponentially declining SFHs, with the SSP models, and successfully recovered the true metallicity within 0.1 dex. This result indicates that SSP-equivalent metallicity agrees well with light- or mass-weighted metallicity. We emphasize that the goal of this paper is to investigate whether the high-mass MZR measured in SSP-equivalent [Fe/H] has systematic offsets relative to the low-mass MZR. As demonstrated in Appendix, although the choice of SSP or CSP model will affect the stellar age estimates, the measured metallicity is much less affected. Therefore, adopting SSP models will not affect the major conclusion of this paper. We focus on the recovery of SSP-equivalent parameters in the following analyses.

To avoid a possible mismatch in flux calibration between the blue and red spectra, we normalized the spectral continuum on both sides using a spline fit. L19 developed an algorithm to minimize any alteration to the template spectra. The algorithm applies a synthesized continuum to the observed continuum-normalized spectrum. In the first iteration, we fit the continuum-normalized observed spectrum with continuum-normalized model spectra, and generated a synthesized continuum by fitting a B-spline to the quotient of the continuum-normalized spectrum and the best-fit SSP model spectrum. Subsequent fits were performed on the corrected observed spectrum with unaltered model spectra. The method is described in full detail in L19.

For each spectrum, we performed two separate fits. For the first fit, we only masked out wavelength regions contaminated by strong sky line residuals. The Ca II H and K lines were also excluded because of strong nonlocal thermodynamic equilibrium effects, which current models poorly reproduce. After the first fit, we masked out regions that were  $\geq 3\sigma$  from the best-fit model. We visually inspected these newly masked regions to make sure they deviated from the model due to weak sky features rather than to deficiencies in the fitting algorithm. Then, we rescaled the variance derived from Section 2.1.3 by the reduced  $\chi^2$  of the first fit to estimate the true noise. We then fit the spectrum with new masking regions and corrected variance again to measure the stellar population properties.

To validate our measurements, we also compared our results with those derived from `alf` (Conroy et al. 2018), which is another full-spectrum fitting algorithm capable of measuring



**Figure 2.** Two examples of the observed spectra (black), the corresponding best-fit models (red), and the model residuals (bottom panel of each spectrum). The gray regions are not included in the spectrum fitting. The top panel shows the stacked spectrum from the entire field, which has the highest S/N, while the bottom panel is an example of a spectrum at the lower end of the S/N in one Voronoi cell. The S/N is estimated from the residual of the fits.

detailed abundances using the Markov Chain Monte Carlo (MCMC) method of parameter estimation. We used the “simple” mode of `alf`, which also assumes SSPs. The [Fe/H] and [Mg/Fe] results from our own code are consistent with those obtained from `alf` within  $2\sigma$ .

Examples of the observed spectra and best-fit models at the highest and lowest ends of the S/N are shown in Figure 2. We summarize our results for the global spectrum in Table 4 and for the spatially resolved spectra in Table 5.

#### 4. Global Properties

In this section, we report the stellar population properties of NGC 147 from the global integrated-light spectrum and compare the best-fit SSP-equivalent results with the updated measurements of the resolved stars.

**Table 4**

Best-fit Stellar Population Parameters for the Stacked Spectrum of NGC 147

Row	Quantity	Units	Best-fit value
(1)	$\sigma_{\text{SSP}}$	$\text{km s}^{-1}$	$40.45 \pm 1.85$
(2)	$\text{Age}_{\text{SSP}}$	Gyr	$8.11^{+0.74}_{-0.54}$
(3)	$[\text{Fe}/\text{H}]_{\text{SSP}}$	dex	$-1.069^{+0.029}_{-0.031}$
(4)	$[\text{Mg}/\text{Fe}]_{\text{SSP}}$	dex	$0.024^{+0.20}_{-0.20}$
(5)	$\sigma_{\text{alf}}$	$\text{km s}^{-1}$	$23.59^{+3.35}_{-3.68}$
(6)	$\text{Age}_{\text{alf}}$	Gyr	$12.50^{+0.32}_{-0.31}$
(7)	$[\text{Fe}/\text{H}]_{\text{alf}}$	dex	$-1.105^{+0.011}_{-0.010}$
(8)	$[\text{Mg}/\text{Fe}]_{\text{alf}}$	dex	$0.090^{+0.017}_{-0.017}$

**Note.** (1)–(4) are derived from L19’s method; (5)–(8) are obtained from the *alf* simple mode.

#### 4.1. SSP Age

The best-fit SSP age derived from our method is  $8.11^{+0.74}_{-0.54}$  Gyr, while *alf* obtains an age of  $12.50^{+0.32}_{-0.31}$  Gyr. We suspect the difference ( $\sim 0.14$  dex) may be attributed to systematic discrepancies of the SSP models used in the fitting codes. First, *alf* uses the newer MIST isochrones (Choi et al. 2016; Dotter 2016) with metallicity coverage larger than that of the Padova isochrones used in our method. Second, the intrinsic spectral resolution of the stellar libraries in *alf* is lower than the instrumental resolution of PCWI, so the observed spectra needed to be smoothed by another  $\sigma = 100 \text{ km s}^{-1}$  before being input to *alf*. Third, even in the simple mode, *alf* includes six more elements than our code does. Degeneracy between the extra free parameters may also affect the posterior distributions and the final results.

Ideally, we would compare the age of the stellar population measured from integrated light to the average ages of individual stars. Unfortunately, it is extremely challenging to measure the ages of individual RGB stars. Generally, the stellar age can only be indirectly estimated by comparing the observed data with stellar models (e.g., Soderblom 2010). The most commonly used method is to place the observations on a color–magnitude diagram (CMD), and match them with isochrones in a Hertzsprung–Russell diagram. This method works best for main-sequence and subgiant stars and is less effective for old, metal-poor RGB stars. For example, the difference in color between a 9 Gyr and a 12 Gyr red giant is on the order of the difference in color between isochrone sets modeling the same star.

We therefore infer the accuracy of the SSP age by comparing it with the age measured from CMD fitting via Hubble Space Telescope (HST) photometry by Weisz et al. (2014). They obtained a mass-weighted age of 7.53 Gyr, more consistent with the SSP age derived from our algorithm than with that derived from *alf*. The agreement suggests that even though NGC 147 has an extended SFH with several stellar populations formed at distinct epochs, the SSP assumption may be valid to recover the mass-weighted age in some cases, though the recovered stellar age is highly model-dependent. Therefore, obtaining a more reliable and consistent measurement of stellar age for galaxies sometimes requires a model of the SFH more complex than an SSP. As suggested in Appendix, CSP models may be more appropriate to recover the stellar age of complicated systems, although this method would also require higher data quality.

#### 4.2. Iron Abundance

Before comparing the metallicity determined from integrated light ( $[\text{Fe}/\text{H}]_{\text{SSP}}$ ) to that of resolved stars ( $[\text{Fe}/\text{H}]_{\text{stars}}$ ), it is

important to understand whether the method of averaging  $[\text{Fe}/\text{H}]_{\text{stars}}$  would affect the final results. The median, mean, and inverse variance-weighted mean of  $[\text{Fe}/\text{H}]_{\text{stars}}$  are  $-1.047 \pm 0.044$ ,  $-1.106 \pm 0.031$ , and  $-0.983 \pm 0.024$ , respectively. As shown in Figure 3, the  $[\text{Fe}/\text{H}]_{\text{SSP}} = -1.069^{+0.029}_{-0.031}$  from integrated light is consistent with the median and the mean of  $[\text{Fe}/\text{H}]_{\text{stars}}$  within  $1\sigma$ . On the other hand,  $[\text{Fe}/\text{H}]_{\text{SSP}}$  is slightly lower than the weighted mean, although the difference is smaller than 0.1 dex.

We calculated the uncertainties of  $[\text{Fe}/\text{H}]_{\text{SSP}}$  according to the methods of L19, which assume perfect model spectra. This assumption is not true in reality. First, NGC 147 has multiple stellar populations, contrary to the model’s SSP assumption. Second, even for a simple system with an SSP, synthetic spectra are generated from the empirical stellar libraries and the theoretical isochrones, none of which is perfect. Therefore, we attribute the small difference ( $\lesssim 0.1$  dex) between the two techniques to systematic error that is not included in the uncertainties of the best-fit parameters. Still, it is appropriate to conclude that the metallicity measured from the resolved stellar spectroscopy is consistent with that derived from the integrated-light spectroscopy for NGC 147. In addition, the consistency between the two methods indicates that the SSP assumption is sufficient to recover the true metallicity of a galaxy with a complicated SFH.

The agreement between the integrated-light and resolved stellar metallicities also implies that the metallicity measured from the center of NGC 147 (covered by PCWI) is representative of the much larger region covered by DEIMOS. Although NGC 147 has a slightly negative metallicity gradient (see Section 5), the two methods agree despite their drastically different spatial coverage. If this holds true for more distant galaxies, the limited spatial coverage of galaxy surveys with fiber spectrographs would have less impact on the measurement of stellar metallicity for galaxies with metallicity gradients as shallow as those of NGC 147. A large-sample study is essential to further investigating the limited-aperture effect on stellar metallicity recovery but it is beyond the scope of this work.

#### 4.3. Magnesium Enhancement

Ideally, we would compare Mg abundances measured from integrated light to those measured from resolved stars. It is possible to measure  $[\text{Mg}/\text{Fe}]$  from DEIMOS spectra of individual stars (e.g., Kirby et al. 2010). However, the spectral range of the DEIMOS configuration employed by the archival spectra used in this paper does not reach the Mg b triplet or other strong Mg lines. Instead, the available lines have relatively high excitation potential, and they are consequently weak. This makes it difficult to measure Mg in low-S/N spectra, including those of stars in galaxies as distant as NGC 147. The result of attempting to measure  $[\text{Mg}/\text{Fe}]$  in such galaxies is a bias toward high values of  $[\text{Mg}/\text{Fe}]$ <sup>8</sup> because low values of  $[\text{Mg}/\text{Fe}]$  result in the Mg lines becoming lost in the noise.

An alternative way to validate  $[\text{Mg}/\text{Fe}]_{\text{SSP}}$  is to use the MZR for  $[\text{Mg}/\text{H}]$ . The low-mass MZR for  $[\text{Mg}/\text{H}]$  can be deduced from resolved stellar spectroscopy of MW satellite galaxies (Kirby et al. 2010), which are near enough to allow S/N that overcome the bias against low values of Mg abundances. The high-mass MZR can be measured from integrated-light spectroscopy of more distant galaxies (L19). The low-mass and high-mass MZRs for  $[\text{Mg}/\text{H}]$  predict different values at the mass of NGC 147, as is the case for  $[\text{Fe}/\text{H}]$  (see Section 6.2). If

<sup>8</sup> See Figure 10 of Kirby et al. (2020) for a demonstration of this effect.



**Table 5**  
Best-fit Stellar Population Parameters for Spatially Resolved Integrated Spectra

Pointing	R.A. (J2000)	Decl. (J2000)	$R_{\text{ellip}}$ (arcmin (kpc))	$\text{Age}_{\text{SSP}}^a$ (Gyr)	$[\text{Fe}/\text{H}]_{\text{SSP}}^a$ (dex)	$[\text{Mg}/\text{Fe}]_{\text{SSP}}^a$ (dex)	$\text{Age}_{\text{alf}}$ (Gyr)	$[\text{Fe}/\text{H}]_{\text{alf}}$ (dex)	$[\text{Mg}/\text{Fe}]_{\text{alf}}$ (dex)	$\text{S/N}^b$ ( $\text{\AA}^{-1}$ )
Central	00 <sup>h</sup> 33 <sup>m</sup> 11.72	+48°30′33″.8	0.14 (0.03)	7.08 <sup>+0.69</sup> <sub>-0.44</sub>	-1.11 <sup>+0.03</sup> <sub>-0.04</sub>	0.13 <sup>+0.24</sup> <sub>-0.28</sub>	7.86 <sup>+0.34</sup> <sub>-0.43</sub>	-0.99 <sup>+0.04</sup> <sub>-0.03</sub>	0.14 <sup>+0.05</sup> <sub>-0.05</sub>	40.6
	00 <sup>h</sup> 33 <sup>m</sup> 12.37	+48°30′40″.2	0.17 (0.04)	7.55 <sup>+0.89</sup> <sub>-0.55</sub>	-1.11 <sup>+0.04</sup> <sub>-0.04</sub>	0.44 <sup>+0.20</sup> <sub>-0.25</sub>	11.67 <sup>+0.63</sup> <sub>-0.52</sub>	-1.20 <sup>+0.03</sup> <sub>-0.03</sub>	0.35 <sup>+0.05</sup> <sub>-0.05</sub>	39.9
	00 <sup>h</sup> 33 <sup>m</sup> 13.20	+48°30′35″.8	0.25 (0.05)	7.43 <sup>+0.89</sup> <sub>-0.54</sub>	-1.09 <sup>+0.04</sup> <sub>-0.04</sub>	-0.04 <sup>+0.28</sup> <sub>-0.29</sub>	9.35 <sup>+0.57</sup> <sub>-0.70</sub>	-1.12 <sup>+0.04</sup> <sub>-0.04</sub>	0.09 <sup>+0.06</sup> <sub>-0.06</sub>	38.4
	00 <sup>h</sup> 33 <sup>m</sup> 12.97	+48°30′27″.0	0.29 (0.06)	7.63 <sup>+1.19</sup> <sub>-0.68</sub>	-1.05 <sup>+0.04</sup> <sub>-0.04</sub>	-0.37 <sup>+0.32</sup> <sub>-0.31</sub>	9.40 <sup>+0.55</sup> <sub>-0.39</sub>	-0.96 <sup>+0.03</sup> <sub>-0.03</sub>	-0.11 <sup>+0.05</sup> <sub>-0.05</sub>	42.1
	00 <sup>h</sup> 33 <sup>m</sup> 13.28	+48°30′48″.2	0.34 (0.07)	7.85 <sup>+1.55</sup> <sub>-0.88</sub>	-1.08 <sup>+0.04</sup> <sub>-0.05</sub>	-0.18 <sup>+0.32</sup> <sub>-0.33</sub>	12.14 <sup>+0.81</sup> <sub>-0.81</sub>	-1.21 <sup>+0.03</sup> <sub>-0.03</sub>	-0.03 <sup>+0.06</sup> <sub>-0.06</sub>	36.0
	00 <sup>h</sup> 33 <sup>m</sup> 10.62	+48°30′20″.6	0.35 (0.07)	7.83 <sup>+1.17</sup> <sub>-0.71</sub>	-1.15 <sup>+0.04</sup> <sub>-0.04</sub>	-0.36 <sup>+0.28</sup> <sub>-0.28</sub>	11.05 <sup>+0.54</sup> <sub>-0.41</sub>	-1.28 <sup>+0.03</sup> <sub>-0.03</sub>	-0.10 <sup>+0.05</sup> <sub>-0.05</sub>	44.1
	00 <sup>h</sup> 33 <sup>m</sup> 11.43	+48°30′42″.2	0.37 (0.08)	7.62 <sup>+0.85</sup> <sub>-0.54</sub>	-1.11 <sup>+0.04</sup> <sub>-0.04</sub>	0.09 <sup>+0.25</sup> <sub>-0.28</sub>	12.01 <sup>+0.65</sup> <sub>-0.64</sub>	-1.19 <sup>+0.03</sup> <sub>-0.03</sub>	0.10 <sup>+0.06</sup> <sub>-0.06</sub>	39.6
	00 <sup>h</sup> 33 <sup>m</sup> 12.49	+48°30′15″.4	0.42 (0.09)	8.87 <sup>+1.63</sup> <sub>-1.12</sub>	-1.14 <sup>+0.05</sup> <sub>-0.05</sub>	-0.33 <sup>+0.32</sup> <sub>-0.31</sub>	13.90 <sup>+0.08</sup> <sub>-0.35</sub>	-1.31 <sup>+0.03</sup> <sub>-0.04</sub>	0.01 <sup>+0.06</sup> <sub>-0.06</sub>	36.3
	00 <sup>h</sup> 33 <sup>m</sup> 10.98	+48°30′08″.6	0.43 (0.09)	7.03 <sup>+0.69</sup> <sub>-0.42</sub>	-1.03 <sup>+0.03</sup> <sub>-0.03</sub>	-0.42 <sup>+0.32</sup> <sub>-0.31</sub>	7.17 <sup>+0.34</sup> <sub>-0.28</sub>	-1.03 <sup>+0.03</sup> <sub>-0.03</sub>	-0.02 <sup>+0.04</sup> <sub>-0.05</sub>	39.7
	00 <sup>h</sup> 33 <sup>m</sup> 11.91	+48°30′10″.6	0.44 (0.09)	7.87 <sup>+1.39</sup> <sub>-0.88</sub>	-1.10 <sup>+0.04</sup> <sub>-0.05</sub>	-0.09 <sup>+0.30</sup> <sub>-0.31</sub>	9.96 <sup>+0.57</sup> <sub>-0.56</sub>	-1.07 <sup>+0.03</sup> <sub>-0.03</sub>	-0.05 <sup>+0.06</sup> <sub>-0.06</sub>	36.1
	00 <sup>h</sup> 33 <sup>m</sup> 12.18	+48°30′53″.6	0.48 (0.10)	7.79 <sup>+1.42</sup> <sub>-0.86</sub>	-1.07 <sup>+0.04</sup> <sub>-0.04</sub>	-0.05 <sup>+0.30</sup> <sub>-0.32</sub>	10.87 <sup>+0.70</sup> <sub>-0.60</sub>	-1.12 <sup>+0.04</sup> <sub>-0.03</sub>	0.17 <sup>+0.06</sup> <sub>-0.06</sub>	35.1
	00 <sup>h</sup> 33 <sup>m</sup> 13.95	+48°30′30″.2	0.51 (0.11)	7.98 <sup>+1.38</sup> <sub>-0.92</sub>	-1.11 <sup>+0.04</sup> <sub>-0.04</sub>	0.14 <sup>+0.26</sup> <sub>-0.29</sub>	12.43 <sup>+0.65</sup> <sub>-0.70</sub>	-1.12 <sup>+0.03</sup> <sub>-0.03</sub>	0.05 <sup>+0.06</sup> <sub>-0.06</sub>	39.1
	00 <sup>h</sup> 33 <sup>m</sup> 09.99	+48°30′11″.7	0.51 (0.11)	8.00 <sup>+1.50</sup> <sub>-0.92</sub>	-1.11 <sup>+0.05</sup> <sub>-0.05</sub>	-0.19 <sup>+0.33</sup> <sub>-0.32</sub>	12.60 <sup>+0.76</sup> <sub>-0.89</sub>	-1.18 <sup>+0.03</sup> <sub>-0.03</sub>	-0.06 <sup>+0.06</sup> <sub>-0.06</sub>	34.4
	00 <sup>h</sup> 33 <sup>m</sup> 14.36	+48°30′42″.1	0.52 (0.11)	7.22 <sup>+0.94</sup> <sub>-0.58</sub>	-1.06 <sup>+0.04</sup> <sub>-0.04</sub>	-0.01 <sup>+0.28</sup> <sub>-0.30</sub>	9.25 <sup>+0.55</sup> <sub>-0.47</sub>	-1.06 <sup>+0.04</sup> <sub>-0.03</sub>	0.07 <sup>+0.06</sup> <sub>-0.06</sub>	36.9
	00 <sup>h</sup> 33 <sup>m</sup> 13.59	+48°30′21″.7	0.54 (0.11)	7.72 <sup>+1.26</sup> <sub>-0.80</sub>	-1.10 <sup>+0.04</sup> <sub>-0.04</sub>	-0.09 <sup>+0.30</sup> <sub>-0.30</sub>	12.90 <sup>+0.61</sup> <sub>-0.69</sub>	-1.18 <sup>+0.03</sup> <sub>-0.03</sub>	0.02 <sup>+0.06</sup> <sub>-0.06</sub>	37.2
	00 <sup>h</sup> 33 <sup>m</sup> 12.22	+48°30′01″.1	0.69 (0.14)	7.94 <sup>+1.19</sup> <sub>-0.73</sub>	-1.07 <sup>+0.04</sup> <sub>-0.04</sub>	-0.04 <sup>+0.29</sup> <sub>-0.30</sub>	13.29 <sup>+0.52</sup> <sub>-0.78</sub>	-1.20 <sup>+0.03</sup> <sub>-0.03</sub>	0.13 <sup>+0.05</sup> <sub>-0.05</sub>	42.6
South	00 <sup>h</sup> 33 <sup>m</sup> 08.32	+48°29′22″.1	1.33 (0.28)	8.92 <sup>+1.57</sup> <sub>-1.07</sub>	-1.05 <sup>+0.03</sup> <sub>-0.04</sub>	-0.00 <sup>+0.25</sup> <sub>-0.29</sub>	13.93 <sup>+0.05</sup> <sub>-0.05</sub>	-1.05 <sup>+0.02</sup> <sub>-0.02</sub>	0.03 <sup>+0.05</sup> <sub>-0.05</sub>	38.7
	00 <sup>h</sup> 33 <sup>m</sup> 09.93	+48°29′19″.5	1.38 (0.29)	8.36 <sup>+1.41</sup> <sub>-0.92</sub>	-1.07 <sup>+0.04</sup> <sub>-0.04</sub>	-0.04 <sup>+0.29</sup> <sub>-0.31</sub>	13.56 <sup>+0.33</sup> <sub>-0.68</sub>	-1.08 <sup>+0.03</sup> <sub>-0.03</sub>	0.05 <sup>+0.05</sup> <sub>-0.04</sub>	39.9
	00 <sup>h</sup> 33 <sup>m</sup> 07.47	+48°29′04″.0	1.67 (0.35)	8.73 <sup>+1.72</sup> <sub>-1.25</sub>	-1.10 <sup>+0.04</sup> <sub>-0.04</sub>	-0.09 <sup>+0.29</sup> <sub>-0.32</sub>	13.46 <sup>+0.42</sup> <sub>-0.98</sub>	-1.08 <sup>+0.03</sup> <sub>-0.03</sub>	0.05 <sup>+0.05</sup> <sub>-0.05</sub>	38.2
	00 <sup>h</sup> 33 <sup>m</sup> 09.16	+48°28′59″.7	1.76 (0.37)	8.56 <sup>+1.52</sup> <sub>-0.95</sub>	-1.08 <sup>+0.04</sup> <sub>-0.04</sub>	0.05 <sup>+0.26</sup> <sub>-0.31</sub>	13.81 <sup>+0.16</sup> <sub>-0.56</sub>	-1.10 <sup>+0.03</sup> <sub>-0.03</sub>	0.06 <sup>+0.04</sup> <sub>-0.04</sub>	40.0

#### Notes.

<sup>a</sup> Measured from L19’s method.

<sup>b</sup> The S/N is calculated from the residuals of the best-fit model.

$[\text{Mg}/\text{H}]_{\text{SSP}}$  (the sum of the best-fit  $[\text{Mg}/\text{Fe}]_{\text{SSP}}$  and  $[\text{Fe}/\text{H}]_{\text{SSP}}$ ) falls on the low-mass MZR ( $[\text{Mg}/\text{H}]$ ) for dwarf galaxies, then we have circumstantial evidence that the two methods give similar values of Mg abundance.

We constructed the MZR for  $[\text{Mg}/\text{H}]$  from the measurements of eight MW dwarf spheroidal (dSph) satellites (Kirby et al. 2010). We excluded more distant galaxies due to the aforementioned weak Mg lines, and we excluded most ultrafaint dwarf galaxies due to small sample sizes. We constructed the MZR from an orthogonal, least-squares linear regression taking into account measurement uncertainties in both axes (Akritas & Bershady 1996). The resulting MZR is  $\langle [\text{Mg}/\text{H}] \rangle = -3.31 + 0.32 \log(M_*/M_\odot)$ .

As seen in Figure 4, the  $[\text{Mg}/\text{H}]_{\text{SSP}}$  of NGC 147 is much more likely to fall on the low-mass MZR (Mg) of dwarf galaxies than on the high-mass one. Considering the low-mass MZR (Mg) was determined only from a few dwarf galaxies at  $M_* < 10^8 M_\odot$ , the small difference ( $< 0.2$  dex) between  $[\text{Mg}/\text{H}]_{\text{SSP}}$  and the predicted value might be attributed to an extrapolation error at the mass of NGC 147 or to intrinsic dispersion in the MZR. We therefore conclude that we have tentative evidence that  $[\text{Mg}/\text{Fe}]_{\text{SSP}}$  is representative of the whole stellar population in the galaxy.

## 5. Spatial Distributions of Stellar Population

### 5.1. Population Gradients

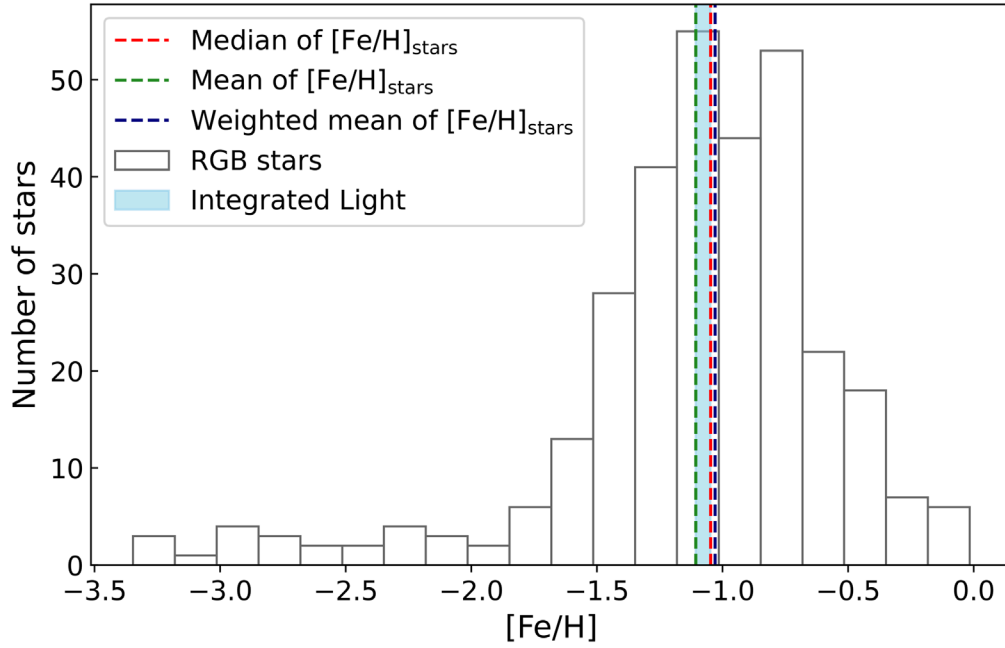
Taking advantage of IFS data, we were able to study the spatial distribution of the stellar population in the innermost region for the first time. We measured the stellar population in each Voronoi cell (Section 2.1.3). Because the north pointing lacks reliable blue spectra below 4700 Å, which are essential to

determining the stellar population age, we excluded the north pointing for spatial distribution studies.

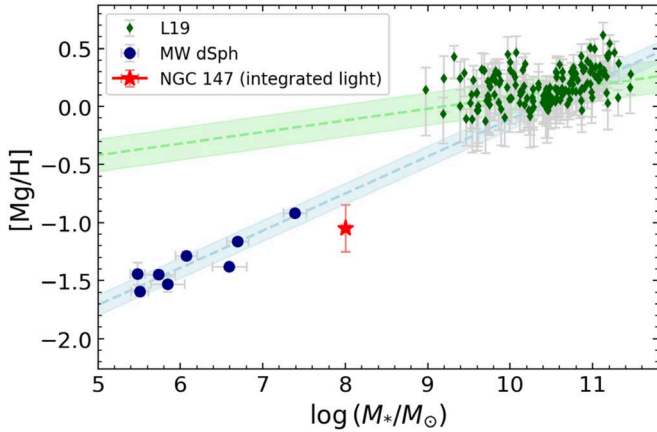
Figure 5 shows the 2D maps and the radial trends for the SSP-equivalent stellar population age,  $[\text{Fe}/\text{H}]$ , and  $[\text{Mg}/\text{Fe}]$ . As demonstrated by Koleva et al. (2011), SSP-equivalent parameters are sufficient to recover the true population gradients. We calculated the galactocentric elliptical radius using the position angle and ellipticity in Crnojević et al. (2014). We fit a linear model to the gradients for age,  $[\text{Fe}/\text{H}]$ , and  $[\text{Mg}/\text{Fe}]$ , with the `emcee` MCMC sampler (Foreman-Mackey et al. 2013).<sup>9</sup> The uncertainties of each parameter were taken to be the asymmetric 68% confidence level of the posteriors.

In the PCWI field, we marginally detected a positive age gradient within 0.4 kpc as  $m_{\text{age}} = \Delta \left( \frac{\text{age}}{R} \right) = 4.5 \pm 2.5 \text{ Gyr kpc}^{-1}$ . The south field is  $\sim 1.5$  Gyr older than the innermost bin. The positive age gradient may suggest that star formation persisted longer in the central region compared to that in more extended regions. However, no age gradient is present in the `alf` measurements. As shown in Table 5, around half of the best-fit results are very close to the upper limit of the stellar population age (i.e., the age of the universe), suggesting the age derived from `alf` is not well constrained. We suspect the factors discussed in Section 4.1 may also account for the difference. Meanwhile, the detected positive age gradient is consistent with previous literature. Han et al. (1997) also found that the younger stars in NGC 147 are more centrally concentrated than the majority of the old stars. Weisz et al. (2014) measured the SFH from HST/

<sup>9</sup> For each gradient, the initial parameters were obtained from  $\chi^2$  minimization. Broad, uniform priors were used. We ran 5000 steps with 32 walkers. Burn-in usually happened after around 40 steps, so the posteriors were well constrained.



**Figure 3.** The metallicity distribution for the resolved stars analyzed in this sample. The value of  $[\text{Fe}/\text{H}]_{\text{SSP}}$  (light blue region) is consistent with the mean (green dashed line), the median (red dashed line) and the inverse-variance weighted mean (blue dashed line) of  $[\text{Fe}/\text{H}]_{\text{stars}}$ .



**Figure 4.** The MZR measured from  $[\text{Mg}/\text{H}]$  at  $z \sim 0$ . The green dashed line shows the high-mass MZR for Mg determined from the integrated spectra of massive quiescent galaxies (green diamonds) by L19. The low-mass MZR for Mg is determined only from the MW dSph satellites (blue circles). The rms for the two best-fit lines are shown in the green and blue bands. The  $[\text{Mg}/\text{H}]$  of NGC 147 measured in this work (red star) is more likely to fall on the low-mass MZR than on the high-mass MZR.

WFPC2 photometry in the central regions of NGC 147, which demonstrates that  $\sim 20\%$  of the stars were formed within the past 2.5 Gyr. On the other hand, the SFH determined from an HST/ACS field targeted at much larger radii suggests NGC 147 formed 80% of its stars ( $6_{-1}^{+2}$ ) Gyr ago and its star formation ceased at least 3 Gyr ago (Geha et al. 2015). The different SFHs derived from HST data at different radii are qualitatively consistent with the positive age gradients we measured from integrated-light spectroscopy.

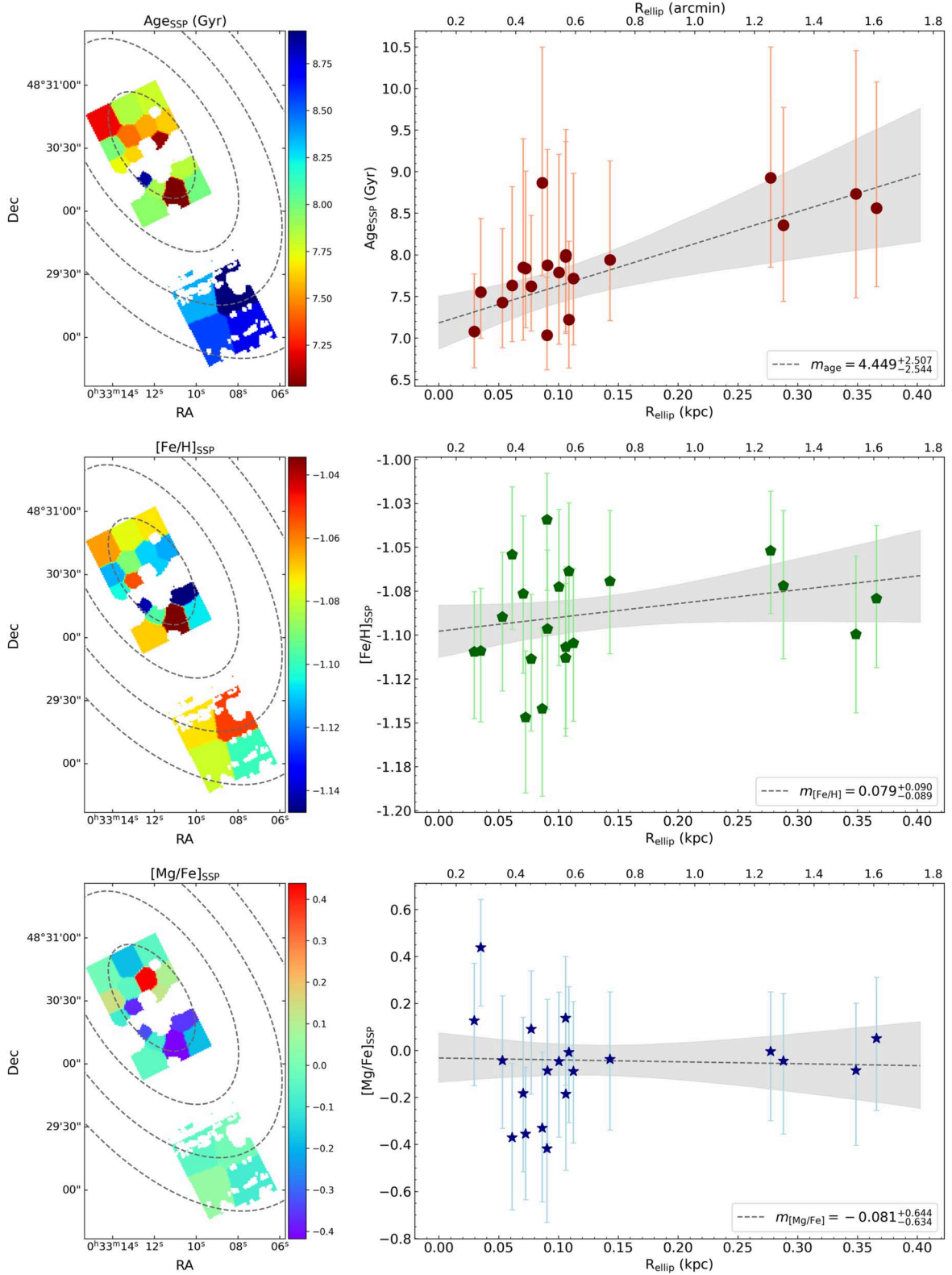
The metallicity is relatively uniform in the PCWI fields, with a slope of  $m_{[\text{Fe}/\text{H}]} = \Delta \left( \frac{[\text{Fe}/\text{H}]}{R} \right) = 0.08 \pm 0.09 \text{ dex kpc}^{-1}$ . On the other hand, we find a large scatter for  $[\text{Mg}/\text{Fe}]$ , with a slope of  $m_{[\text{Mg}/\text{Fe}]} = \Delta \left( \frac{[\text{Mg}/\text{Fe}]}{R} \right) = -0.08_{-0.63}^{+0.64} \text{ dex kpc}^{-1}$ . The large uncertainty on the  $[\text{Mg}/\text{Fe}]$  gradient precludes us from drawing

conclusions about the relevance of this gradient to the galaxy's SFH. Furthermore, the gradient was estimated from the very central region of the galaxy so it may not represent the whole stellar population.

To investigate the population gradient at larger scales, we made use of the stellar metallicities of resolved RGB stars. As shown in Figure 6, the resolved metallicities for the innermost region ( $\sim 400$  pc) show large scatter. There are only 17 stars within the outermost radius of the PCWI fields, so the resolved metallicities may not indicate the characteristic metallicity of the stellar population in this region. The integrated metallicities derived from PCWI data are between the lowest and highest values of the resolved metallicities. The resolved study reveals a negative gradient of  $\Delta \frac{\langle [\text{Fe}/\text{H}] \rangle}{R} \sim -0.13 \pm 0.03 \text{ dex kpc}^{-1}$ , different from the flat metallicity profile revealed in the PCWI fields. Because the resolved study extends to a much larger radius, we believe that the negative metallicity gradient is more representative of the galaxy population.

The negative metallicity gradient is inconsistent with the flat to positive gradient found by Crnojević et al. (2014) and Vargas et al. (2014). Crnojević et al. (2014) estimated the photometric metallicity gradient from the metallicity distribution functions in three radial bins from 2.5 to 6.4 kpc. Their sample consists of stars at much larger radii compared to ours and lacks measurements within 2.5 kpc. In addition, the authors suggested that their gradient may be biased because stars from tidal tails can contribute to the mean stellar metallicity in the outermost region. On the other hand, Vargas et al. (2014) measured spectroscopic metallicities from Keck/DEIMOS spectra. Similar to that in K13, the photometry used to estimate the effective temperature and surface gravity possibly had errors in its zero-points, leading to unreliable metallicity estimates. In fact, Vargas et al. (2014) reported a mean  $[\text{Fe}/\text{H}]$  for NGC 147 of  $\sim -0.5$  dex, which is around twice the values derived in other literature, including photometric and spectroscopic estimates (Han et al. 1997; Davidge 2005; Gonçalves et al. 2007; Geha et al. 2010;



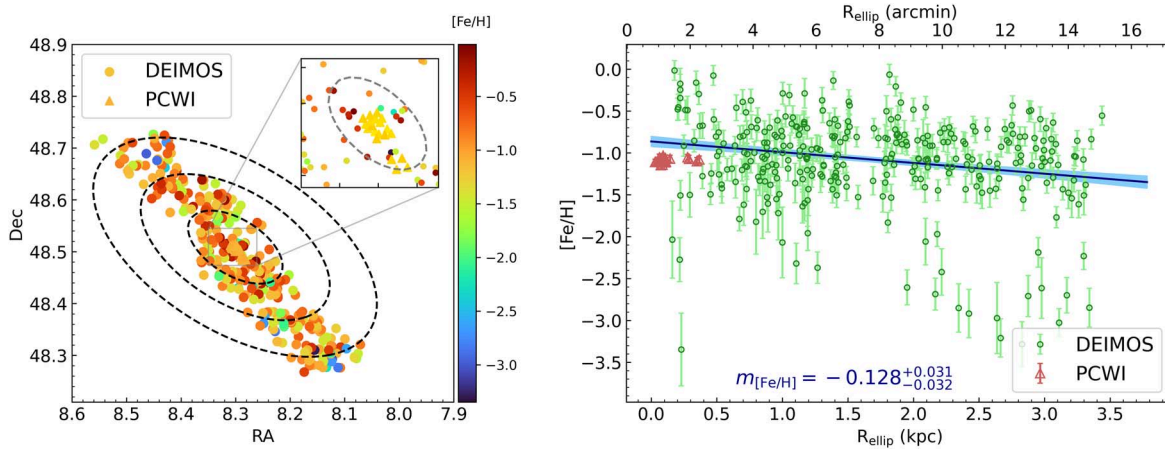


**Figure 5.** The spatially resolved measurements of SSP parameters for the central and south pointings. Left: The 2D maps of the best-fit stellar population age,  $[\text{Fe}/\text{H}]$ , and  $[\text{Mg}/\text{Fe}]$ . The contours indicate the (projected) galactocentric elliptical radii with 0.1 kpc intervals. Right: The radial gradients of the stellar population. The dashed lines represent the best-fit lines from MCMC. The gray shaded region indicates the ranges between the 16th and 84th percentiles.

Crnojević et al. 2014). We recovered similarly high values of  $[\text{Fe}/\text{H}]$  when we used the questionable photometry. We conclude that we should not compare our gradient to that measured by Vargas et al.

### 5.2. Outside-in Formation of NGC 147?

Stellar population gradients can provide us insight into the evolution of a galaxy. Observational studies have suggested



**Figure 6.** Left panel: The metallicity map for the resolved stars (circles) and integrated light (triangles). The black contours demonstrate the galactocentric radii at 4 kpc, 8 kpc, 12 kpc, and 16 kpc. The inset shows a zoomed-in map of the galaxy central part. The gray dashed ellipse is the maximum elliptical radius of the PCWI spatial coverage. The metallicities obtained from the integrated light are very close to those of the individual RGB stars. There is no prominent spatial variance of the stars’ metallicities. Right panel: The radial gradient of the metallicity. The best-fit linear gradient of the resolved stars’ metallicities is shown by the purple line. A negative slope of  $\sim -0.13$  dex  $\text{kpc}^{-1}$  is marginally detected.

that dSph<sup>10</sup> galaxies in the Local Group generally have positive age gradients and negative metallicity gradients (e.g., Kirby et al. 2011; Koleva et al. 2011). The observed gradients are usually explained as a result of outside-in formation (e.g., Benítez-Llambay et al. 2016; Revaz & Jablonka 2018; Genina et al. 2019), where the metals accumulate at the center of the galaxy, and the star formation becomes more concentrated with time. In particular, Genina et al. (2019) identified that ram pressure—imposed by the massive host galaxy when the satellite galaxy falls in—can compress the gas at the center of the satellite galaxy while stripping the gas from the outskirts. Consequently, star formation at the center can be reignited while the outer regions are quenched, giving rise to the gradients observed in Local Group dSphs. The age and metallicity gradients of NGC 147 found in this work are in alignment with the outside-in scenario, if the age gradient detected in the PCWI field is real.

As NGC 147 has no HI detection (Young & Lo 1997), tidal stripping and ram pressure stripping are very likely to have played an important role in quenching the star formation and shaping the observed radial gradients in this galaxy. Crnojević et al. (2014) uncovered the presence of extended tidal tails in NGC 147. The tails indicate a strong tidal interaction with its host galaxy, M31. Geha et al. (2006) found kinematic evidence for tidal disruption in another M31 dE satellite, NGC 205. Mayer et al. (2006) showed that the combined effects of tides and ram pressure are significantly more successful at removing gas from dwarfs than tidal stripping alone. In addition, Kormendy (1985; Kormendy & Bender 2012) also suggested that ram-pressure stripping would be one of the major quenching mechanisms for dSph galaxies in the Local Group.

## 6. Discussion

### 6.1. Comparison with Previous Work

So far, most works investigating the validity of recovering the stellar population, especially the metallicity, from integrated-light spectroscopy have studied relatively simple systems like globular clusters (e.g., González Delgado & Cid Fernandes 2010;

Barber et al. 2014; Conroy et al. 2018), which tend to have formed stars at the same epoch. Very few attempts have been made to directly compare stellar populations measured from integrated and resolved spectroscopic analyses for complex systems with multiple stellar populations.

Ruiz-Lara et al. (2018) obtained an integrated spectrum of the central part of Leo A, one of the Local Group’s star-forming dIrr galaxies with  $M_* \sim 3 \times 10^6 M_\odot$ . They derived the SFH and the metallicity of each age bin from the integrated spectrum. They found that while the SFH is consistent with that derived from CMD fitting, the derived metallicity at all age bins is higher than the average value of individual stars obtained by Kirby et al. (2017). The differences in metallicity range from  $\sim 0.25$  dex in the intermediate-age bins to  $> 1.0$  dex in younger-age bins, in contrast to our results for NGC 147.

However, such a discrepancy can be explained by factors other than the intrinsic systematic effects between the two techniques. First, Leo A is still star-forming and therefore has nebular emission. If the modeling of the Balmer emission lines from the ISM was not perfect, it could have affected the depths of the Balmer absorption lines, which heavily affect the measurement of the stellar population age. In turn, this could have affected the measured metallicity according to the age-metallicity degeneracy. As an old, quiescent galaxy, NGC 147 would not suffer the same source of uncertainty. Second, Leo A is metal-poor ( $[\text{Fe}/\text{H}] \sim -1.5$ ). Most stellar libraries used in the modeling of integrated-light spectra do not contain enough stars with  $[\text{Fe}/\text{H}] < -1$ . If the incompleteness of metal-poor stars in stellar libraries causes the discrepancy found by Ruiz-Lara et al. (2018), it will not affect the existing metallicities of more massive and more metal-rich galaxies like NGC 147 derived from integrated light. In addition, Leo A exhibits a stronger metallicity gradient ( $\sim -0.33$  dex  $\text{kpc}^{-1}$ ; Kirby et al. 2017) than NGC 147, which may explain the discrepancy between Kirby et al. (2017) and Ruiz-Lara et al. (2018). Kirby et al. (2017) measured stars at larger radii ( $\sim 7'$ ), whereas Ruiz-Lara et al. (2018) observed a total area of  $19''.2 \times 7''$  in the central region, corresponding to a region within a 0.15 half-light radius.

On the other hand, Boecker et al. (2020) recovered an age-metallicity distribution of M54 measured from integrated-light

<sup>10</sup> Some studies call fainter dEs spheroidal galaxies (Kormendy et al. 2009).

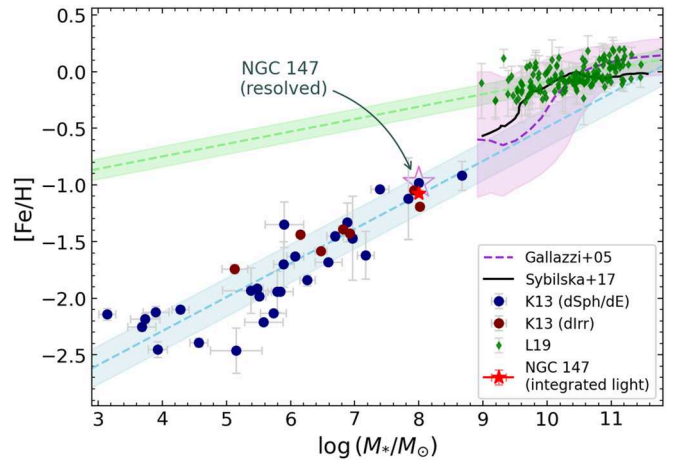
spectra that was consistent with resolved stellar spectra obtained on MUSE. Unlike most other star clusters, M54 has known multiple stellar populations formed at different epochs (Siegel et al. 2007; Alfaro-Cuello et al. 2019). Boecker et al. fit for a linear combination of SSP models to account for the SFH rather than assuming an SSP, and they found excellent agreement between the mass-weighted age and metallicity. The mass-weighted metallicity of the integrated-light spectra is 0.2 dex lower than the mean for the resolved stars, whereas the difference in the light-weighted metallicity is as small as 0.1 dex. The results are similar to ours, suggesting that taking the SFH into account or not does not significantly affect the recovery of metallicity from integrated spectra. However, their data set is slightly different from ours. First, a number of stars are resolved in the MUSE field of M54, so that the resolved stars would contribute greatly to the integrated-light spectra. Since the resolved stars are the brightest ones among the whole stellar population, one is more likely to obtain consistent light-weighted values from integrated and resolved analyses even though systematic offsets are present. Second, both integrated and resolved spectra are obtained on the same instrument; therefore, instrumental effects would be minimized. Our work moves one step toward validating the comparability of integrated-light spectra of distant galaxies to resolved spectra of local galaxies because we directly compare measurements from completely unresolved and completely resolved populations.

## 6.2. Implications for the MZR

Measuring the MZR across the full range of galaxy mass is extremely important to unveil how the physical processes regulating metal retention and interaction with the surrounding environment vary as the galaxies grow in mass. However, possible systematic effects between different methods of determining the metallicity obfuscate the exact form of the MZR.

Constructing the MZR from gas-phase metallicities has taught us an important lesson on understanding the role of systematic effects when measuring metallicities in different galaxy types and ages. Previous studies have found that different methods used to measure gas metallicity suffer from systematic offsets (e.g., Nagao et al. 2006; Kewley & Ellison 2008; Andrews & Martini 2013; Yates et al. 2020), and the discrepancy can be as large as 0.7 dex (Kewley & Ellison 2008). These systematic effects can lead to quantitatively and qualitatively different shapes for the gas-phase MZR. For instance, most studies favor a continuous gas-phase MZR that exhibits a single power law at low and intermediate mass, and gradually flattens at the high-mass end (e.g., Tremonti et al. 2004; Kewley & Ellison 2008; Zahid et al. 2013), while Blanc et al. (2019) discovered a sharp transition in the gas-phase MZR, which has a much steeper slope at  $10^{9.5}$ – $10^{10.5} M_{\odot}$ . Therefore, it is extremely important to understand the systematic offsets between different methods before placing the MZRs on the same absolute scale.

The literature discussion of systematic effects on galactic stellar metallicities is small relative to the extensive studies on the discrepancies of gas metallicity diagnostics. The different treatment exists partly because it is difficult to apply resolved stellar spectroscopy and integrated-light spectroscopy to the same galaxies. It is impossible to resolve RGB stars in massive galaxies beyond Andromeda, while most local dwarf galaxies are too large to fit in a long-slit spectrograph for integrated-light spectroscopy. So far, only Ruiz-Lara et al. (2018) have



**Figure 7.** The MZR measured from  $[\text{Fe}/\text{H}]$  at  $z \sim 0$ . This figure is the same as Figure 4, except that the low-mass MZR is obtained from dLrrs in addition to dSphs/dEs. The purple region shows the MZR of local Sloan Digital Sky Survey (SDSS) galaxies using spectrophotometric indices (Gallazzi et al. 2005). The weighted mean of the  $[\text{Fe}/\text{H}]$  of NGC 147 for resolved stars, which is highlighted by the purple star, is remeasured in this work using updated photometry (Section 2.2). The  $[\text{Fe}/\text{H}]$  of NGC 147 measured from integrated light (red star) falls on the low-mass MZR, indicating the systematic difference between the two techniques cannot explain the discrepancy in the MZRs between the two mass ranges.

compared the stellar metallicity of Leo A determined from integrated light to that obtained from resolved stars. As discussed in Section 6.1, the large discrepancy may result from factors other than systematic offsets. The agreement we obtained between resolved stars and integrated light in NGC 147 is encouraging because we adopted the same fitting algorithm used to derive the high-mass MZR (L19), which indicates relatively small systematic offsets in the stellar metallicity determination.

As seen in Figure 7, the high-mass MZR predicts an  $[\text{Fe}/\text{H}]$  at the mass of NGC 147 that is higher by 0.6 dex relative to the low-mass MZR. Our consistent resolved and integrated metallicity estimates tentatively suggest that the discrepancy between the low-mass MZR and high-mass MZR cannot be attributed to systematic effects intrinsic to the two techniques. We also exclude the possibility that the consistency in the two approaches is biased due to the SSP assumption in Appendix. Instead, our results support a real break related to physical processes governing the metal content. If the systematic offsets of the resolved and integrated-light methods are indeed as low as what we obtained for NGC 147, it is possible that the stellar MZR may also have a sharp transition, like the gas-phase MZR derived by Blanc et al. (2019). This break in the stellar MZR was also observed by Gallazzi et al. (2005), who measured the stellar MZR from a single method (spectrophotometric indices) rather than comparing MZRs measured from different methods in different mass ranges. However, the uncertainties of  $[\text{Fe}/\text{H}]$  below  $10^{10} M_{\odot}$  are too large to distinguish between the low-mass and high-mass MZRs.

A break in the stellar MZR was also reported by Panter et al. (2008), who measured the stellar MZR for a large sample of local SDSS galaxies above  $10^8 M_{\odot}$  via full-spectrum fitting and discovered a steeper slope for galaxies between  $10^8$  and  $10^{10} M_{\odot}$  than observed at higher mass. They suggested that the steeper slope was an artificial bias introduced by the limited aperture of the SDSS fibers, which cannot observe the outer region of low-mass galaxies. However, Sybilka et al. (2017)



measured the stellar metallicities for dwarf and giant quiescent galaxies from IFS data cubes obtained with the SAURON IFU, which suffers much less from aperture bias. A break was still present in their MZR in the same mass range, further suggesting that the break is very likely to be real.

The two MZRs may indicate a change in the feedback mechanisms as the galaxies grow in mass. Hayward & Hopkins (2017) proposed an analytic model to explain how stellar feedback simultaneously regulates star formation and drives outflows in a turbulent ISM. They argued that the mass-loading factor could decrease dramatically for galaxies above  $\sim 10^{10} M_{\odot}$ , so that high-mass galaxies become much less efficient in expelling the metal-enriched ISM. Because the stellar metallicity represents the metals locked inside the stars from the surrounding ISM averaged over the SFH, feedback changes can also affect the shape of the stellar MZR indirectly. Unfortunately, the current stellar MZR (Figure 7) has few individual precise measurements for intermediate-mass galaxies ( $10^8 < M_*/M_{\odot} < 10^{10}$ ), so we cannot determine how the shape of the stellar MZR would change across the full range of galaxy mass. Further studies focusing on the MZR of individual galaxies at intermediate mass are necessary to unveil the nature of the processes regulating the metal cycle and shaping the MZR. Nevertheless, the stellar MZRs of individual dwarf and massive galaxies can be a useful tool to help us understand the physical processes related to galactic chemical evolution. Our work presents the first evidence that it is reasonable to put the two MZRs on the same absolute scale to jointly constrain the shape of the stellar MZR.

Our investigation of systematic effects focused on just one galaxy: NGC 147. It is possible that our result is fortuitous and that a larger ensemble of galaxies would reveal some systematic differences between resolved spectroscopy and integrated light. Nonetheless, our results for NGC 147 suggest that these systematic effects do not play an important role in causing the discrepancy observed between different mass ranges of the stellar MZR. We anticipate observing more galaxies in the future to retire once and for all the possibility that the difference in technique—rather than in astrophysics—is responsible for the discrepancy.

## 7. Conclusions

We measured the stellar population age,  $[\text{Fe}/\text{H}]$ , and  $[\text{Mg}/\text{Fe}]$  of the central region ( $\lesssim 400$  pc, i.e.,  $2'$ ) in NGC 147 from integrated-light spectra obtained with PCWI. We also provided updated metallicity estimates for 317 RGB stars in the sample of K13. We compared the mean metallicity of the resolved stars with the measurement from integrated light to investigate systematic effects between the two techniques. A summary of our findings is as follows:

1. We recovered the global SSP-equivalent stellar population from integrated spectra and measured the stellar population age =  $8.11^{+0.74}_{-0.54}$  Gyr,  $[\text{Fe}/\text{H}] = -1.069^{+0.029}_{-0.031}$ , and  $[\text{Mg}/\text{Fe}] = 0.024^{+0.20}_{-0.20}$ . The iron and magnesium abundances are consistent with the best-fit parameters obtained from `alf`'s simple mode.
2. The best-fit SSP age is only 0.03 dex higher than the mass-weighted age determined from the CMD fitting of Weisz et al. (2014). The difference between  $[\text{Fe}/\text{H}]_{\text{SSP}}$  and the weighted mean of  $[\text{Fe}/\text{H}]_{\text{stars}}$  is less than 0.1 dex, implying that the SSP assumption is valid to recover the mean stellar

metallicity for a complex system with multiple stellar populations. It also suggests that the SSP assumption may recover the stellar age in some cases, although it strongly depends on the choice of SSP templates.

3. Making use of the IFS cubes, we detected a marginally positive age gradient in the central regions. Combining this age gradient with the negative metallicity gradient in the resolved stars, it is possible that NGC 147 formed outside in.
4. The resolved and integrated metallicities of NGC 147 are consistent within 0.1 dex, suggesting that systematic effects cannot explain the  $\sim 0.6$  dex discrepancy between the low-mass (K13) and high-mass (L19) stellar MZRs. Therefore, we tentatively conclude that the break in the MZRs at different mass ranges implies some physical mechanism transitions at  $10^8 M_{\odot} < M_* < 10^{10} M_{\odot}$ .

There is only one target in this work, so it is possible that such a small systematic offset ( $< 0.1$  dex) is underestimated. Nevertheless, our analysis of NGC 147 is strong evidence that the systematic effects of different techniques of determining metallicity will not be the major cause of discrepancy in the stellar MZRs. Future observations are necessary to expand this study to a larger sample in order to verify that the underlying astrophysical processes are ultimately responsible for the changing behavior of the stellar MZR.

The authors acknowledge the constructive feedback from the anonymous referee, which helped us improve the manuscript. We thank Yuguang Chen, Donal O'Sullivan, and James (Don) Neill for useful discussions on PCWI data reduction and covariance correction. We gratefully thank the staff at the Palomar Observatory, including support astronomers Carolyn Heffner and Kevin Rykoski, and telescope operator Kajsa Pfeffer, for assisting in the observations.

This material is based on work supported by the National Science Foundation (NSF) under grant No. AST-1847909. E.N.K. gratefully acknowledges support from a Cottrell Scholar Award administered by the Research Corporation for Science Advancement. M.A.C.d.I.R. acknowledges the financial support of the NSF Graduate Research Fellowship Program.

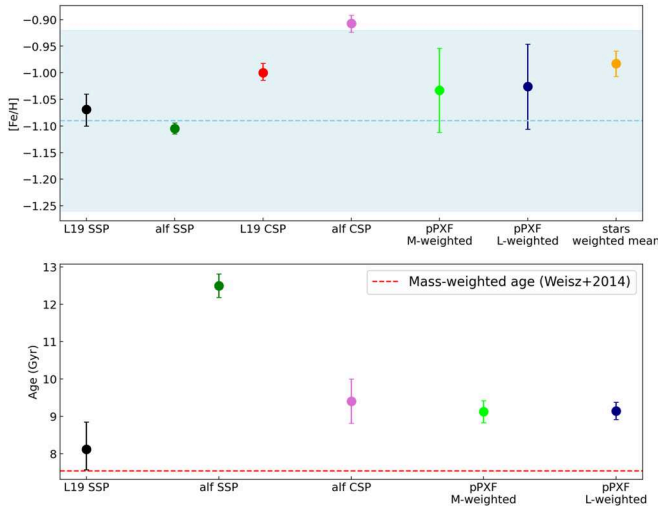
We are grateful to the many people who have worked to make the Keck Telescope and its instruments a reality and to operate and maintain the Keck Observatory. The authors wish to extend special thanks to those of Hawaiian ancestry, on whose sacred mountain we are privileged to be guests. Without their generous hospitality, none of the observations presented herein would have been possible. This research has made use of the Keck Observatory Archive, which is operated by the W. M. Keck Observatory and the NASA Exoplanet Science Institute, under a contract with the National Aeronautics and Space Administration.

*Facilities:* Hale (CWI), Keck II (DEIMOS).

*Software:* CWITools (O'Sullivan & Chen 2020), FSPS (Conroy et al. 2009), `alf` (Conroy et al. 2018), MPFIT (Markwardt 2012), Astropy (Astropy Collaboration et al. 2013, 2018), NumPy (Harris et al. 2020), SciPy (Virtanen et al. 2020), pPXF (Cappellari 2017).

## Appendix Comparison with CSP Models

In this section, we investigate whether fitting the data with more complicated CSP models would affect our primary results.



**Figure 8.** Comparison between the measured  $[\text{Fe}/\text{H}]$  (top panel) and stellar ages (bottom panel) in different stellar population models. Top: The blue dashed line and the shaded region indicate the  $[\text{Fe}/\text{H}]$  predicted by the low-mass MZR (K13) at the mass of NGC 147 and the corresponding  $1\sigma$  confidence interval. The data points show the measured  $[\text{Fe}/\text{H}]$  (the weighted mean for CSP models) via the various methods discussed in this work, which are all consistent with the low-mass MZR. Bottom: The measured ages (the weighted mean for CSP models) via the different methods discussed in this work. Among them, only the SSP age derived from L19’s method is consistent with the mass-weighted age obtained from the CMD fitting in Weisz et al. (2014) (red dashed line).

Because CSP models have more free parameters and thus require higher-quality data to achieve the same precision, we limit our analysis to the stacked spectrum, which has a very high S/N ( $\sim 111 \text{ \AA}^{-1}$ ). Here we present the measured ages and  $[\text{Fe}/\text{H}]$  from three different full spectral fitting codes that fit for multi-component stellar populations: L19’s method (Appendix A.1), *alf* (Appendix A.2), and PPXF (Cappellari 2017; Appendix A.3). As shown in Figure 8, the stellar metallicity estimates are not greatly affected by the choice of models, while the measured stellar ages are more strongly dependent on the assumed SFH. We find that taking the SFH into account leads to greater consistency between different methods and templates than assuming an SSP. Therefore, CSP models may be a more appropriate way to estimate the stellar age of complex systems when the data quality is good enough.

#### A.1. Comparison to the CSP Models in L19’s Fitting Code

To investigate whether adopting CSP models would affect the recovered metallicities, we fit for a multicomponent stellar population with ages and weights determined from the SFH measured from CMD fitting to HST photometry (Weisz et al. 2014). Instead of fitting for linear combinations of SSP models with different metallicities, we choose to unrealistically assume that all components have the same metallicity. This can reduce the degeneracy of free parameters and approximately estimate the mass-weighted metallicities to the greatest accuracy.

The best-fit metallicities are  $[\text{Fe}/\text{H}]_{\text{CSP}} = -1.000^{+0.018}_{-0.014}$  and  $[\text{Mg}/\text{Fe}]_{\text{CSP}} = 0.096^{+0.15}_{-0.37}$ . Compared to  $[\text{Fe}/\text{H}]_{\text{SSP}} = -1.069^{+0.029}_{-0.031}$  and  $[\text{Mg}/\text{Fe}]_{\text{SSP}} = 0.024^{+0.20}_{-0.20}$ , the best-fit metallicities derived from CSP models are higher and closer to the resolved metallicities, further diminishing the systematic offsets between the two techniques. The differences in the recovered  $[\text{Fe}/\text{H}]$  and  $[\text{Mg}/\text{Fe}]$  from SSP models and CSP models are within 0.1 dex; therefore, we conclude that metallicity measurements are

not affected significantly by the choices of SSP and CSP models, consistent with the results of Mentz et al. (2016) and Leethochawalit et al. (2018).

#### A.2. Comparison to the Results from *alf* Full Mode

In addition to SSP models in simple mode, *alf* provides a full mode that fits for two-component stellar populations with an older ( $>3$  Gyr) and a younger population ( $<3$  Gyr), which allows us to approximate the SFH of the galaxy. We fit the stacked spectrum in *alf* full mode.

The mass-weighted age derived in full mode (CSP-equivalent) is  $\sim 9.4$  Gyr. Compared to the SSP-equivalent age (12.5 Gyr) derived in simple mode, the difference between the CSP age and the mass-weighted age from CMD fitting by Weisz et al. (2014) is smaller. The consistency suggests that more complicated stellar population models are necessary to obtain more accurate stellar age measurements in galaxies. On the other hand, the best-fit results from *alf* full mode indicate that less than 0.5% of the stellar population was formed in the past 3 Gyr, while the SFH revealed by the CMD fitting in Weisz et al. (2014) suggests that  $\sim 20\%$  of stars were formed in the past 2.5 Gyr. We also notice that the CSP age is surprisingly younger than the SSP age, whereas Choi et al. (2014) found that SSP-equivalent ages are younger than both light- and mass-weighted ages derived in models with extended SFHs. The discrepancy may arise from the fact that *alf* assumes two independent starbursts at different epochs rather than extended SFHs. The disagreements above may also suggest that the recovered stellar age is sensitive to the adopted CSP models and their underlying assumptions on SFH, and thus future investigations are important to explore which CSP models are more suitable to measuring the stellar ages of galaxies.

Unlike the large difference seen in ages, the metallicities derived in the two modes are very similar.  $[\text{Mg}/\text{Fe}]_{\text{CSP}} = 0.064^{+0.018}_{-0.016}$  is consistent with  $[\text{Mg}/\text{Fe}]_{\text{SSP}}$  within  $1\sigma$ , while  $[\text{Fe}/\text{H}]_{\text{CSP}} = -0.907^{+0.015}_{-0.017}$  is slightly higher than  $[\text{Fe}/\text{H}]_{\text{SSP}}$ . The difference between  $[\text{Fe}/\text{H}]_{\text{CSP}}$  and  $[\text{Fe}/\text{H}]_{\text{stars}}$  is still within 0.1 dex, which once again demonstrates that metallicity measurements are less affected by the assumed stellar population models.

#### A.3. Comparisons to PPXF Full Spectral Fitting Code

Apart from the two fitting methods discussed in the main text, we test our CSP measurements by running another independent code, PPXF (Cappellari 2017), which fits for linear combinations of different SSPs to account for complex SFHs in galaxies. For simplicity,  $\alpha$ -enhancement is not included. We use the SSP models from the E-MILES library (Vazdekis et al. 2016), generated from the BaSTI isochrones (Pietrinferni et al. 2004) and Kroupa (2001) IMF. The metallicities and ages of the templates range from  $-2.27$  to  $0.4$  dex and from  $0.03$  to  $14$  Gyr, respectively. Because the SSP models assume solar-scaled abundances, the metallicity  $[\text{M}/\text{H}]$  can be interpreted as  $[\text{Fe}/\text{H}]$ . Regularization is used to address some of the deficiencies in PPXF: that it does not necessarily return unique solutions, and that it is very sensitive to noise in data due to the vast parameter space explored in fitting. The mass- and light-weighted stellar population parameters can be computed from the weights of each SSP component returned by PPXF.

The mass- and light-weighted  $[\text{Fe}/\text{H}]$  are  $-1.033 \pm 0.079$  and  $-1.026 \pm 0.08$ , respectively, consistent with the metallicities

measured from other methods. The uncertainties are calculated as the weighted variance of the ages and metallicities of each SSP component. On the other hand, the mass- and light-weighted ages are both  $9.96 \pm 0.01$  Gyr, higher than the SFH-based age measured by Weisz et al. (2014) but similar to the mass-weighted age derived in `alf` full mode, which also uses CSP models. The results suggest that stellar age estimates from integrated-light spectroscopy depend more heavily than metallicity estimates on the assumed SFH.

### ORCID iDs

Zhuyun Zhuang  <https://orcid.org/0000-0002-1945-2299>

Evan N. Kirby  <https://orcid.org/0000-0001-6196-5162>

Nicha Leethochawalit  <https://orcid.org/0000-0003-4570-3159>

Mithi A. C. de los Reyes  <https://orcid.org/0000-0002-4739-046X>

### References

- Akritas, M. G., & Bershad, M. A. 1996, *ApJ*, **470**, 706
- Alfaro-Cuello, M., Kacharov, N., Neumayer, N., et al. 2019, *ApJ*, **886**, 57
- Andrews, B. H., & Martini, P. 2013, *ApJ*, **765**, 140
- Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, *AJ*, **156**, 123
- Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, *A&A*, **558**, A33
- Barber, C., Courteau, S., Roediger, J. C., & Schiavon, R. P. 2014, *MNRAS*, **440**, 2953
- Benítez-Llambay, A., Navarro, J. F., Abadi, M. G., et al. 2016, *MNRAS*, **456**, 1185
- Blanc, G. A., Lu, Y., Benson, A., Katsianis, A., & Barraza, M. 2019, *ApJ*, **877**, 6
- Boecker, A., Alfaro-Cuello, M., Neumayer, N., Martín-Navarro, I., & Leaman, R. 2020, *ApJ*, **896**, 13
- Calura, F., Pipino, A., Chiappini, C., Matteucci, F., & Maiolino, R. 2009, *A&A*, **504**, 373
- Cappellari, M. 2017, *MNRAS*, **466**, 798
- Cappellari, M., & Copin, Y. 2003, *MNRAS*, **342**, 345
- Cappellari, M., Emsellem, E., Krajnović, D., et al. 2011, *MNRAS*, **413**, 813
- Carnall, A. C. 2017, arXiv:1705.05165
- Choi, J., Conroy, C., Moustakas, J., et al. 2014, *ApJ*, **792**, 95
- Choi, J., Dotter, A., Conroy, C., et al. 2016, *ApJ*, **823**, 102
- Conn, A. R., Ibata, R. A., Lewis, G. F., et al. 2012, *ApJ*, **758**, 11
- Conroy, C., Graves, G. J., & van Dokkum, P. G. 2014, *ApJ*, **780**, 33
- Conroy, C., Gunn, J. E., & White, M. 2009, *ApJ*, **699**, 486
- Conroy, C., Villaume, A., van Dokkum, P. G., & Lind, K. 2018, *ApJ*, **854**, 139
- Conroy, C., White, M., & Gunn, J. E. 2010, *ApJ*, **708**, 58
- Cooper, M. C., Griffith, R. L., Newman, J. A., et al. 2012, *MNRAS*, **419**, 3018
- Crnojević, D., Ferguson, A. M. N., Irwin, M. J., et al. 2014, *MNRAS*, **445**, 3862
- Davidge, T. J. 2005, *AJ*, **130**, 2087
- Dekel, A., & Silk, J. 1986, *ApJ*, **303**, 39
- Dotter, A. 2016, *ApJS*, **222**, 8
- Faber, S. M., Phillips, A. C., Kibrick, R. I., et al. 2003, *Proc. SPIE*, **4841**, 1657
- Finlator, K., & Davé, R. 2008, *MNRAS*, **385**, 2181
- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *PASP*, **125**, 306
- Gallazzi, A., Charlot, S., Brinchmann, J., White, S. D. M., & Tremonti, C. A. 2005, *MNRAS*, **362**, 41
- Geha, M., Guhathakurta, P., Rich, R. M., & Cooper, M. C. 2006, *AJ*, **131**, 332
- Geha, M., van der Marel, R. P., Guhathakurta, P., et al. 2010, *ApJ*, **711**, 361
- Geha, M., Weisz, D., Grocholski, A., et al. 2015, *ApJ*, **811**, 114
- Genina, A., Frenk, C. S., Benítez-Llambay, A., et al. 2019, *MNRAS*, **488**, 2312
- Gonçalves, D. R., Magrini, L., Leisy, P., & Corradi, R. L. M. 2007, *MNRAS*, **375**, 715
- González Delgado, R. M., & Cid Fernandes, R. 2010, *MNRAS*, **403**, 797
- Han, M., Hoessel, J. G., Gallagher, J. S., et al. 1997, *AJ*, **113**, 1001, 1
- Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, *Natur*, **585**, 357
- Hayward, C. C., & Hopkins, P. F. 2017, *MNRAS*, **465**, 1682
- Hopkins, P. F., Quataert, E., & Murray, N. 2012, *MNRAS*, **421**, 3522
- Kewley, L. J., & Ellison, S. L. 2008, *ApJ*, **681**, 1183
- Kirby, E. N., Cohen, J. G., Guhathakurta, P., et al. 2013, *ApJ*, **779**, 102
- Kirby, E. N., Gilbert, K. M., Escala, I., et al. 2020, *AJ*, **159**, 46
- Kirby, E. N., Guhathakurta, P., Bolte, M., Sneden, C., & Geha, M. C. 2009, *ApJ*, **705**, 328
- Kirby, E. N., Guhathakurta, P., Simon, J. D., et al. 2010, *ApJS*, **191**, 352
- Kirby, E. N., Guhathakurta, P., & Sneden, C. 2008, *ApJ*, **682**, 1217
- Kirby, E. N., Lanfranchi, G. A., Simon, J. D., Cohen, J. G., & Guhathakurta, P. 2011, *ApJ*, **727**, 78
- Kirby, E. N., Rizzi, L., Held, E. V., et al. 2017, *ApJ*, **834**, 9
- Koleva, M., Prugniel, P., De Rijcke, S., & Zeilinger, W. W. 2011, *MNRAS*, **417**, 1643
- Kormendy, J. 1985, *AJ*, **295**, 73
- Kormendy, J., & Bender, R. 2012, *ApJS*, **198**, 2
- Kormendy, J., Fisher, D. B., Cornell, M. E., & Bender, R. 2009, *ApJS*, **182**, 216
- Kroupa, P. 2001, *MNRAS*, **322**, 231
- Lee, H., Skillman, E. D., Cannon, J. M., et al. 2006, *ApJ*, **647**, 970
- Leethochawalit, N., Kirby, E. N., Ellis, R. S., Moran, S. M., & Treu, T. 2019, *ApJ*, **885**, 100
- Leethochawalit, N., Kirby, E. N., Moran, S. M., Ellis, R. S., & Treu, T. 2018, *ApJ*, **856**, 15
- Lequeux, J., Peimbert, M., Rayo, J. F., Serrano, A., & Torres-Peimbert, S. 1979, *A&A*, **500**, 145
- Magrini, L., Hunt, L., Galli, D., et al. 2012, *MNRAS*, **427**, 1075
- Marigo, P., & Girardi, L. 2007, *A&A*, **469**, 239
- Markwardt, C. 2012, MPFIT: Robust non-linear least squares curve fitting, ascl:1208.019
- Martin, N. F., Ibata, R. A., McConnachie, A. W., et al. 2013, *ApJ*, **776**, 80
- Mayer, L., Mastroiello, C., Wadsley, J., Stadel, J., & Moore, B. 2006, *MNRAS*, **369**, 1021
- McClure, R. D., & van den Bergh, S. 1968, *AJ*, **73**, 313
- Mentz, J. J., La Barbera, F., Peletier, R. F., et al. 2016, *MNRAS*, **463**, 2819
- Murray, N., Quataert, E., & Thompson, T. A. 2005, *ApJ*, **618**, 569
- Nagao, T., Maiolino, R., & Marconi, A. 2006, *A&A*, **459**, 85
- Newman, J. A., Cooper, M. C., Davis, M., et al. 2013, *ApJS*, **208**, 5
- O'Sullivan, D., & Chen, Y. 2020, arXiv:2011.05444
- O'Sullivan, D. B., Martin, C., Matuszewski, M., et al. 2020, *ApJ*, **894**, 3
- Panther, B., Jimenez, R., Heavens, A. F., & Charlot, S. 2008, *MNRAS*, **391**, 1117
- Pietrinferni, A., Cassisi, S., Salaris, M., & Castelli, F. 2004, *ApJ*, **612**, 168
- Revaz, Y., & Jablonka, P. 2018, *A&A*, **616**, A96
- Ruiz-Lara, T., Gallart, C., Beasley, M., et al. 2018, *A&A*, **617**, A18
- Sánchez-Blázquez, P., Peletier, R. F., Jiménez-Vicente, J., et al. 2006, *MNRAS*, **371**, 703
- Siegel, M. H., Dotter, A., Majewski, S. R., et al. 2007, *ApJL*, **667**, L57
- Soderblom, D. R. 2010, *ARA&A*, **48**, 581
- Sybilski, A., Lisker, T., Kuntschner, H., et al. 2017, *MNRAS*, **470**, 815
- Tremonti, C. A., Heckman, T. M., Kauffmann, G., et al. 2004, *ApJ*, **613**, 898
- Vargas, L. C., Geha, M. C., & Tollerud, E. J. 2014, *ApJ*, **790**, 73
- Vazdekis, A., Koleva, M., Ricciardelli, E., Röck, B., & Falcón-Barroso, J. 2016, *MNRAS*, **463**, 3409
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *Nature Methods*, **17**, 261
- Weisz, D. R., Dolphin, A. E., Skillman, E. D., et al. 2014, *ApJ*, **789**, 147
- Yates, R. M., Schady, P., Chen, T. W., Schweyer, T., & Wiseman, P. 2020, *A&A*, **634**, A107
- Young, L. M., & Lo, K. Y. 1997, *ApJ*, **476**, 127
- Zahid, H. J., Geller, M. J., Kewley, L. J., et al. 2013, *ApJL*, **771**, L19