The Astrophysical Journal, 968:24 (37pp), 2024 June 10 https://doi.org/10.3847/1538-4357/ad39e5

PHANGS-ML: Dissecting Multiphase Gas and Dust in Nearby Galaxies Using Machine Learning

Dalya Baron[1,33], Karin M. Sandstrom[2], Erik Rosolowsky[3], Oleg V. Egorov[4], Ralf S. Klessen[5,6], Adam K. Leroy[7,8], Médéric Boquien[9], Eva Schinnerer[10], Francesco Belfiore[11], Brent Groves[12], Jérémy Chastenet[13], Daniel A. Dale[14], Guillermo A. Blanc[1,15], José E. Méndez-Delgado[4], Eric W. Koch[16], Kathryn Grasha[17,18,34], Mélanie Chevance[19,20,32], David A. Thilker[21], Dario Colombo[22,23], Thomas G. Williams[24], Debosmita Pathak[7], Jessica Sutter[2], Toby Brown[25], John F. Wu[26,27], Josh E. G. Peek[26,27], Eric Emsellem[28,29], Kirsten L. Larson[30], and Justus Neumann[31]
[1] The Observatories of the Carnegie Institution for Science, 813 Santa Barbara Street, Pasadena, CA 91101, USA; dalyabaron@gmail.com
[2] Department of Astronomy & Astrophysics, University of California, San Diego. 9500 Gilman Drive, La Jolla, CA 92093, USA
[3] Department of Physics, University of Alberta, Edmonton, Alberta, T6G 2E1, Canada
[4] Astronomisches Rechen-Institut, Zentrum für Astronomie der Universität Heidelberg, Mönchhofstraße 12-14, 69120 Heidelberg, Germany
[5] Universität Heidelberg, Zentrum für Astronomie, Institut für Theoretische Astrophysik, Albert-Ueberle-Straße 2,69120 Heidelberg, Germany
[6] Universität Heidelberg, Interdisziplinäres Zentrum für Wissenschaftliches Rechnen, Im Neuenheimer Feld 205, 69120 Heidelberg, Germany
[7] Department of Astronomy, Ohio State University, 180 W. 18th Avenue, Columbus, OH 43210, USA
[8] Center for Cosmology and Astroparticle Physics, 191 West Woodruff Avenue, Columbus, OH 43210, USA
[9] Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, Laboratoire Lagrange, 06000, Nice, France
[10] Max Planck Institute for Astronomy, Königstuhl 17, D-69117, Germany
[11] INAF—Osservatorio Astrofisico di Arcetri, Largo E. Fermi 5, I-50157, Firenze, Italy
[12] International Centre for Radio Astronomy Research, University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia
[13] Sterrenkundig Observatorium, Ghent University, Krijgslaan 281-S9, 9000 Gent, Belgium
[14] Department of Physics and Astronomy, University of Wyoming, Laramie, WY 82071, USA
[15] Departamento de Astronomía, Universidad de Chile, Camino del Observatorio 1515, Las Condes, Santiago, Chile
[16] Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, 02138 Cambridge, MA, USA
[17] Research School of Astronomy and Astrophysics, Australian National University, Canberra, ACT 2611, Australia
[18] ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Australia
[19] Zentrum für Astronomie der Universität Heidelberg, Institut für Theoretische Astrophysik, Albert-Ueberle-Str. 2, 69120 Heidelberg, Germany
[20] Cosmic Origins Of Life (COOL) Research DAO, Germany
[21] Center for Astrophysical Sciences, Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218, USA
[22] Argelander-Institut für Astronomie, Universität Bonn, Auf dem Hügel 71, 53121 Bonn, Germany
[23] Max-Planck-Institut für Radioastronomie, Auf dem Hügel 69, 53121 Bonn, Germany
[24] Sub-department of Astrophysics, Department of Physics, University of Oxford, Keble Road, Oxford OX1 3RH, UK
[25] Herzberg Astronomy and Astrophysics Research Centre, National Research Council of Canada, 5071 West Saanich Road, Victoria, BC, V9E 2E7, Canada
[26] Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA
[27] Department of Physics & Astronomy, Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218, USA
[28] European Southern Observatory, Karl-Schwarzschild-Straße 2, 85748, Garching, Germany
[29] University Lyon, Univ Lyon1, ENS de Lyon, CNRS, Centre de Recherche Astrophysique de Lyon UMR5574, 69230, Saint-Genis-Laval, France
[30] AURA for the European Space Agency (ESA), Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA
[31] Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany
Received 2024 February 2; revised 2024 March 21; accepted 2024 April 1; published 2024 June 5

Abstract

The PHANGS survey uses Atacama Large Millimeter/submillimeter Array, Hubble Space Telescope, Very Large Telescope, and JWST to obtain an unprecedented high-resolution view of nearby galaxies, covering millions of spatially independent regions. The high dimensionality of such a diverse multiwavelength data set makes it challenging to identify new trends, particularly when they connect observables from different wavelengths. Here, we use unsupervised machine-learning algorithms to mine this information-rich data set to identify novel patterns. We focus on three of the PHANGS-JWST galaxies, for which we extract properties pertaining to their stellar populations; warm ionized and cold molecular gas; and polycyclic aromatic hydrocarbons (PAHs), as measured over 150 pc scale regions. We show that we can divide the regions into groups with distinct multiphase gas and PAH properties. In the process, we identify previously unknown galaxy-wide correlations between PAH band and optical line ratios and use our identified groups to interpret them. The correlations we measure can be naturally explained in a scenario where the PAHs and the ionized gas are exposed to different parts of the same radiation field that varies spatially across the galaxies. This scenario has several implications for nearby galaxies: (i) The uniform PAH ionized fraction on 150 pc scales suggests significant self-regulation in the interstellar medium, (ii) the PAH 11.3/7.7 μm band ratio may be used to constrain the shape of the non-ionizing far-ultraviolet

[32] coolresearch.io
[33] Carnegie-Princeton Fellow.
[34] ARC DECRA Fellow.

to optical part of the radiation field, and (iii) the varying radiation field affects line ratios that are commonly used as PAH size diagnostics. Neglecting this effect leads to incorrect or biased PAH sizes.

Unified Astronomy Thesaurus concepts: Astrostatistics techniques (1886); Astronomy data visualization (1968); Warm ionized medium (1788); Interstellar dust (836); Polycyclic aromatic hydrocarbons (1280)

1. Introduction

Over the past several decades, astronomy has been going through a data revolution. It was pioneered by the Sloan Digital Sky Survey, which imaged roughly one-third of the sky in five photometric bands, providing measured photometry for billions of objects and spectroscopy for millions (York et al. 2000; Eisenstein et al. 2011; Dawson et al. 2013). Since then, past and ongoing surveys have been producing massive data sets that include millions to billions of objects with astrometric, photometric, spectroscopic, or time-series observations (e.g., Pan-STARRS; Zwicky Transient Facility; Gaia; APOGEE; MANGA; DESI; Kaiser et al. 2010; Bellm 2014; Bundy et al. 2015; DESI Collaboration et al. 2016; Gaia Collaboration et al. 2016; Majewski et al. 2017; Abdurro'uf et al. 2022; DESI Collaboration et al. 2023; Gaia Collaboration et al. 2023). These observations, along with numerous derived products from them, were made publicly accessible through efficient and convenient interfaces and changed the way astronomers interact with observations, marking the beginning of the big data era in astronomy. In the near future, surveys by the Vera Rubin Observatory, Roman Space Telescope, Euclid, SDSS-V, and the Square Kilometre Array (e.g., Dewdney et al. 2009; Kollmeier et al. 2017; Ivezić et al. 2019; Euclid Collaboration et al. 2022), to name a few, are expected to make another order-of-magnitude increase in data volume.

The big data era in astronomy is not only characterized by an increase in data volume, related to the total number of observed sources, but is also characterized by an increase in data complexity, which is related to the increased information content of a single astronomical source. With numerous surveys conducted using different telescopes and instruments, astronomical sources today often have multiwavelength observations, from radio to X-ray, and in some wave bands, also as a function of time. In the nearby Universe, over the past several decades, surveys have mapped tens to hundreds of nearby galaxies from ultraviolet (UV) to radio using imaging and spectroscopy (e.g., SINGS; KINGFISH; THINGS; xCOLD GASS; Kennicutt et al. 2003; Dale et al. 2007; Walter et al. 2008; Moustakas et al. 2010; Kennicutt et al. 2011; Saintonge et al. 2011; Dale et al. 2017; Saintonge et al. 2017). Examples of galaxy clusters include the multiwavelength mapping of the Virgo cluster galaxies (e.g., Côté et al. 2004; Boselli et al. 2011, 2018; Brown et al. 2021). Outside the local Universe, there are several multiwavelength surveys that mapped galaxies at different redshifts in deep fields (e.g., GOODS; COSMOS; Ferguson et al. 2000; Giavalisco et al. 2004; Scoville et al. 2007).

The increase in data complexity raises some challenges, but also presents some new opportunities. On the one hand, it raises the question of how to incorporate the different types of observations, each with different sensitivities, spatial and spectral resolutions, and noise properties, within the same framework, in a sufficiently general manner to be applied to a variety of astronomical objects. In addition, the high dimensionality of the data makes it challenging to identify trends, especially when they tie observables from different wavelengths or instruments. On the other hand, the increase in information content offers the unique opportunity to use the data itself to form novel hypotheses, a core approach in the field of data science.

In the more traditional, model-driven or physics-driven approach, a scientific study starts with a hypothesis. Observations are planned and conducted to test the hypothesis, and their analysis leads to new insights, often resulting in new hypotheses. In data science, various statistical tools are used to visualize and dissect the high-dimensional space spanned by the data set. When the information content of the data is large, such tools may uncover previously unknown trends or groups of objects, allowing one to form hypotheses directly from the data. Since this process relies less on a physical model or on prior knowledge, it may lead to unexpected discoveries.

With the advent of multiwave band opportunities, surveys have been producing larger and more complex data sets. The Physics at High Angular resolution in Nearby GalaxieS (PHANGS; Leroy et al. 2021a; Emsellem et al. 2022; Lee et al. 2022, 2023) survey is an example of a modern astronomical survey that pushes the limits of data complexity. With the goal of constraining the physics near or at the molecular cloud scale, the survey has been making high-resolution observations of nearby galaxies across the electromagnetic spectrum, utilizing various telescopes and instruments. Nineteen of the PHANGS galaxies have high-resolution maps obtained with the following telescopes: Atacama Large Millimeter/submillimeter Array (ALMA; mm interferometry; Leroy et al. 2021a), Hubble Space Telescope (HST; UV and optical imaging; Lee et al. 2022), JWST (near-infrared and mid-infrared imaging; Lee et al. 2023), and Very Large Telescope (VLT; optical integral field spectroscopy with MUSE; Emsellem et al. 2022). Each of the galaxies has 10^5–10^7 independent spatial resolution elements,[35] with each pixel/spaxel probing the conditions of multiphase gas, dust, and stars on scales of 5–100 pc (see Figure 1). The unique combination of spectral coverage and high spatial resolution makes the information content of a single PHANGS galaxy comparable to that of the Legacy SDSS spectroscopic survey when considering the number of spectra and spectral resolution elements.

The high information content of the PHANGS galaxies makes it an ideal data set for applications of data-science tools. In this pilot study, our goal is to test whether unsupervised machine-learning algorithms can be used to identify previously unknown trends or groups in the data. Such tools have been applied to astronomical data sets in various contexts (see reviews and references in Baron 2019, hereafter B19; and Fluke & Jacobs 2020), with the most relevant and recent examples being (i) identification of underlying correlations in the high-dimensional data of molecular cloud populations constructed from multiwavelength observations from PHANGS (Sun et al. 2022), and (ii) the application of a

[35] The number of independent resolution elements depends on the spatial resolution, with $\sim 10^5$ spaxels for MUSE and ALMA, $\sim 10^6$ pixels for JWST MIRI, and $\sim 10^7$ pixels for HST and JWST NIRCam.

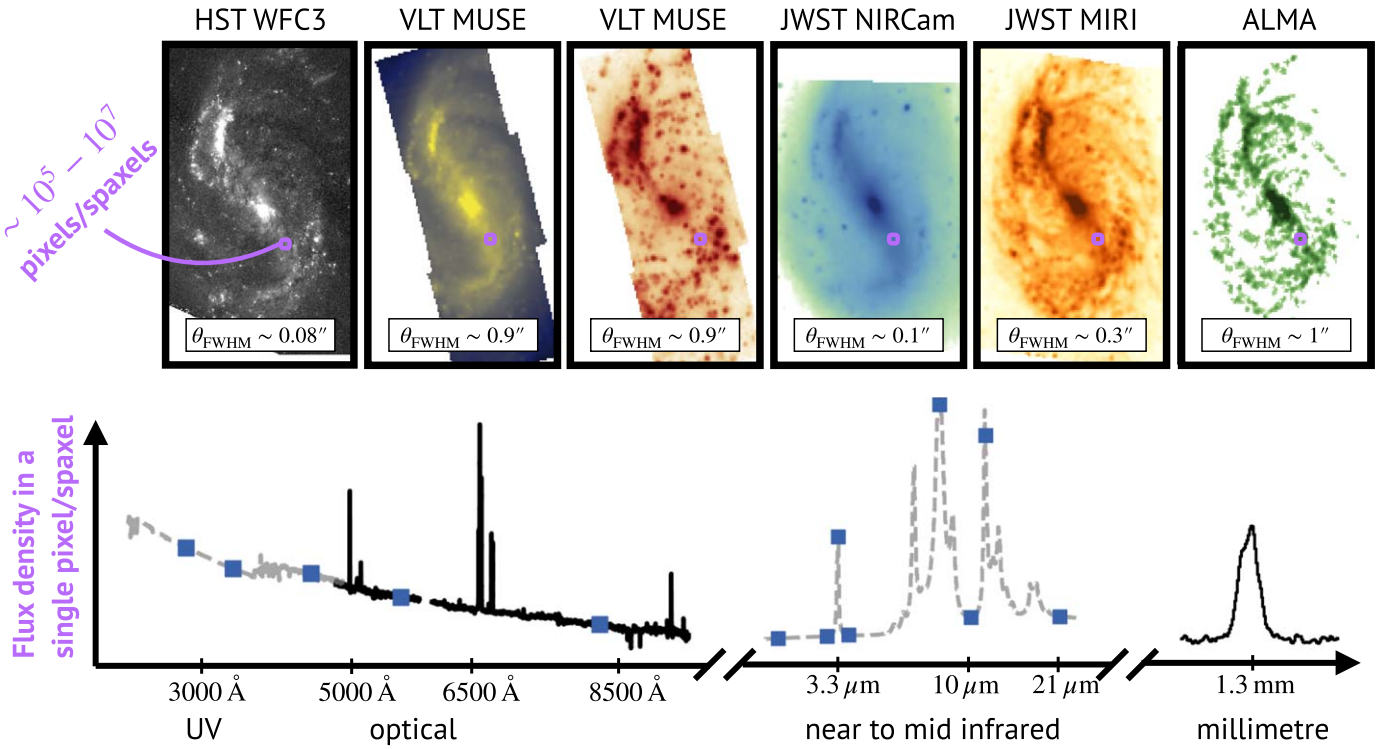

Figure 1. *The information content in a single PHANGS galaxy.* The diagram uses data obtained for NGC 7496 to illustrate the multiscale multiwavelength information available as part of the PHANGS survey. The top row shows images of the galaxy obtained using: HST-WFC3 (F336W filter), VLT-MUSE (stellar continuum emission and Hα surface brightness), JWST-NIRCam (F200W filter), JWST-MIRI (F770W filter), and ALMA. The different images have different spatial resolutions, ranging from $\sim 0\farcs08$ to $\sim 1''$, as indicated at the bottom of each image. Each PHANGS galaxy has 10^5–10^7 pixels/spaxels. The multiwavelength information available in each pixel/spaxel is shown in the bottom row, and it includes several UV-optical photometric bands by HST, a full optical spectrum between 4800 Å and 9300 Å by MUSE, eight JWST near and mid-infrared photometric bands, and a spectrum of the CO line reconstructed from millimeter interferometric observations by ALMA. The black solid lines represent observed spectra, blue rectangles represent photometry, and gray dashed lines represent the expected underlying spectrum. For presentation purposes, the flux densities in each wavelength region have been stretched vertically to cover the entire panel. In this work, we use a set of physically motivated properties measured from these observations.

clustering algorithm to JWST observations of polycyclic aromatic hydrocarbon (PAH) emission in the Orion Bar (Pasquini et al. 2023).

In this work, we estimate properties related to the stellar populations, multiphase gas conditions, and dust properties, on scales of 150 pc. We then use dimensionality reduction and clustering algorithms to divide these 150 pc sized regions into groups. In the process, we identify new relations between PAHs and the warm ionized gas, and use our defined clusters to interpret these relations. This work is therefore complementary to recent studies that explicitly study the connection between PAHs, the ionized gas, and the radiation field, in the PHANGS-JWST galaxies, using a more physics-driven approach. In particular, Egorov et al. (2023) study the PAH-to-total dust fraction and its relation to the strength of the radiation field, parameterized using the gas ionization parameter, in H II regions. Dale et al. (2023) constrain different PAH properties, such as their size and charge distribution, in stellar clusters, while exploring the impact of changing radiation fields. Chastenet et al. (2023a) and J. Sutter et al. (2024, in preparation) study how the PAH fraction depends on local and global conditions, including the phase of the ISM, specific star formation rate, metallicity, and stellar mass.

In Section 2, we describe the data we use in our analysis. In Section 3, we describe our approach, which consists of three main steps: feature construction (Section 3.1), dimensionality reduction (Section 3.2), and clustering (Section 3.3). The resulting clusters, their interpretation, and the new relations found between PAHs and the warm ionized gas are presented in Section 4. The results section stands on its own and does not require a deep understanding of the methods. Therefore, readers who are interested in the results may skip Section 3 and go directly to Section 4. In Section 5, we discuss possible extensions and generalizations of our methodology. We summarize and conclude in Section 6.

2. Data

To study the interplay between interstellar medium (ISM) gas and dust, star formation, and the observed stellar populations, we use multiwavelength observations by ALMA, VLT-MUSE, JWST-NIRCam, and JWST-MIRI, obtained as part of the PHANGS survey (Leroy et al. 2021a; Emsellem et al. 2022; Lee et al. 2023). The first PHANGS-JWST data release includes fully reduced and calibrated NIRCam and MIRI broadband images of the three galaxies: NGC 0628, NGC 1365, and NGC 7496 (e.g., Lee et al. 2023 and references therein). The broadband images have been extensively tested and analyzed in a series of recent works (e.g., Sandstrom et al. 2023a; Chastenet et al. 2023b; Sandstrom et al. 2023b; Belfiore et al. 2023; Dale et al. 2023; Egorov et al. 2023; Leroy et al. 2023), making them an ideal benchmark for applications of machine-learning algorithms. We therefore focus on these three galaxies here.

Since our goal is to combine information from different instruments (ALMA, MUSE, and JWST NIRCam and MIRI),

Table 1
PHANGS-Two-dimensional Galaxy Properties

(1) Galaxy	(2) D (Mpc)	(3) i (deg)	(4) SFR ($M_\odot$ yr^{-1})	(5) $\log M_*$ ($\log M_\odot$)	(6) AGN?	(7) FWHM (arcsec)	(8) $N_{\rm pix,in}$	(9) $N_{\rm pix,samp}$	(10) $N_{\rm pix,final}$
NGC 0628	9.84	9	1.75	10.2	no	3.″14	2,560,000	40,000	6,387
NGC 1365	19.57	55	16.90	10.8	yes	1.″58	346,710	38,523	12,565
NGC 7496	18.72	36	2.26	9.8	yes	1.″65	99,000	24,750	5,055

Note. (1)–(5) Galaxy properties from Lee et al. (2023): name, distance, inclination, star formation rate (SFR), and stellar mass. (6) Indicator of AGN presence. (7) Effective spatial resolution of the convolved multiwavelength maps. (8) Number of pixels in the initial ALMA WCS grid of the galaxy. (9) Number of pixels in the grid after downsampling the ALMA grid to have at least two pixels per 150 pc. (10) Number of pixels that are not masked out in the pixel mask. The machine-learning algorithms are applied to these sets of pixels.

each with a different spatial resolution, and to consider pixels from different galaxies within the same analysis, we have to ensure that the pixels trace information originating from the same physical scale. We therefore use PHANGS data products that are based on maps convolved to a resolution of 150 pc, as described in each of the subsections below. We list the effective angular resolution (″) of the convolved maps for each of the galaxies in Table 1.

We project all the convolved maps into the world coordinate system (WCS) defined by the ALMA observations using REPROJECT.EXACT by ASTROPY (Robitaille et al. 2020; Astropy Collaboration et al. 2022). We then downsample the R.A.–decl. coordinate grid to have two pixels per resolution element of 150 pc in each of the galaxies.[36] We list the initial number of pixels in the ALMA observation grid and the number of pixels after the downsampling in Table 1.

As an illustration, we show different properties of NGC 1365 derived from the ALMA, MUSE, and JWST maps in Figure 2. These include properties pertaining to the cold molecular gas, warm ionized gas, stellar populations, PAHs, and large dust grains. In the rest of the section, we describe the data products we use and their analysis.

2.1. ALMA

To trace the cold molecular gas properties, we use the PHANGS-ALMA survey (Leroy et al. 2021a, 2021b). The survey mapped the ^{12}CO(2 − 1) (CO hereafter) line emission at a spatial resolution of $\sim 1'' \sim 100$ pc in 90 nearby galaxies. After imaging, the cubes are convolved to a succession of physical resolutions. We use the PHANGS-ALMA data products obtained after convolving the cubes to a spatial resolution of 150 pc.

The catalog includes two main types of products. The "strict" mask products include two-dimensional maps generated from the data cubes after applying stringent signal identification criteria. Since these require a detection of the signal with high confidence, these maps have low noise, but they include less of the total CO flux and are thus somewhat incomplete. The "broad" mask products include two-dimensional maps generated using all the sight lines where signal is identified at any resolution, making them more complete than the "strict" mask products. However, since these include regions with faint emission, they appear noisier and can contain false positives. In Figure 2, we show the CO moment 0 (total intensity) derived using the "strict" and "broad" masks, as well as the CO effective width ($W_{\rm eff}$, a line width measure) calculated using the "strict" masks. Since our analysis requires high completeness and can tolerate some additional noise, we use the CO flux derived using the "broad" mask in our feature extraction (see details in Section 3.1).

2.2. MUSE

To trace the warm ionized gas conditions and the stellar population properties, we use the PHANGS-MUSE survey (Emsellem et al. 2022). This survey mapped 19 of the PHANGS-ALMA galaxies with the integral field spectrograph MUSE at a spatial resolution of $\sim 1''$. The data analysis pipeline of the survey includes (i) mosaicking and homogenization of individual MUSE pointings for a given galaxy, (ii) performing spatial binning of the spectra to reach sufficient signal-to-noise ratio (S/N) for stellar population synthesis modeling, (iii) fitting the stellar continuum to extract the stellar kinematics, reddening, and stellar populations, and (iv) fitting the optical emission lines to extract gas kinematics and line fluxes. In this work, we use various properties (see below) derived from the MUSE cubes after they have been convolved to a resolution of 150 pc (PHANGS DR2.2 release).

For the warm ionized gas properties, we use the surface brightness maps of the Balmer lines Hβ and Hα, and the following emission lines: [O III]λ 5007Å, [O I]λ 6300Å, [N II] λ 6584Å, and [S II]$\lambda\lambda$ 6717Å + 6731Å ([O III], [O I], [N II], and [S II] hereafter). These are estimated by fitting the spectra in the convolved cubes with a set of Gaussians (see details in Emsellem et al. 2022). To estimate the dust-corrected Hα surface brightness, we assume case-B recombination ($T = 10^4$ K), a dusty-screen, and the Cardelli et al. (1989) extinction law, with the color excess given by the following:

$$E(B-V) = 2.33 \times \log\left[\frac{(\mathrm{H}\alpha/\mathrm{H}\beta)_{\rm obs}}{2.86}\right] \mathrm{mag}, \quad (1)$$

where $(\mathrm{H}\alpha/\mathrm{H}\beta)_{\rm obs}$ is the observed Hα/Hβ surface brightness ratio. We then correct the observed Hα surface brightness using the derived $E(B-V)$ values.

In our analysis, we consider the dust-corrected Hα surface brightness, the Hα gas velocity dispersion, and the line ratios log([O III]/Hβ), log([N II]/Hα), log([S II]/Hα), and log([O I]/Hα), which are typically used to constrain the main source of ionizing radiation. These line ratios are based on lines close in wavelength, so they are nearly reddening independent. We thus do not correct the lines for reddening prior to the line ratio estimation. We do correct the Hα surface brightness for

[36] The pixels in the downsampled grid therefore trace distances of 76.3, 85.4, and 89.6 pc, for NGC 0628, NGC 1365, NGC 7496, respectively. We ensured that using only one pixel per resolution element of 150 pc results in a low-dimensional embedding comparable to that which we obtain with two pixels.

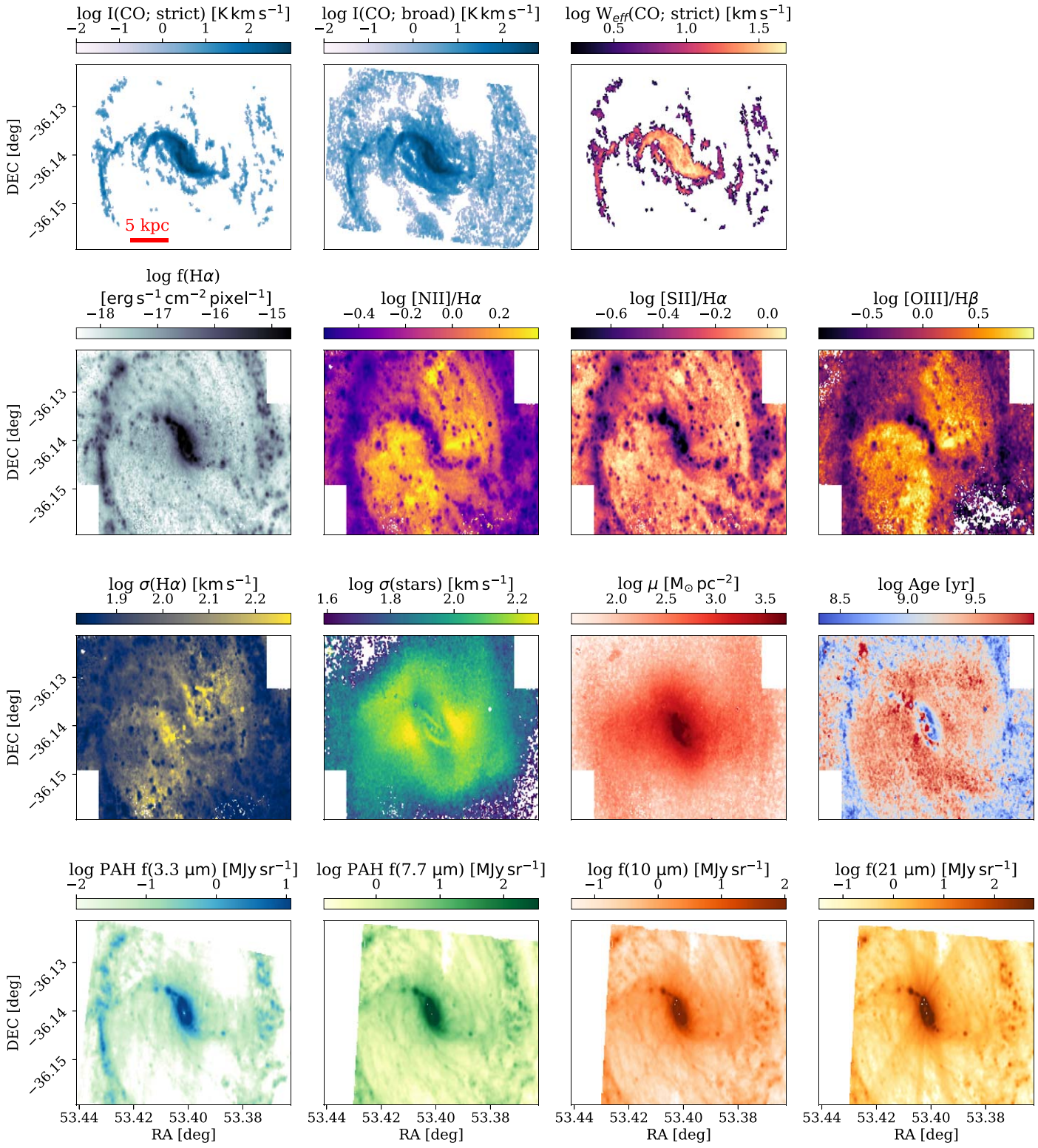

Figure 2. Different properties derived from the ALMA, MUSE, and JWST observations of NGC 1365. The first row shows cold molecular gas properties from ALMA: ^{12}CO(2 – 1) moment 0 (intensity) derived using the "strict" (lower noise but less complete) and "broad" (noisier and more complete) masks, and the CO effective width derived using the "strict" masks. The second and third rows show gas and stellar population properties from MUSE: dust-corrected Hα surface brightness, the line ratios log([N II]/Hα), log([S II]/Hα), and log([O III]/Hβ), Hα velocity dispersion, stellar velocity dispersion, stellar mass surface density, and the age of the stellar population. The fourth row shows dust grain properties from JWST: PAH 3.3 and 7.7 μm emission features, and broad-band mid-infrared emission from hot dust at 10 and 21 μm. The 21 μm image is saturated in the center of the galaxy and shows noticeable diffraction spikes. All the maps are convolved to a common resolution of 150 pc.

reddening since our derived features in Section 3.1 include ratios of Hα to CO or PAH emission. We propagate the surface brightness uncertainties to the feature uncertainties, and pixels in which these properties are measured with S/N <3 masked out (see discussion about feature missingness in Section 5.2).

For the stellar population properties, we use maps of the stellar velocity dispersion, stellar mass density, mass-weighted stellar age, mass-weighted stellar metallicity, and reddening toward the stars. These properties were derived using stellar population synthesis fits of binned stellar spectra, using Voronoi bins of the convolved MUSE cubes (Emsellem et al. 2022; Pessa et al. 2023).

2.3. JWST

To trace PAH properties and dust-reprocessed stellar or active galactic nucleus (AGN) light, we use the Cycle 1 PHANGS-JWST survey data (Lee et al. 2023; Williams et al. 2024). It is a JWST treasury survey aimed at collecting imaging data in eight bands from 2 to 21 μm of the 19 galaxies observed as part of PHANGS-ALMA, PHANGS-MUSE, and PHANGS-HST. The broadband images include the four NIRCam bands F200W, F300M, F335M, and F360M, probing near-infrared emission at 2, 3, 3.35, and 3.6 μm respectively, and the four MIRI bands F770W, F1000W, F1130W, and F2100W, probing mid-infrared emission at 7.7, 10, 11.3, and 21 μm. These filters are designed to cover different PAH emission bands, which are expected to be sensitive to different grain size and charge distribution, to capture dust continuum emission, and the silicate 9.7 μm absorption feature.

We use the same version of the surface brightness maps presented and described in the PHANGS-JWST Cycle 1 Focus Issue[37] (version 0.8; see, e.g., Sandstrom et al. 2023a; Chastenet et al. 2023b; Sandstrom et al. 2023b; Belfiore et al. 2023; Dale et al. 2023; Lee et al. 2023). These maps are convolved to a spatial resolution of 150 pc using the approach described in Aniano et al. (2011). Williams et al. (2024) describes the full data reduction pipeline.

To study the PAH properties, we use the MIRI F770W and F1130W broadband images, which are expected to be dominated by the 7.7 and 11.3 μm PAH features. We also use the NIRCam bands F300M, F335M, and F360M, and follow the procedure outlined by Sandstrom et al. (2023a) to estimate the 3.3 μm PAH flux ($F335M_{PAH}$ hereafter). This procedure uses the F300M and F360M bands to subtract the starlight contribution from the F335M band, thus isolating the flux from the 3.3 μm PAH feature.

For the dust continuum emission, we considered both the MIRI F1000W and F2100W filters, although both suffer from different limitations. The F1000W filter shows strong correlations with the PAH-tracing filters F770W and F1130W, and a weaker correlation with the hot dust-tracing filter F2100W, suggesting that the filter is in fact dominated by PAH emission in a large fraction of the pixels (e.g., Belfiore et al. 2023; Leroy et al. 2023). The F2100W filter traces only dust continuum emission, but it is saturated in the central pixels of NGC 1365 and NGC 7496, presenting significant diffraction spikes (see Figure 2 and Chastenet et al. 2023b). Since using the F2100W filter would require us to mask out the saturated galaxy centers and the diffraction spikes, we do not include it in our feature construction. We do include the F1000W filter, but note that it traces PAH emission much more than the hot dust continuum. We use both F1000W and F2100W in our interpretation of the resulting clusters and trends.

[37] https://iopscience.iop.org/collections/2041-8205_PHANGS-JWST-First-Results

Following the papers in the Focus Issue, we use the convolved maps of NGC 7496 to estimate the noise level in all the relevant bands ($F335M_{PAH}$, F770W, F1130W, F1000W, F2100W).[38] NGC 7496 is the only target with sufficiently empty space that is not contaminated by emission from the source. The estimated noise rms levels, in units of MJy sr^{-1}, are 0.0071, 0.021, 0.023, 0.013, and 0.082, for the $F335M_{PAH}$, F770W, F1130W,F1000W, F2100W bands, respectively. We use these values to mask out pixels in which the measured flux is lower than 3 times the noise rms in all three galaxies.

3. Methods

In this section, we apply unsupervised machine-learning algorithms to the maps constructed from the ALMA, MUSE, and JWST observations. These algorithms are used to divide the pixels from the different galaxies into groups according to their multiwavelength properties, where pixels in a given group show distinct values or correlations between their stellar, gas, and dust properties, from pixels in other groups. This allows us to explore the complex multiwavelength PHANGS data set without an initial model-driven hypothesis. Instead, we form data-driven hypotheses by inspecting the resulting groups and identifying previously unknown trends and correlations within them.

In our analysis, each pixel from each of the galaxies is considered as a separate *object* with a set of measured *features*. The measured features are constructed from the multiwavelength maps, and they trace the stellar, gas, and dust properties within each of the pixels. The features do not include information related to the galaxy a pixel belongs to, or its location within the galaxy, although both are used when interpreting the results. Once a final list of objects with measured features is constructed, we apply a dimensionality reduction algorithm to this data. The output of the dimensionality reduction is an embedding of the high-dimensional data into a two-dimensional space, where every object (pixel) is represented by two numbers. We then apply a clustering algorithm to the distribution of the objects in the two-dimensional space, which allows us to assign a class to each of the objects. Therefore, this three-step procedure divides the PHANGS pixels into groups according to their multiwavelength observations.

We start by describing our feature construction scheme in Section 3.1, which includes our definition of features, and their scaling and normalization. Since the data contains a non-negligible fraction of nondetections, the section also describes our adopted pixel masking strategy aimed at minimizing missing feature values, and our strategy to handle the remaining missing values. In Section 5.2, we discuss the issue of missing values more generally, and propose methods to handle nondetections and upper limits in future works. In Section 3.2, we apply the dimensionality reduction algorithm uniform manifold approximation and projection (UMAP) and obtain a two-dimensional embedding of the input data set. We also discuss the impact of different hyperparameter choices on the resulting two-dimensional embedding. In Section 3.3, we apply several different clustering algorithms to

[38] Our estimated noise levels of the convolved, lower-resolution maps are consistent with those estimated by Lee et al. (2023) and Chastenet et al. (2023b) for the high-resolution maps given the convolution kernel size.

Table 2
Selected Features

Feature	Unit	Missing Fraction after Pixel Mask
logf(Hα)/I(CO)	log(erg s^{-1} cm^{-2} pixel^{-1}/K km s^{-1})	...
logf(Hα)/f(10 μm)	log(erg s^{-1} cm^{-2} pixel^{-1}/MJy sr^{-1})	...
logI(CO)/f_{PAH}(7.7 μm)	log(K km s^{-1}/MJy sr^{-1})	...
logf([N II])/f(Hα)	...	0.033%
logf([S II])/f(Hα)	...	...
logf([O I])/f(Hα)	...	0.36%
logf([O III])/f(Hβ)	...	1.53%
logAge	log(yr)	...
gas E($B-V$)	mag	...
logf_{PAH}(3.3 μm)/f_{PAH}(7.7 μm)	...	...
logf_{PAH}(3.3 μm)/f_{PAH}(11.3 μm)	...	...
logf_{PAH}(11.3 μm)/f_{PAH}(7.7 μm)	...	...
log (f_{PAH}(7.7 μm) + f_{PAH}(11.3 μm))/f(10 μm)	...	...
log σ(Hα) *	log(km s^{-1})	0.066%
log μ *	log($M_\odot$ pc^{-2})	0.12%
log σ(Hα)/σ(stars) **	...	12.4%

Note. Summary of the selected features, their units, and their missing fraction *after* applying the pixel mask. The data was first scaled, then clipped, and then normalized. All features except the gas $E(B-V)$ are scaled logarithmically. Then, all features are clipped to be between the 0.5th and 99.5th percentiles of the distribution. All the features are then normalized by subtracting the mean and dividing by the standard deviation of the clipped distribution. The last three features were excluded from the data after some initial tests: *, these features dominated the two-dimensional embedding and resulted in a trivial embedding with three clusters corresponding to the three different galaxies; **, high fraction of missing values and little impact on the resulting embedding.

the two-dimensional embedding and define the final clusters that will be used to group the pixels.

3.1. Feature Extraction

We use the convolved, reprojected, and downsampled maps obtained from the ALMA, MUSE, and JWST observations described in Section 2. Since our goal is to study the interplay between stellar, gas, and dust properties, we require >3σ detection of the Hα and CO emission lines, the PAH bands, and the dust continuum emission.[39] For that requirement, we construct a pixel mask map for each of the galaxies. In each pixel mask map, a pixel value is set to 1 if this pixel is not masked out in *every one* of the following maps: dust-corrected Hα surface brightness, CO intensity using the "broad" mask, F335M$_{PAH}$, F770W, and F1000W. Otherwise, the pixel value is set to 0. Since the Hα surface brightness is dust corrected, requiring a 3σ detection ensures the detection of the weaker Hβ line. While this requirement does not ensure the detection of the [O III], [N II], [S II], and [O I] lines, their detection fraction is quite high (see missing fractions in Table 2).

We list the number of pixels in the pixel mask map in Table 1. The included pixels are also marked with colors in Figure 3 in the results. They constitute only 15%, 32%, and 20%, of all the pixels in the downsampled maps of NGC 0628, NGC 1365, and NGC 7496, respectively. There are two main factors contributing to the large number of masked-out pixels. The first is the incomplete overlap between the fields of view of the different instruments (see, e.g., Figure 1 in Lee et al. 2023). The second factor is our requirement of CO and PAH detection, which masks out around 30% of the pixels (see, e.g., Figure 3). These pixels are masked out because the CO and PAH 3.3 μm emission are below the sensitivity limit.

Among the maps we consider in our masking, the dust-corrected Hα and the 7.7 and 10 μm surface brightness maps are quite complete, with a small fraction of masked-out pixels. The CO and 3.3 μm maps have a larger fraction of masked-out pixels, and they typically have the same pixels masked out in both. Since we are interested in studying multiphase gas and dust, which cannot be done without a CO and 3.3 μm PAH emission detection,[40] we choose to exclude such pixels from the analysis. Our choice to use the CO emission identified using the "broad" mask is motivated by the CO detection requirement, since the CO line is detected in a larger fraction of the pixels compared to that of the "strict" mask. If instead we had used the CO intensity identified using the "strict" mask, we would have had to mask out approximately 80%–90% of the pixels in each galaxy, excluding most of the diffuse ISM. In Section 5.2, we discuss the implications of excluding such pixels from the analysis and propose possible methods to include upper limits and nondetections in future analyses of this kind.

Table 2 summarizes the features we considered, most of which are surface brightness *ratios*. Although the machine-learning algorithms we use can in principle be applied to measured surface brightness values directly, we choose to work with ratios as we find them more easy to interpret. The first three features, Hα to CO, Hα to 10 μm, and CO to 7.7 μm PAH, trace the interplay between star formation (traced by Hα), different gas phases (traced by Hα and CO), and dust grains. We also consider the warm ionized gas line ratios log([N II]/Hα), log([S II]/Hα), log([O I]/Hα), and log([O III]/Hβ) as features, as these are known to be sensitive to the source of ionizing radiation through standard Baldwin,

[39] The stellar population properties are measured using high-S/N binned spectra, and their measured uncertainties are generally smaller than 3 times the measured value. Therefore, the stellar properties are quite complete throughout the field of view.

[40] We require a detection of the 3.3 μm PAH feature as this band is particularly useful when constraining the PAH size distribution. See Section 4.2 for additional details.

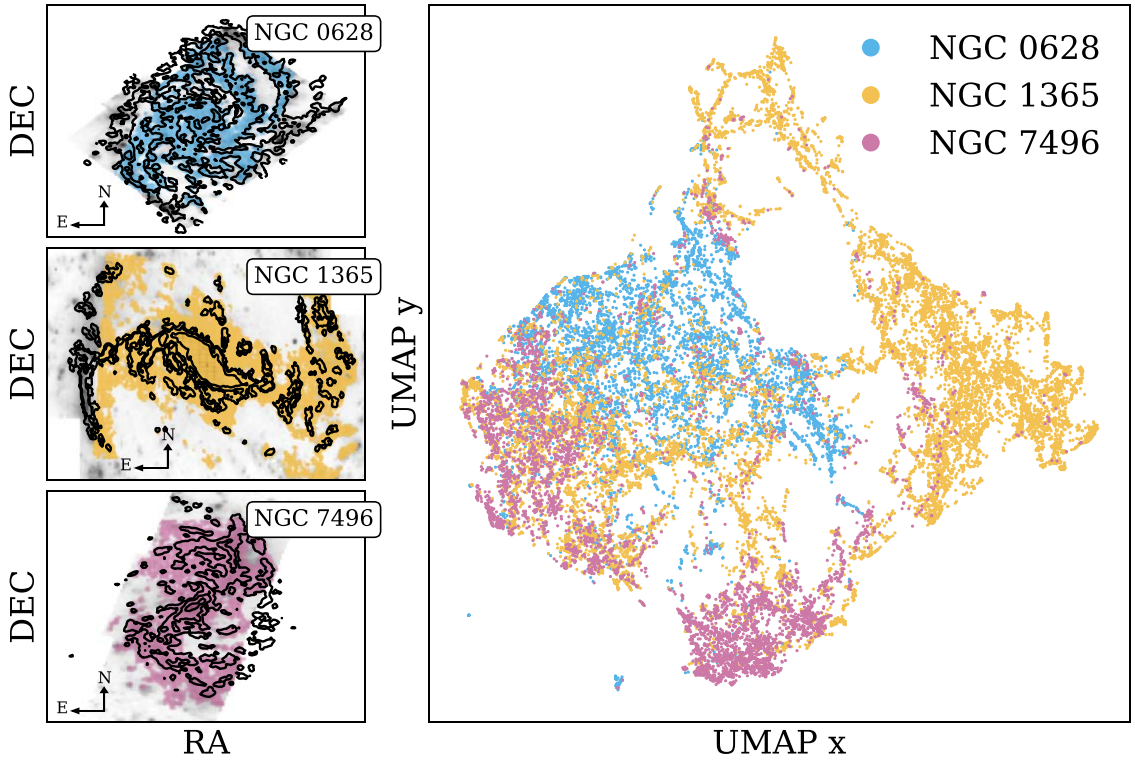

Figure 3. UMAP dimensionality reduction of the PHANGS multiwavelength pixels. The left panels show the three galaxies NGC 628, NGC 1365, and NGC 7496, where pixels that are included in the analysis are marked with colors. In these panels, the gray-scale background represents the Hα surface brightness, and the black contours represent the CO intensity. The right panel shows our adopted UMAP embedding. Every point corresponds to an object (pixel) from our input data set, which includes 24,007 objects with 13 features each (features trace 150 pc size regions). The two-dimensional embedding was obtained using the following hyperparameters: `metric=correlation`, `n_neighbors=10`, and `min_dist=0`, although we show in Appendix A that the global structure in the two-dimensional space remains stable when changing the metric and the `n_neighbors` parameter. Each pixel is colored according to the galaxy it belongs to, information that is not included in the input data set. The embedding shows several overdense regions that may be interpreted as clusters, connected through filamentary structures of points, suggesting that the features form continuous relations in the complex high-dimensional space.

Phillips, and Terlevich (1981; BPT) diagrams (Baldwin et al. 1981; Veilleux & Osterbrock 1987; Kewley et al. 2001; Kauffmann et al. 2003), as well as to metallicity, ionization parameter, and more (e.g., Kewley et al. 2019).

We consider properties pertaining to the dynamics and stellar population traced by the MUSE observations: the ionized gas velocity dispersion measured with the Hα, the gas-to-stellar velocity dispersion ratio, the stellar mass surface density, the age of the stellar population, and the reddening toward the line-emitting gas (see Equation (1)). Three of these features were later excluded from the analysis—(i) the Hα velocity dispersion, (ii) and the stellar mass surface density, since they dominated the dimensionality reduction and resulted in a trivial embedding where the pixels were divided into three clusters according to the galaxy they belong to,[41] even after correcting for galaxy inclination, and (iii) the gas to stellar velocity dispersion ratio, which had a large fraction of missing values and had little impact on the low-dimensional embedding.

We include several flux ratios that trace different PAH properties. The 3.3 μm PAH feature primarily traces small and neutral PAHs, while the 7.7 μm feature traces larger and positively charged ions, and the 11.3 μm feature represents grains that are larger and neutral (e.g., Boersma et al. 2016, 2018; Maragkoudakis et al. 2020; Draine et al. 2021; Rigopoulou et al. 2021; Maragkoudakis et al. 2022). Therefore, the 3.3/11.3 μm and 3.3/7.7 μm PAH ratios are sensitive to the PAH size distribution, although they are also sensitive to the shape of the far-ultraviolet (FUV)-optical radiation field (see, e.g., Draine et al. 2021; and Appendix C.1). The 11.3/7.7 μm PAH ratio is primarily sensitive to the PAH charge distribution, although also to the shape of the radiation. To trace the PAH abundance, several recent studies defined $R_{PAH} = (F770W + F1130W)/F2100W$, which is a ratio of PAH to dust-continuum flux (e.g., Chastenet et al. 2023a, 2023b; Egorov et al. 2023; J. Sutter et al. 2024, in preparation). Since the F2100W is saturated in the centers of NGC 1356 and NGC 7496, we use an alternative feature, (F770W + F1130W)/F1000W, where F1000W is used to trace the hot dust continuum. However, we found that the (F770W + F1130W)/F1000W feature does not correlate with R_{PAH}, most likely since the F1000W band is dominated by PAH emission, rather than by hot dust continuum emission. In addition, the band may also be affected by 9.7 μm silicate absorption (e.g., Smith et al. 2007).

Most of the features we consider are distributed over several orders of magnitude. In such a case, any dimensionality reduction algorithm that starts by measuring pairwise distances between the features will be dominated by features with larger

[41] The clipping and normalization we perform (see below) to the features should have, in principle, ensured that the feature values do not differ significantly between the different galaxies. In practice, however, these two features showed non-Gaussian distributions with significant tails, which affected the normalization and resulted in nonstandard distributions that are different for the different galaxies. Since these tails correspond to robust measurements that represent extreme dynamical environments, we do not clip them. To include such features in future studies, it might be necessary to perform histogram equalization prior to the clipping and normalization.

values (see, e.g., B19). To give even weights to small and large feature values, we use a logarithmic scaling of the features. That is, the Hα to CO flux ratio feature is represented by $\log[f(\mathrm{H}\alpha)/I(\mathrm{CO})]$ rather than $f(\mathrm{H}\alpha)/I(\mathrm{CO})$. The only exception is the feature $E(B-V)$, which is not distributed over several orders of magnitude.

We then apply clipping and normalization to the scaled features. For each feature, we clip its values to be between the 0.5th and 99.5th percentiles of the distribution. The clipping ensures that catastrophic outliers do not have a significant impact on the normalization of the feature, which is the next and final stage of the feature construction. These outliers are mostly the result of problems in observations, their reduction, or the estimation of a feature value from them. By definition, the fraction of objects with clipped feature values is very low, and thus, the clipping of their values does not have an impact on the two-dimensional embedding of the rest of the objects. Indeed, we found the two-dimensional embedding by UMAP to be stable to different clipping schemes, ranging from (0.1%, 99.9%) to (1%, 99%). However, even a low fraction of catastrophic outliers can have a significant impact on the estimated standard deviation of a feature, which can affect the normalization of the whole feature, and thus the UMAP embedding. It is therefore essential to perform some clipping before the normalization.

Finally, each scaled and clipped feature x is normalized as $(x-\mu_x)/\sigma_x$, where μ_x is the average feature value, and σ_x is the standard deviation. The normalization is done to ensure that the dimensionality reduction will not be dominated by features with a large dynamical range (e.g., see discussion in B19).

Since we exclude pixels where the CO and/or 3.3 μm PAH are not detected, our analysis is complete in H II regions and much of the dense ISM as these show brighter CO and mid-infrared emission, but is incomplete in the most diffuse part of the ISM, where CO and mid-infrared emission are much fainter. Therefore, our results may not be applicable to the most diffuse parts of local galaxies. In Table 2, we list the fraction of missing values in the features we consider after applying the pixel mask. Due to our pixel mask, the fraction of missing values is very low. We compared two different imputation methods to replace these missing values (see additional details in Section 5.2 in the discussion): K nearest neighbor (KNN) search (Hruschka et al. 2003; Jonsson & Wohlin 2004) and random forest regression (Stekhoven & Bühlmann 2011; Shah et al. 2014) by SKLEARN (Pedregosa et al. 2011). The two methods gave comparable results, and neither had a significant impact on the resulting two-dimensional embedding. The results shown in Section 4 are based on a data set where the missing values have been replaced using the KNNImputer.[42]

3.2. Dimensionality Reduction with UMAP

In this section, we use the dimensionality reduction algorithm UMAP (McInnes et al. 2018) to embed our input data set into a two-dimensional space.

UMAP is a nonlinear dimensionality reduction algorithm that aims to represent high-dimensional data in a lower-dimensional space while preserving the underlying structure and relationship among data points. It operates on the assumption that the data lie on a manifold, which is a low-dimensional curved surface embedded within the high-dimensional space. It attempts to learn an approximation of this manifold and then project the data onto the lower-dimensional space. UMAP has been shown to preserve local as well as global structures in the data, and has been widely used in a variety of fields (see, e.g., Becht et al. 2018; Ali et al. 2019; Cao et al. 2019; Carter et al. 2019; Packer et al. 2019).

The two main use cases of UMAP are (i) visualization of complex high-dimensional data sets, and (ii) dimensionality reduction prior to the application of clustering algorithms. The latter, which is also the use case in this work, is done because many clustering algorithms do not scale well with a large number of features (e.g., Xu & Tian 2015), and thus, UMAP is used as an intermediate stage to reduce the initial dimensions of the data set. This intermediate step also improves the interpretability of the final clustering output, as the clusters and their properties can be easily visualized in the two-dimensional space given by UMAP.

UMAP has several hyperparameters, and setting different values of the hyperparameters can change the resulting embedding significantly. The first, and probably most important, hyperparameter is `metric`, which defines the distance metric to be used when estimating distances between the objects in the high-dimensional space. Different metrics are sensitive to different aspects and scales in the feature space. As a result, while one metric may suggest proximity between two objects, another may suggest a considerable separation (see discussion in B19). Given that UMAP relies on pairwise distances between objects for the embedding process, the selection of an appropriate metric becomes crucial.

The second most important hyperparameter of UMAP is `n_neighbors`, which controls how the algorithm balances local versus global structure when learning the manifold of the data (see examples in McInnes et al. 2018). Setting a low value of `n_neighbors` will result in an embedding that is more sensitive to local structure of the data, sometimes at the expense of accurately representing the global structure. On the other hand, increasing the value of `n_neighbors` expands the neighborhoods considered when estimating the manifold, which may result in the loss of fine-grained details.

The third hyperparameter, `min_dist`, controls how tightly points can be packed in the low-dimensional space. Larger values of `min_dist` will force even close neighbors to be separated in the low-dimensional embedding, while lower values will result in clumpier embeddings.

Following the feature construction scheme in Section 3.1, the input data set has 24,007 objects with 13 features each. Each object represents a pixel in one of the three galaxies we consider, and the 13 features (listed in Table 2) trace different properties related to the stellar population, gas, dust, and star formation, over a 150 pc scale.

We applied UMAP to our input data set while exploring a wide range of hyperparameter choices. In particular, we examined the two-dimensional embedding using 11 different distance metrics, and using a wide range of `n_neighbors` values (see Appendix A).[43] We find that the global structure of the data in the two-dimensional space remains stable when

[42] https://scikit-learn.org/stable/modules/generated/sklearn.impute.KNNImputer.html

[43] One could, in principle, apply UMAP to reduce the dimensions to three or four, and then apply clustering to the lower-dimensional embedding. We find it challenging to visualize and interpret the low-dimensional embedding for three or four dimensions, and to identify a suitable set of hyperparameters for the clustering algorithms. We therefore only explore two-dimensional embeddings.

changing the metric and/or n_neighbors. We reach a similar conclusion for the min_dist parameter, which primarily changes the density of points. This suggests that the clusters that would have been identified in the two-dimensional embeddings by UMAP with different hyperparameters should be roughly the same. Therefore, the results presented in Section 4 should not depend significantly on the assumed hyperparameters.

For the rest of the paper, we adopt the two-dimensional UMAP embedding obtained using metric=correlation, n_neighbors=10, and min_dist=0. This set of hyperparameters was adopted after a visual inspection of the resulting embeddings, where we selected an embedding where we expect cluster identification to be less challenging technically, as we describe below. However, since different sets of hyperparameters result in embeddings that satisfy the criteria described below, the selection of this particular set of hyperparameters is somewhat arbitrary.

To select an embedding where identifying clusters is expected to be less challenging, we favor embeddings where clusters that we identify by eye are more separated from each other, and are separated by roughly the same distances, one from the other. In addition, we prefer embeddings where the density of points in different clusters does not vary significantly, and try to avoid embeddings with a large number of filamentary structures, as these tend to result in clusters that do not match our visual perception (many clustering algorithms are designed to work effectively in a flat geometry, and filamentary structures are a strong departure from this assumption).

In Figure 3, we show our adopted two-dimensional UMAP embedding, where every pixel is colored according to the galaxy it belongs to. In Figure 4, we show our adopted embedding color-coded by the features in the input data set, where strong gradients can be seen in many of the features throughout the embedding.

3.3. Clustering

This section includes the final phase of our three-step process, where we apply a clustering algorithm to the two-dimensional embedding by UMAP to divide the PHANGS pixels into groups. To partition the two-dimensional distribution, we examine several different clustering algorithms: K-means, Birch, OPTICS, DBSCAN, and hierarchical clustering (see Xu & Tian 2015; and B19 for reviews about clustering algorithms). The different algorithms have different optimization objectives and different stopping criteria, making them sensitive to different aspects of the data. For example, by construction, K-means identifies evenly sized clusters and can only operate in a flat geometry, while OPTICS and DBSCAN may identify unevenly sized clusters and can operate in a nonflat geometry. The output of hierarchical clustering depends on the assumed linkage method, which defines how clusters are linked to different clusters. We considered the four linkage methods: Ward, complete, average, and single, each being sensitive to different types of structures (see Figure 11 in B19 and related text).

Each of the clustering algorithms has different hyperparameters, and changing the values of these parameters can have a significant impact on the resulting clusters (see, e.g., B19). For example, in K-means, Birch, and hierarchical clustering, the number of clusters is a hyperparameter of the algorithm and is set by the user. When applying clustering algorithms directly to high-dimensional data sets, it may be challenging to select the ideal number of clusters,[44] since it is not straightforward to visualize the detected clusters in the high-dimensional space. In our case, since the clustering algorithms are applied to the two-dimensional embedding by UMAP, we were able to visualize the resulting clusters, color-coded by different features (Figure 4), and select a suitable number of clusters. The chosen numbers, which were selected to not be larger than $\sim$10, so we can inspect the clusters manually and compare between their properties, and not smaller than $\sim$4, so that objects with different properties will not be grouped together, are six for K-means and hierarchical clustering, and seven for Birch.

While K-means, Birch, and hierarchical clustering divide all the points into clusters, OPTICS and DBSCAN may divide only some points into clusters, leaving others unclustered. The hyperparameters of OPTICS and DBSCAN primarily control the minimum cluster membership and neighborhood size (see, e.g., Xu & Tian 2015), and we set these parameters to result in roughly six–eight clusters. For OPTICS, we used the hyperparameters (min_samples, xi, min_cluster_size) to be (300, 0.001, 0.04). For DBSCAN, we set the hyperparameters (min_samples, min_cluster_size) to be (10, 300). We used the python SKLEARN library to apply these different clustering algorithms (Pedregosa et al. 2011).

Figure 5 shows the different clustering algorithms applied to our adopted two-dimensional embedding by UMAP. One of the methods, hierarchical clustering with the single linkage, can be excluded immediately as it clusters the majority of the points into a single cluster, and marks five very small outlier groups as the remaining clusters. We also exclude OPTICS and DBSCAN as they leave a significant fraction of points unclustered, and we wish to include as many objects as possible in the analysis. Next, we exclude K-means, Birch, and hierarchical clustering with the complete linkage as they cluster together groups that appear separate upon visual inspection (yellow group in K-means, yellow group in Birch, and pink and orange groups in hierarchical clustering). This leaves hierarchical clustering with Ward linkage versus the average linkage, both of which roughly agree on three of the clusters (topmost, rightmost, and bottommost), but disagree about the division of points in the leftmost overdensity of points. Our choice to adopt the average linkage was motivated by the distributions of features seen in Figure 4, where it is apparent that the average linkage divides the points in the leftmost region into points with lower PAH 3.3/11.3 μm ratios and points with higher ratios. The results reported in Section 4 are therefore based on the clusters identified with hierarchical clustering with the average linkage method.

4. Results

We apply dimensionality reduction and clustering algorithms to the PHANGS multiwavelength data set of NGC 0628, NGC 1365, and NGC 7496, and divide the pixels into six groups. This data set is constructed from maps extracted from ALMA, MUSE, and JWST observations, and it traces the

[44] There are various suggested techniques to select a suitable number of clusters automatically, but these techniques are usually tailored to a specific clustering algorithm, and cannot be applied broadly and generally with all considered clustering algorithms (e.g., Xu & Tian 2015).

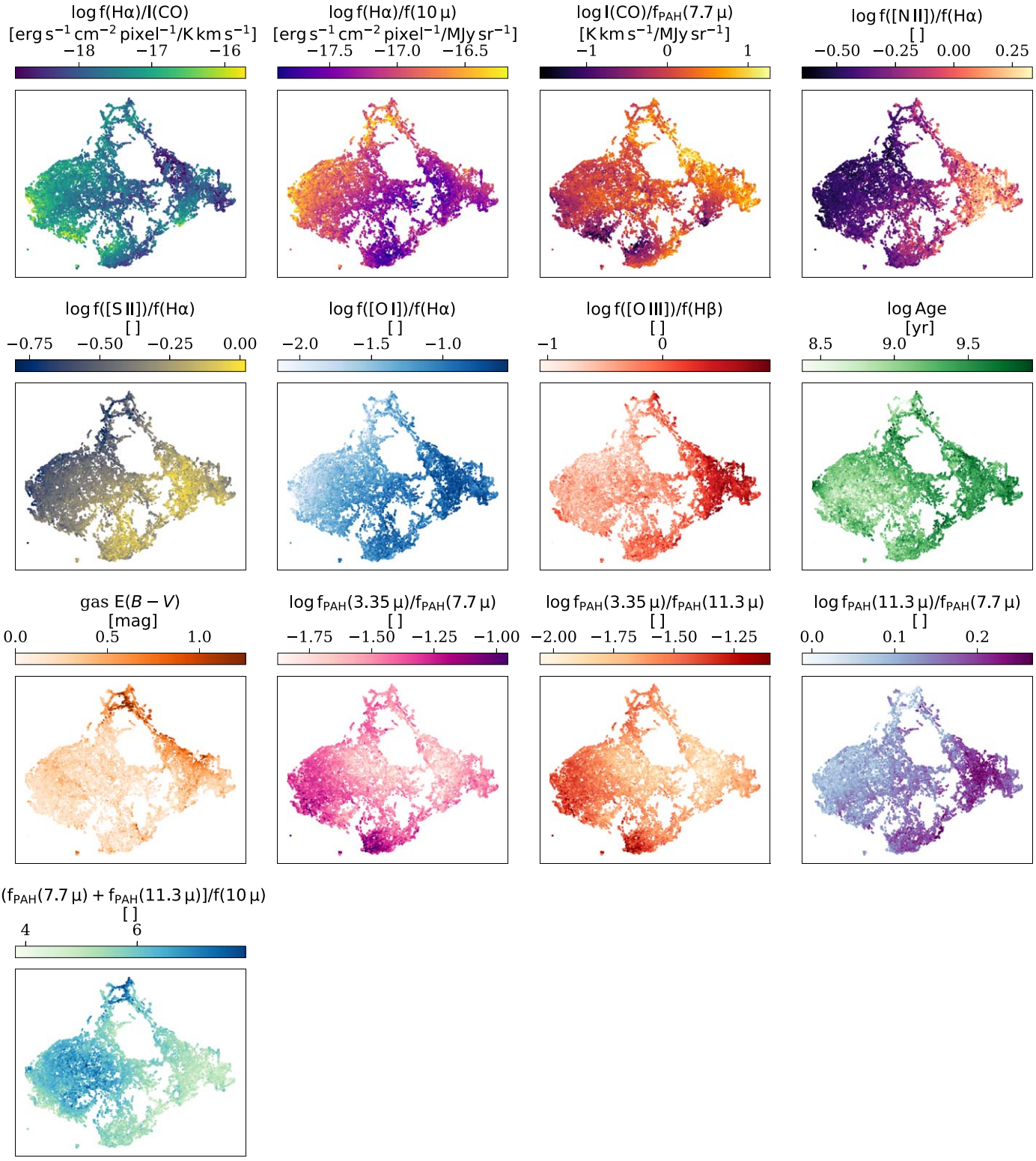

Figure 4. Our adopted two-dimensional UMAP embedding color-coded by the features in the input data set.

properties of the stellar population, multiphase gas, dust, and star formation, on a scale of 150 pc in these galaxies.

The top left panel of Figure 6 shows our adopted two-dimensional embedding by UMAP (Section 3.2), where every point represents a pixel from one of the three galaxies. The distribution of points in the two-dimensional space shows several regions with an overdensity of points, which may be interpreted as separate clusters, connected to each other through filamentary structures. This highly connected filamentary structure,[45] seen for different UMAP hyperparameter choices (see Appendix A),

[45] We observe the same highly connected filamentary structure in the two-dimensional embedding when using one pixel per spatial resolution element (150 pc) instead of two (Section 2). This suggests that this structure is not driven by our pixels subsampling the 150 pc resolution.

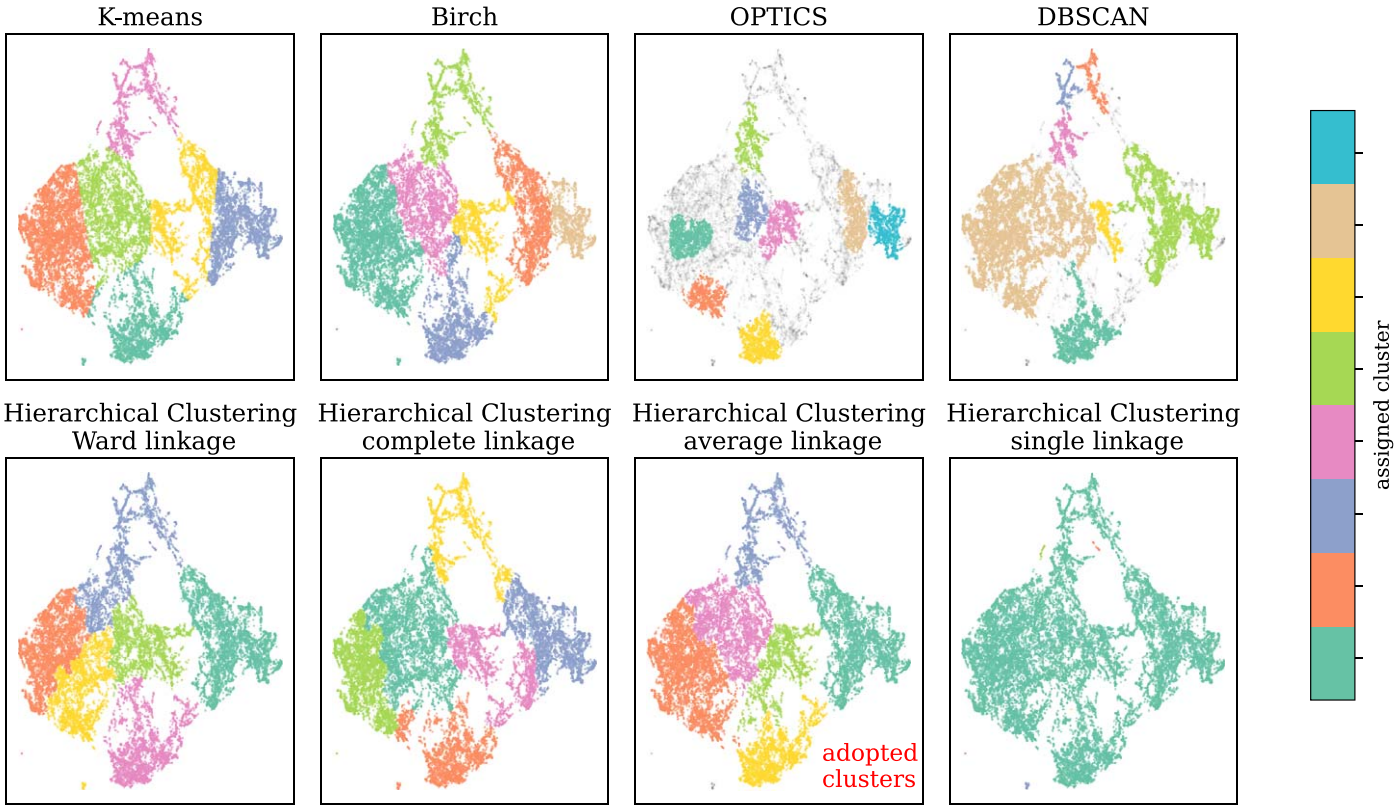

Figure 5. Different clustering algorithms applied to the two-dimensional embedding of the PHANGS pixels. Each panel shows the application of a different clustering algorithm to our adopted two-dimensional embedding by UMAP. In each panel, the points are color-coded according to their assigned cluster (all points for K-means, Birch, and hierarchical clustering), or are marked with light gray if they are not clustered (in the case of OPTICS and DBSCAN). The figure shows that, while some regions in the two-dimensional map are always considered as separate and well-defined clusters (e.g., the clusters at the bottommost and at the topmost), others are more ambiguous, with different clustering algorithms dividing the objects into different groups. We adopt the clusters identified using hierarchical clustering with the average linkage.

indicates that the multiwavelength features of the PHANGS pixels form continuous relations in the high-dimensional space they span. It suggests that the PHANGS pixels do not represent different entities with distinct physical properties (e.g., the difference between a star and a quasar), but rather the same entity with varying physical conditions (e.g., optical spectra of stars with different temperatures). This is not surprising given that the features we constructed trace different physical properties of gas and dust, averaged over a 150 pc scale.

The points in the two-dimensional space are color-coded according to their assigned group using our adopted clustering algorithm (Section 3.3). Since the points do not form well-separated clusters in the two-dimensional space, the resulting groups are somewhat arbitrary, with different clustering algorithms and different hyperparameter choices changing the resulting groups. Figure 5 and the top left panel of Figure 6 demonstrate the arbitrary nature of partitioning objects that form a continuous sequence into distinct groups—there is more than one way to divide the objects into groups, each resulting in groups that are distinct in different aspects. Despite this ambiguity, dividing objects into groups is a common practice in science and in astronomy in particular, with examples ranging from classifying the stellar sequence into distinct stellar types, (O, B, A, F, G, K, M), the classification of core collapse supernovae according to their light curves or spectra, classification of AGN into type I and II, and more. The practice of dividing objects, even when they form a sequence, into groups is useful as it allows one to compare the properties of objects in different groups, and by that, gain insight into the physics that drive the continuous variation in their properties. The difference of this work is the use of statistical tools to dissect the high-dimensional space, rather than using predefined physics-motivated properties, such as line ratios, metallicity, mass, luminosity, temperature, etc.

In this section, we describe the properties of the adopted groups, showing that they each have distinct gas ionization and PAH properties (Section 4.1). We then present newly identified galaxy-wide correlations between PAH band and optical line ratios and use the adopted groups to interpret them (Section 4.2).

4.1. Distinct Gas Ionization and PAH Properties in the Different Clusters

In this section, we interpret the six groups using different observables. We use the spatial distributions of pixels in different groups, as well as the feature values and their relation to other features. In particular, we use optical line diagnostic diagrams, PAH band ratios, and the relations between the Hα, CO, and 10 μm emission.

Figure 6 shows the spatial distribution of the adopted groups. Although the input data set did not include information regarding the galaxy a region belongs to, nor information regarding the relative location of a region within the galaxy, the groups map onto large-scale coherent structures within the galaxies.

In Figure 7, we show the distribution of the groups in standard optical line diagnostic diagrams (BPT diagrams; Baldwin et al. 1981; Veilleux & Osterbrock 1987; Kewley et al. 2001). These

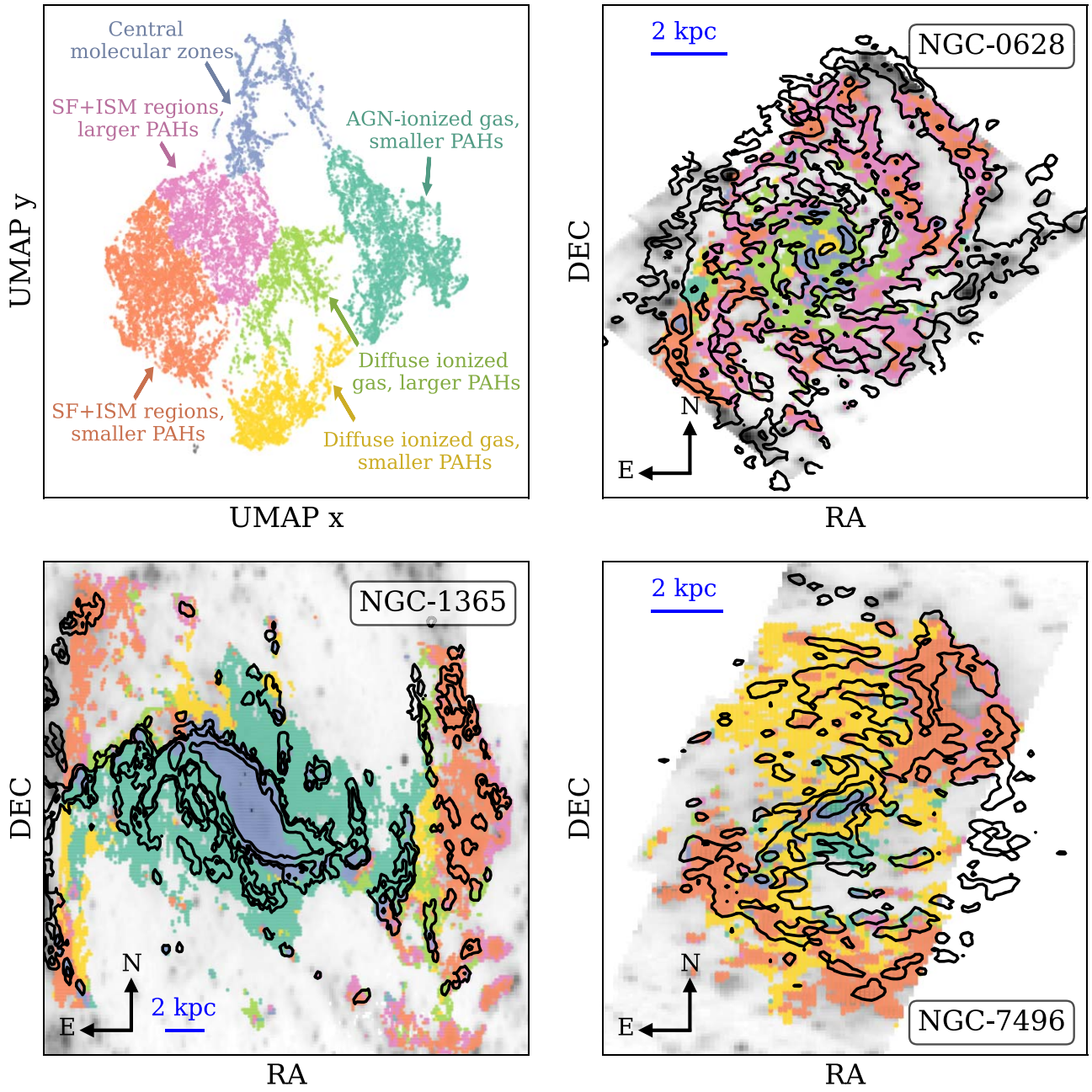

Figure 6. PHANGS multiwavelength pixels partitioned into six groups using dimensionality reduction and clustering algorithms. The top left panel shows the two-dimensional embedding by UMAP of the input data set, which includes 24,007 PHANGS pixels with 13 measured features that trace the stellar population, gas, dust, and star formation properties. Each point in the two-dimensional space represents a pixel from one of the galaxies we consider. The objects are divided into groups using the hierarchical clustering algorithm with the average linkage, and each point is color-coded according to its assigned group. The other three panels show the spatial distribution of the six groups, which map to large-scale coherent structures within the galaxies. In these panels, the gray-scale background represents the Hα surface brightness, and the black contours represent the CO emission.

diagrams are used to constrain the main source of ionizing radiation, by classifying a set of emission lines into one of the three classes: (i) H II regions, where the ratios are consistent with ionization by O and early B-type stars, (ii) Seyfert, where the ratios are consistent with ionization by AGN, and (iii) LINER/LIER, where the ratios may be consistent with either AGN ionization, photoionization by hot and evolved stars, or shock-excited gas (e.g., Kewley et al. 2001; Kauffmann et al. 2003; Kewley et al. 2006; Allen et al. 2008; Cid Fernandes et al. 2010; Rich et al. 2011, 2015; Belfiore et al. 2022).

In Figure 8, we show the distribution of the groups in the PAH 11.3/7.7 μm versus 3.3/11.3 μm plane (hereafter PAH 11.3/7.7 and 3.3/11.3 ratios). These band ratios are sensitive to the ionized fraction of PAHs, the PAH size distribution, and the shape of the incident FUV-optical radiation (e.g., Draine et al. 2021; Rigopoulou et al. 2021; and Section C.1).

Figures 7 and 8 show that the different groups have distinct ionized gas and PAH properties, as probed by the different optical line and PAH band ratios. Below, we describe these general properties[46]:

[46] The order in which we describe these clusters also corresponds to the order in which they are presented in all the figures from this point forward.

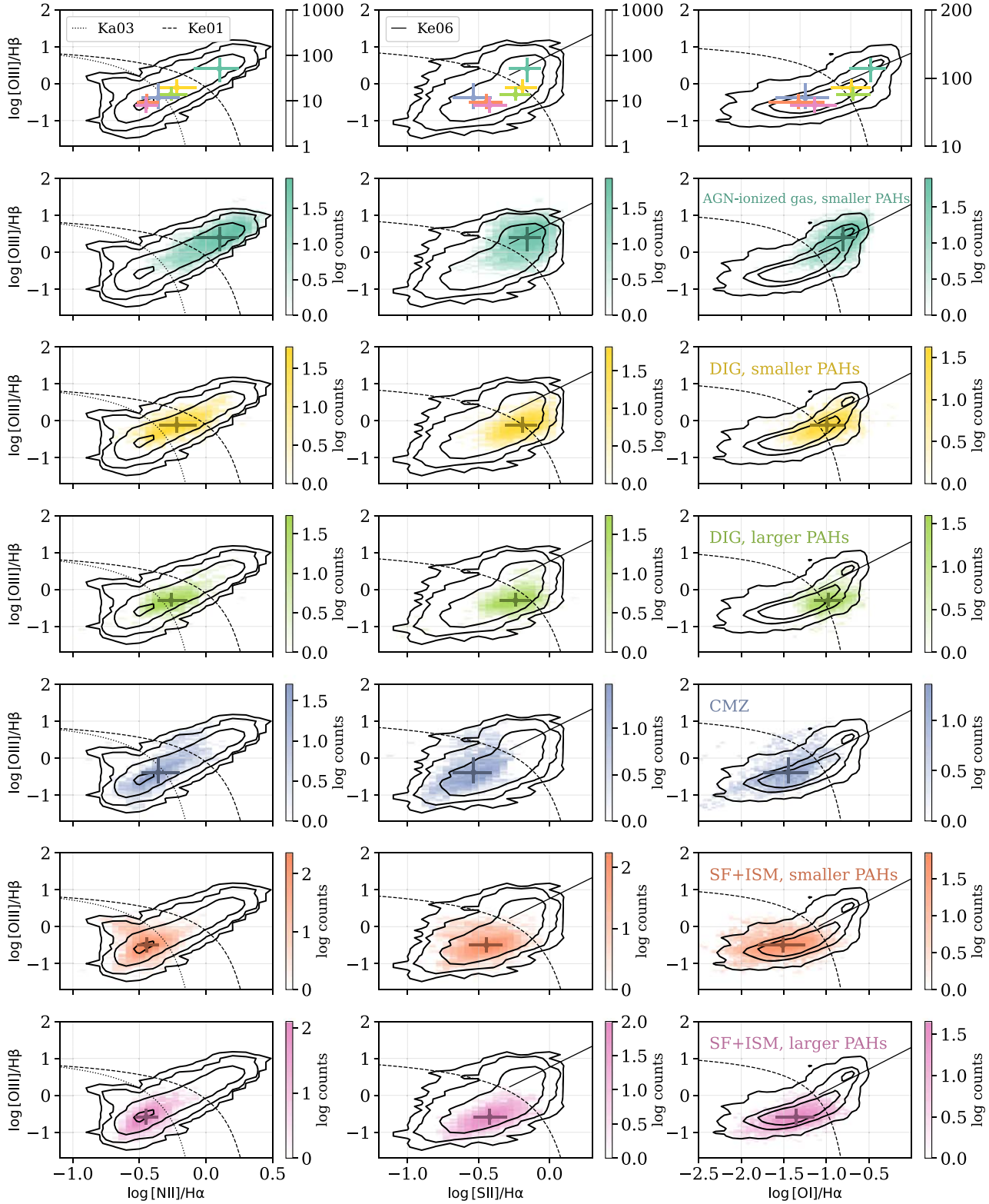

Figure 7. Location of the identified groups in optical line diagnostic diagrams. Each row shows the optical line ratios on standard diagnostic diagrams (e.g., Baldwin et al. 1981; Veilleux & Osterbrock 1987; Kewley et al. 2001): log([O III]/Hβ) vs. log([N II]/Hα), log([S II]/Hα), log([O I]/Hα). These diagrams are used to constrain the main source of ionizing radiation. In the left panel, we show the separating criteria by Kewley et al. (2001) and Kauffmann et al. (2003), which are used to separate ionization by young massive stars from AGN. In the middle and right panels, we show the LINER-Seyfert separating criteria by Kewley et al. (2006). The black contours represent the distribution of line ratios in all the pixels we considered in our analysis. The colormaps represent two-dimensional histograms of the line ratios for a given group, in logarithmic scale. The crosses represent the 16th, 50th, and 84th percentiles of the distributions in each of the band ratios. The different groups show different distributions in these diagnostic diagrams, suggesting different sources of ionizing radiation, as described in the text. Details on the classification of each group can be found in Section 4.1.

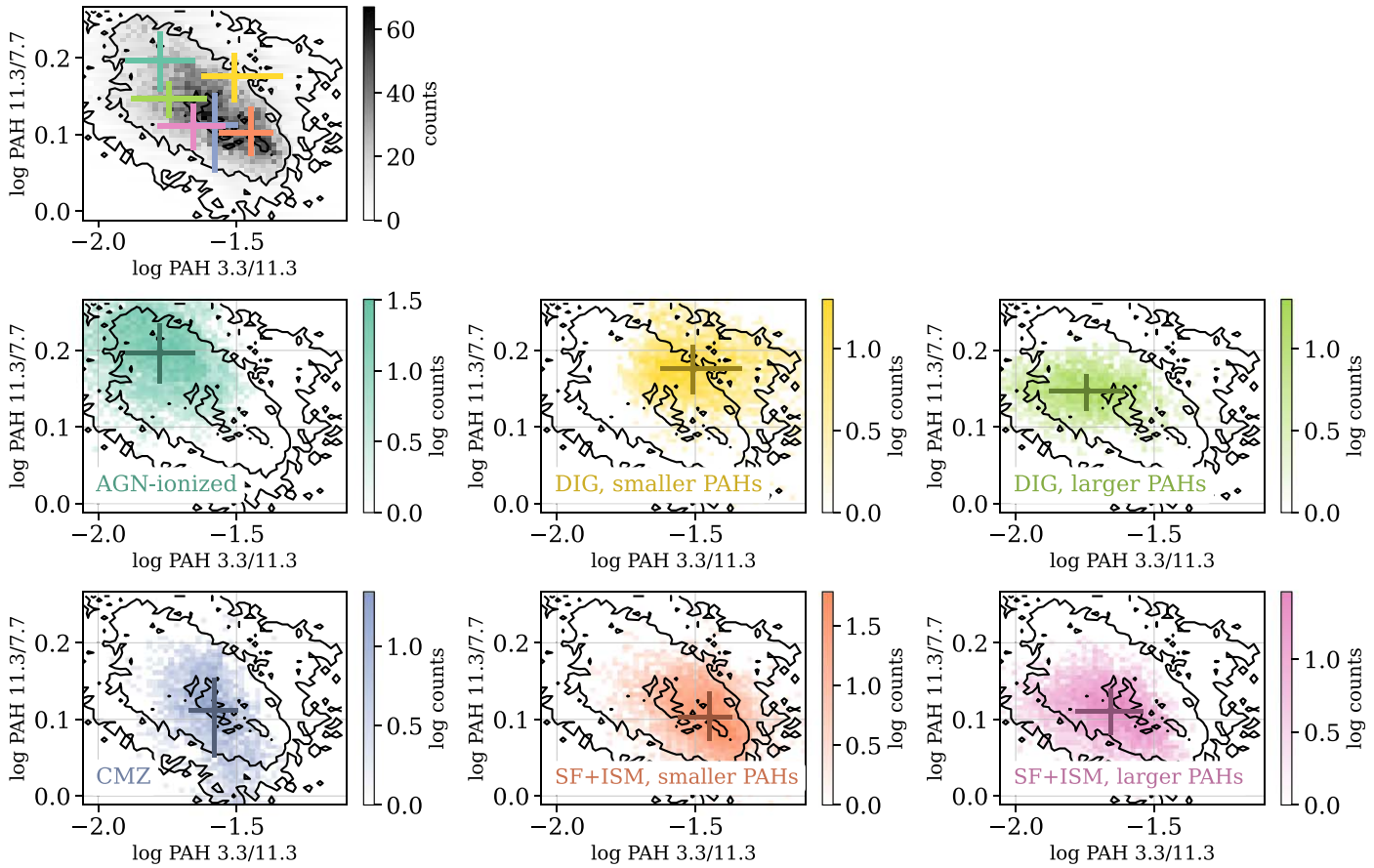

Figure 8. Distribution of the identified groups in the PAH band ratios plane. The panels show the distribution of pixels in the PAH 11.3/7.7 μm vs. 3.3/11.3 μm plane. The top left panel shows the distribution of all the pixels we consider using gray-scale color-coding, and the rest of the panels show the distributions in each individual group. While the gray-scale colormap on top represents the two-dimensional histogram counts in a linear scale, the individual group panels show the counts in a logarithmic scale. The black contours represent the distribution of all the pixels we consider, and they are the same in the different panels. The crosses represent the 16th, 50th, and 84th percentiles of the distributions in each of the band ratios for each group. Details on the classification of each group can be found in Section 4.1.

(1) AGN-photoionized gas (turquoise group). These pixels can primarily be found in NGC 1365 and NGC 7496,[47] both of which host AGN in their center (e.g., Morganti et al. 1999). The optical line ratios place most of the pixels in the Seyfert/LINER region in the BPT diagram, with log([O III]/Hβ) ratios too high to be powered solely by hot and evolved stars (e.g., Cid Fernandes et al. 2010; Byler et al. 2019; Belfiore et al. 2022). The pixels are located primarily along the AGN ionization cones, suggesting that the gas is photoionized by the AGN. Some of the pixels in this group are in the bars of NGC 1365, and these pixels show properties intermediate between the AGN-photoionized group and the central molecular zone (CMZ) group below. Interestingly, other clustering algorithms (e.g., BIRCH and OPTICS in Figure 5) divide this group into two—one that corresponds to pixels in the bars and the other that corresponds to pixels within the AGN ionization cones. From here forward, when referring to this group, we will focus on the pixels that are within the AGN ionization cones, located on kiloparsec scales, in the bulge that is dominated by an older stellar population. The pixels in this group show the highest PAH 11.3/7.7 band ratio, and the lowest PAH 3.3/11.3 band ratio. Importantly, while the AGN seems to be dominating the ionizing radiation, resulting in the Seyfert-like line ratios, the old stellar population in the bulge probably dominates the FUV-optical radiation, and is thus responsible for the PAH heating.

(2+3) Diffuse ionized gas (yellow and green groups). These pixels are seen in all three galaxies, and they primarily probe regions with faint Hα and CO emission (e.g., Figures B2 and B3 in Appendix B). The optical line ratios are consistent with LINER/LIER-like emission. In addition, these pixels spatially coincide with regions identified by Belfiore et al. (2022) as regions dominated by diffuse ionized gas, where the gas is ionized by a combination of radiation leaking from H II regions along with emission from hot and evolved stars, called the HOLMES mixing sequence. Both of the groups show a quite high 11.3/7.7 PAH band ratio. The two groups differ in their 3.3/11.3 PAH band ratio, with the yellow group showing significantly larger values than those of the green group.

(4) Central molecular zone (CMZ; slate blue group). Most of the pixels in this group belong to the central part of NGC 1365, a region known for hosting extreme star formation that is powered by a massive molecular gas reservoir (see Schinnerer et al. 2023; and an overview by Henshaw et al. 2023). Interestingly, some of the central pixels of NGC 0628 and NGC 7496 are also assigned to this group. The optical line ratios place the pixels in the H II region of the BPT diagram, suggesting ionization by young massive stars. This group

[47] Some of the pixels of this group belong to NGC 628. These pixels represent the very few pixels that are located in H II regions in the BPT diagram that were assigned to this cluster. In the two-dimensional embedding, they are located at the intersection with the CMZ cluster described below. Their properties are more in line with those of pixels in the CMZ cluster, and we believe that they should not belong to the AGN group.

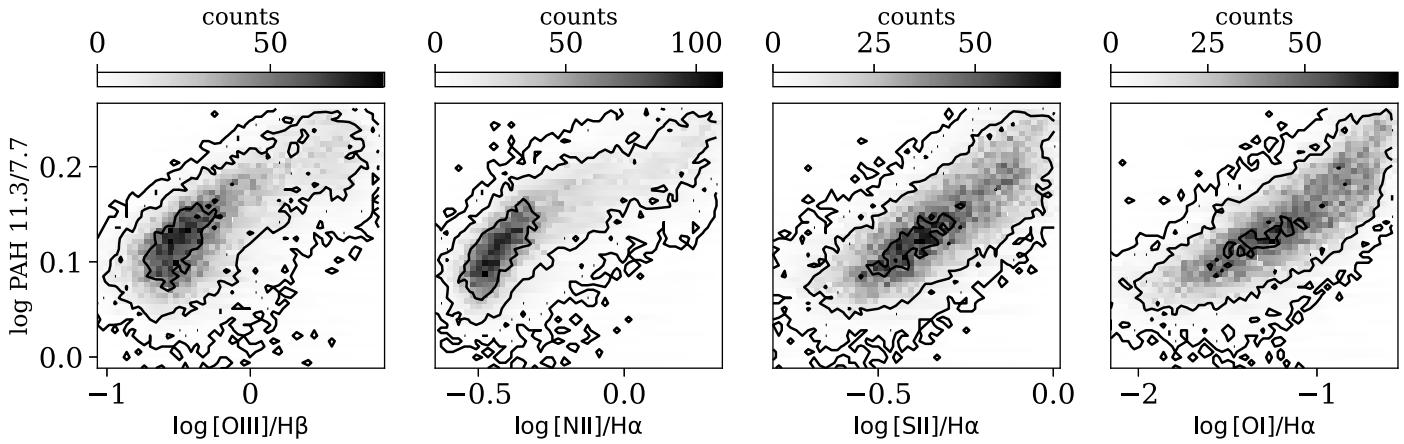

Figure 9. PAH band ratio log(11.3/7.7) vs. optical line ratios. The different panels show the distribution of the PAH band ratio log(11.3/7.7) vs. the optical line ratios log([O III]/Hβ), log([N II]/Hα), log([S II]/Hα), and log([O I]/Hα), for all the PHANGS pixels.

differs from the two star formation (SF)+ISM groups below primarily due to its very bright CO emission (Figure B3 in Appendix B), in particular with respect to the observed Hα, and significant dust extinction. It shows quite a low 11.3/7.7 PAH band ratio, consistent with those observed in the SF+ISM groups below, and an intermediate 3.3/11.3 ratio, in between those of the SF+ISM groups. Some of the pixels of this group deviate from the strong correlation between 3.3/7.7 and 3.3/11.3 (see Figure B4 in Appendix B), which may suggest the presence of silicate 9.7 μm absorption that reduces the observed flux in F1130W.

(5+6) Star-forming regions and ISM (orange and pink groups). These pixels are primarily seen in the spiral arms of the three galaxies. Their Hα, CO, and 10 μm emission show strong correlations, suggesting standard star formation that is powering the observed Hα and 10 μm emission. The optical line ratios are consistent with ionization by young massive stars. Similarly to the CMZ group, they show comparable and low 11.3/7.7 PAH band ratios. The first difference between the two groups is in their 3.3/11.3 PAH band ratio, which is significantly larger in the orange group than in the pink group. In addition, the orange group has a higher Hα/CO ratio than the pink group, which may suggest that it corresponds to more dispersed and evolved clouds compared to those of the pink group.

The three galaxies we consider dominate different regions of the two-dimensional embedding by UMAP (see Figure 3). This can be explained in the context of the groups we identify—(i) NGC 1365 is unique in having a large number of pixels that trace the CMZ, the bar, and AGN-photoionized gas on kiloparsec scales. It therefore dominates two out of the six groups. (ii) The diffuse ionized gas in NGC 7496 differs from that of NGC 0628 and NGC 1365 in PAH band ratios, thus dominating one of the groups. (iii) The star-forming and ISM regions of NGC 0628 differ from from those of NGC 1365 and NGC 7496 in PAH band ratios. It therefore dominates another group. At this point, it is unclear whether adding the rest of the 19 PHANGS galaxies would fill-in the space, making pixels that originate from a given galaxy indistinguishable from pixels of other galaxies. For that, it may be necessary to correct some of the features for inclination. We plan to address this question in a future publication where we plan to include all 19 PHANGS galaxies.

4.2. Close Connection between the Heating of PAHs and the Ionization of the Warm Ionized Gas

We identify significant and tight correlations between different PAH band and optical line ratios (see the full correlation matrix in Figure B1 in Appendix B). These correlations are seen across the entire data set, extending from the star-forming regions and the ISM, through the diffuse ionized gas, to the AGN-photoionized gas. The correlations are also detected in individual groups in which the dynamical range is large enough. Ionizing radiation is expected to destroy PAHs, and observations suggest much weaker PAH emission in regions dominated by ionized gas, such as H II regions (e.g., Chastenet et al. 2019, 2023a; Chown et al. 2023; Egorov et al. 2023; Lee et al. 2023; Peeters et al. 2023; and reviews Tielens 2008; Li 2020), although, at our spatial resolution, the PHANGS pixels include contributions from both. These correlations suggest a strong connection between the heating of PAHs and the ionization of the warm ionized gas on 150 pc scales.

In Figure 9, we show the PAH band ratio 11.3/7.7 versus the optical line ratios [O III]/Hβ, [N II]/Hα, [S II]/Hα, and [O I]/Hα, for the PHANGS pixels considered in our analysis. The 11.3/7.7 band ratio shows strong correlations with all of them. In Figure 10, we show the PAH band ratio 3.3/11.3 versus the optical line ratios. The correlations show a larger scatter than those of the 11.3/7.7 band ratio, but they extend over twice as large a dynamical range. In addition, there is a clear difference in the relation seen for the lower-ionization transitions traced by [S II]/Hα and [O I]/Hα compared to the higher-ionization transitions [N II]/Hα and [O III]/Hβ, with the former showing stronger correlations with the 3.3/11.3 band ratio.

Since the PAH band ratios are based on the broadband filter ratios F1130W/F770W and F335M$_{\rm PAH}$/F1130W, we first ensure that these correlations are not due to varying contributions of dust continuum emission to the F1130W filter flux. Since the F1000W filter likely has a large contribution from PAHs under most conditions in the PHANGS galaxies (e.g., Leroy et al. 2023), we use the F2100W filter flux to trace the dust continuum emission. If the observed correlations are due to a varying contribution of hot dust emission, we expect to find significant correlations between F2100W/F770W and F335M$_{\rm PAH}$/F2100W and the optical line ratios (F2100W replaces F1130W). We find no such correlations. Therefore, in what follows, we assume that the change in F1130W/F770W

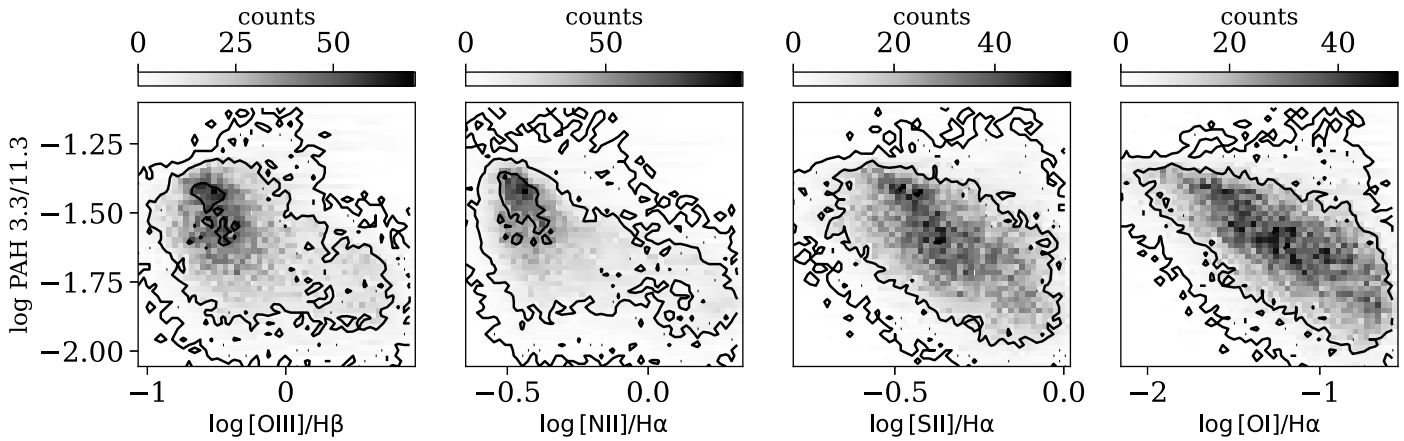

Figure 10. PAH band ratio log(3.3/11.3) vs. optical line ratios. The different panels show the distribution of the PAH band ratio log(3.3/11.3) vs. the optical line ratios log([O III]/Hβ), log([N II]/Hα), log([S II]/Hα), and log([O I]/Hα), for all the PHANGS pixels.

and F335M$_{\rm PAH}$/F1130W band ratios traces changes in PAH emission.

Various PAH band ratios, including those we consider here, have been used to place constraints on the PAH size and charge distribution in different environments (see reviews by Tielens 2008; Li 2020). Theoretical calculations and laboratory measurements suggest that neutral PAHs have significantly different infrared spectra than ionized PAHs, with the former showing strong 3.3 and 11.3 μm bands compared to the 7.7 μm band, and the latter showing stronger 7.7 μm band compared to the 3.3 and 11.3 μm bands (e.g., Allamandola et al. 1999; Tielens 2008). Therefore, the 11.3/7.7 PAH band ratio has been used extensively as a PAH ionization diagnostic, and to a lesser extent, as a PAH size diagnostic (see below; e.g., Kaneda et al. 2005; Smith et al. 2007; Galliano et al. 2008; Diamond-Stanic & Rieke 2010; Vega et al. 2010; Lai et al. 2022; Chastenet et al. 2023b; Dale et al. 2023). Since smaller PAHs have smaller heat capacities than larger PAHs, stochastic heating by single photon absorption raises their peak temperature to higher values (e.g., Draine & Li 2001; Draine 2011), leading to higher luminosity in shorter wavelength bands such as 3.3 μm compared to longer wavelength bands such as 11.3 μm. Therefore, PAH band ratios such as 3.3/11.3, and, to a lesser extent, the 11.3/7.7, have been used as PAH size diagnostics (e.g., Smith et al. 2007; Galliano et al. 2008; Chastenet et al. 2023b; Dale et al. 2023; Lai et al. 2023; Ujjwal et al. 2024).

Draine et al. (2021) and Rigopoulou et al. (2021) showed that various PAH band ratios are also sensitive to the shape of the incident radiation field. The same population of PAHs may show significantly different 11.3/7.7 and 3.3/11.3 band ratios when heated by radiation that is dominated by young versus older stars. A harder FUV-optical radiation field, typical of young and massive stars, is expected to lead to hotter PAHs, resulting in higher luminosity in the 3.3 μm feature, and, to a lesser extent, the 7.7 μm feature, compared to the 11.3 μm. Since the PHANGS 150 pc sized pixels trace a variety of radiation fields, this effect must be taken into account when interpreting PAH band ratios. Dale et al. (2023) demonstrated that, in the PAH band ratio plane available for the PHANGS galaxies (3.3/11.3 μm versus 3.3/7.7 μm; see their Figure 3), the impact of varying PAH size distribution and a varying radiation field are degenerate with each other. In this work, we use the observed optical line ratios to break this degeneracy.

The mid-infrared emission band ratios therefore depend on intrinsic PAH properties such as their size and charge distribution, as well as on extrinsic properties such as the shape of incident radiation field, which affects their temperature. A further complication is that intrinsic PAH properties are expected to be affected by the properties of the radiation field, with numerous studies exploring the impact of photoionization and shocks on the PAH size and charge distribution across different environments (e.g., Hony et al. 2001; O'Halloran et al. 2006; Gordon et al. 2008; Croiset et al. 2016; Peeters et al. 2017; Knight et al. 2021). Lacking a general radiative transfer framework that accounts self-consistently for the change in PAH properties and their heating due to the radiation field, here, we take the simplified approach of examining the impact of intrinsic versus extrinsic processes separately, and neglect the possible impact of the radiation field on the intrinsic PAH properties. This is similar to the approach recently taken by Donnelly et al. (2024) to interpret the PAH band ratio variations in the low-luminosity AGN host NGC 4138. However, it is important to note that this approach may oversimplify the complex picture of PAH heating and emission in the ISM, and indeed, in several cases throughout this section, we find that a more complex picture is required to explain all the observed PAH band variations.

4.2.1. A Varying Radiation Field as the Main Driver of the Correlations

To interpret the observed correlations, we use our adopted groups, as well as the PAH emission models by Draine et al. (2021) and a range of spectral energy distributions (SEDs) of the incident radiation field (see details in Appendix C). Draine et al. (2021) calculated the infrared emission spectra of PAHs for different illuminating radiation SEDs, radiation intensities, and PAH size and charge distributions. To remove the underlying continuum emission from starlight or other hot, small grains, they employed a clipping scheme that focuses on strong PAH emission features. Using this clipping scheme, they studied the sensitivity of PAH band ratios to the SED of the illuminating radiation and the PAH size and charge distribution.

We use the infrared spectra predicted by Draine et al. (2021). In particular, we consider models calculated using the small, standard, and large PAH size distributions, and with low-, standard-, and high-ionization distributions. For the illuminating

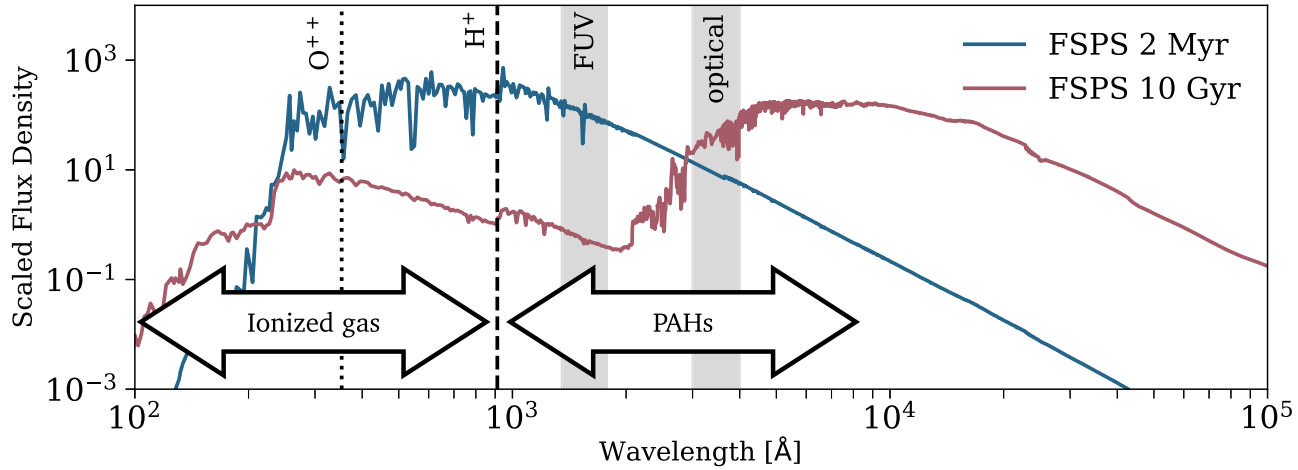

Figure 11. PAHs and ionized gas are sensitive to different parts of the illuminating radiation field. The lines represent two example SEDs calculated using the flexible stellar population synthesis code by Conroy et al. (2009). They correspond to single stellar populations of age 2 Myr and 10 Gyr (see Appendix C.2 for additional details), where the flux density of the 10 Gyr star has been scaled to higher values for representational purposes. The black vertical lines represent the wavelengths that correspond to photon energies that can ionize hydrogen (H^+; dashed) and oxygen twice (O^{++}; dotted). The gray bands correspond to the wavelength ranges 1350–1780 Å and 3000–4000 Å, which are our adopted definitions for FUV and optical luminosities. The conditions in the warm ionized gas primarily depend on the flux densities at $\lambda \leqslant 912$ Å. The PAHs are exposed to nonionizing radiation, and their heating depends on the shape of the FUV-optical radiation field, which we parameterize using νL_ν(FUV)/νL_ν(optical) (see Appendix C.1 for details).

radiation SED, we consider 12 different templates, covering a range of stellar population ages and FUV luminosities. The PAHs are believed to be located in regions that are shielded from ionizing radiation (e.g., Chastenet et al. 2019, 2023a; Egorov et al. 2023; Lee et al. 2023; and reviews Tielens 2008; Li 2020), and their heating is dominated by non-ionizing FUV-optical radiation (see Figure 11). To parameterize the dependence of PAH band ratios on the incident radiation field, we define the FUV-to-optical luminosity ratio to be νL_ν(1350–1780 Å)/νL_ν(3000–4000 Å), as illustrated in Figure 11.

We use the clipping scheme by Draine et al. (2021) to estimate the PAH band ratios 11.3/7.7 and 3.3/11.3, and parameterize in Appendix C.1 how they vary for different FUV-to-optical ratios, PAH size distributions, and PAH charge distributions. Since the JWST photometry includes contributions from non-PAH emission sources, and since the photometric bands do not coincide with the clipping points defined by Draine et al. (2021), we do not expect the absolute PAH band ratios to match those predicted by the models. Instead, we use the clipping scheme only to compare the observed *trends in ratios* with those predicted by the models. Dale et al. (2023) integrated the Draine et al. (2021) models under the JWST filter bandpasses, allowing them to compare the absolute values of the PAH band ratios to those predicted by the model. The observed PAH band ratios in the regions they consider (for the same three PHANGS galaxies) are within the ranges spanned by the models. In particular, they find that the PAH band ratios are more aligned with models of larger and ionized PAHs in compact stellar clusters and stellar associations.

In Figure 12, we show the PAH band ratios 11.3/7.7 and 3.3/11.3 versus [O III]/Hβ ratio, using our adopted groups. We explore the impact of both intrinsic (PAH size and ionization) and extrinsic (incident radiation field) property variation on the PAH band ratios as described below. For the intrinsic PAH properties, the expected changes in PAH band ratios due to varying PAH size or charge distribution are marked with blue and red arrows. Under our simplified assumption that the PAH size and charge distribution do not depend on the radiation field, these changes affect only the PAH bands and have no effect on the optical line ratios, and thus, they form arrows in the vertical direction only. The relative sizes of the arrows represent the expected change in PAH band ratios when changing the size/ionization from low to standard, or from standard to large/high. As for extrinsic processes, we mark with a gray arrow the expected mutual variation in the PAH band and optical line ratios under the assumption that the PAHs and ionized gas are exposed to different parts of the same spatially varying radiation, as described below. For the gray arrow, we assume that the PAH intrinsic properties are fixed.

Variation in intrinsic PAH properties (size and charge) as the main driver of the relation. For a fixed incident radiation field, Figure 12 suggests that PAHs are more ionized in regions with low optical line ratios. In particular, using our groups, PAHs are more ionized in star-forming regions and ISM, and are more neutral in the diffuse ionized gas and in the AGN-photoionized gas. The right panel of the figure shows a negative correlation between 3.3/11.3 and the optical line ratio, suggesting PAHs are smaller in regions with low optical line ratios. This would suggest smaller PAHs in the star-forming and ISM regions, and larger PAHs in the diffuse ionized gas. These are in line with the conclusions of Ujjwal et al. (2024) for the three PHANGS galaxies we consider. However, the 3.3/11.3 shows a strong positive correlation with the Hα/CO ratio (see section 4.2.2 for details). According to this interpretation, it would suggest that PAHs are smaller in regions with a larger fraction of ionized-to-molecular gas. This is the opposite of what would be expected if smaller PAHs are destroyed by ionizing radiation more efficiently than larger PAHs.

Variation in extrinsic properties (incident radiation) as the main driver of the relation ("common varying radiation field" interpretation hereafter). The left panel of Figure 12 shows that the tight correlation between the PAH band ratio 11.3/7.7 and the optical line ratios is in fact a sequence in ionized gas conditions, with the different groups occupying distinct ranges within the correlation. It may suggest that the PAHs and the ionized gas are exposed to *different parts* of the same radiation field. That is, while the ionized gas is exposed to the entire

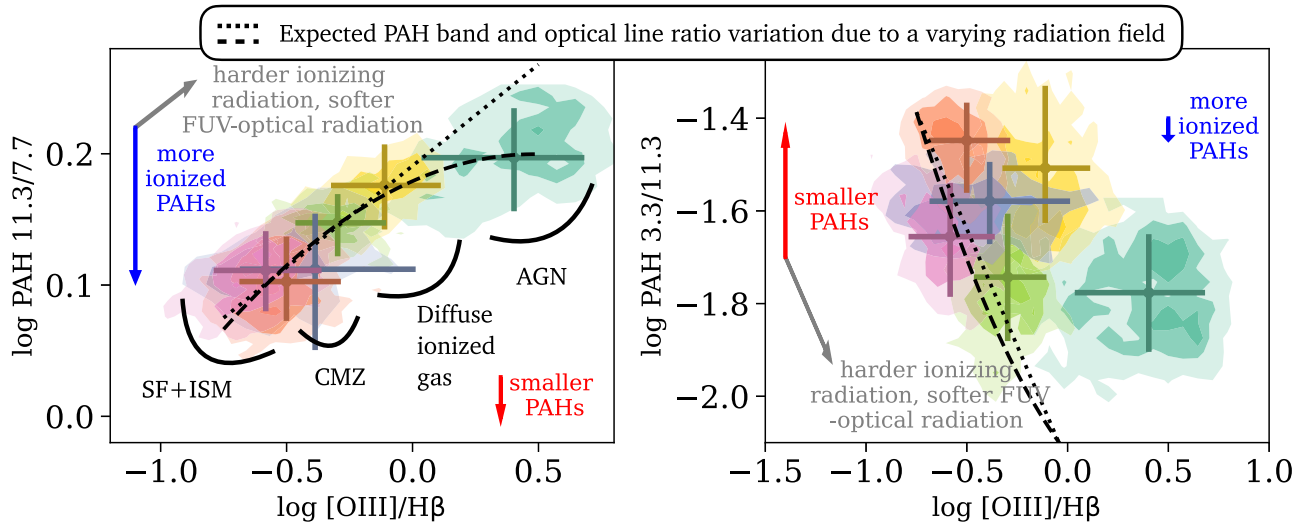

Figure 12. Interpreting the relation between PAH band ratios and log([O III]/Hβ). The panels show the PAH band ratios log(11.3/7.7) and log(3.3/11.3) vs. log([O III]/Hβ), where each group is plotted separately. The contours represent the two-dimensional distribution for each group, and the crosses represent the 16th, 50th, and 84th percentiles. The blue and red arrows represent the expected change in PAH band ratios when changing the PAH size and charge distribution while keeping the radiation field fixed (see Appendix C.1). The gray arrows represent the expected change in PAH band and log([O III]/Hβ) ratios assuming that the PAHs and the ionized gas are exposed to different parts of the same varying radiation field, while keeping the PAH size and charge distribution fixed. The varying radiation field is characterized by harder ionizing radiation and softer FUV-optical slope, a scenario that is expected under a wide range of assumptions regarding what powers it. The relation on the left panel is consistent with the "common varying radiation field" interpretation, without the need to invoke PAHs with different charge distributions. The dashed and dotted black lines in the left panel represent two polynomial fits to the 11.3/7.7 vs. [O III]/Hβ relation. These best fits are then propagated to the expected relations between 3.3/11.3 and [O III]/Hβ (right panel), assuming the "common varying radiation field" interpretation and the PAH models by Draine et al. (2021; see Appendix C.1 for details). Clusters that deviate from these lines in the right panel may represent PAHs with different size distributions. The dotted and dashed lines are given by the following equations: Left panel (y = log PAH 11.3/7.7, x = log [O III]/Hβ): $y = 0.15x + 0.19$, $y = -0.087x^2 + 0.084x + 0.18$. Right panel ($y$ = log PAH 3.3/11.3, x = log [O III]/Hβ): $y = -x - 2.14$, $y = 0.57x^2 - 0.55x - 2.12$.

radiation field, and its properties are set by the *ionizing part* of the radiation, the PAHs are located in regions shielded from the ionizing radiation, and are exposed only to the non-ionizing FUV-optical part of the radiation (see Figure 11). Under this interpretation, the radiation field varies over kiloparsec scales, and its variation drives the changes in both PAH band and optical line ratios. Our simplifying assumption is that the radiation field affects the PAH temperature distribution, but their size and charge distributions are fixed. The PHANGS pixels show significant variations in their stellar and/or AGN properties, with some regions dominated by young stellar populations, while others are dominated by old stars or AGN radiation. Therefore, we argue that this scenario is *unavoidable*. The only question is its impact on the PAH band ratios compared to those of varying PAH size and charge distributions.

The gray arrows in Figure 12 represent the expected relations between the PAH band and optical line ratios under the "common varying radiation field" interpretation. The expected relation depends on several factors: (i) the dependence of PAH bands on the FUV-to-optical luminosity ratio, (ii) the SED shape, in particular, the connection between the FUV-optical and ionizing parts of the radiation field, and (iii) the dependence of optical line ratios on the shape of the ionizing radiation.

We explore the dependence of PAH bands on the FUV-to-optical luminosity ratio using the Draine et al. (2021) models in Appendix C.1. In particular, we find that varying the radiation field from old (1–10 Gyr) to young (2 Myr) stellar populations increases the FUV-to-optical luminosity ratio by 2.5 dex, and at the same time, changes the PAH band ratios 11.3/7.7 and 3.3/11.3 by −0.1 and 0.55 dex respectively. Interestingly, these are quite similar to the ranges spanned by 11.3/7.7 and 3.3/11.3 in the PHANGS pixels.

We explore the connection between the FUV-optical and ionizing parts of the radiation field in Appendix C.2. We use several different stellar libraries, covering single stellar populations with ages of 2 Myr to 10 Gyr, while varying the metallicity, stellar isochrones, and stellar templates. Since Figure 12 suggests a sequence that covers H II regions, the diffuse ionized gas, and AGN-photoionized gas, we also consider models that are constructed to reproduce the optical line ratios in the diffuse ionized gas and AGN-photoionized gas. In particular, we consider the HOLMES mixing sequence introduced by Belfiore et al. (2022) to explain the observed optical line ratios in the diffuse ionized gas in PHANGS galaxies. In this model, the SED is a combination of radiation from a young stellar population leaking from H II regions and from hot and evolved stars, where the latter is used as it contributes the hard ionizing radiation required to power LINER-like emission line ratios. We also construct AGN+SF mixing sequences. We use a standard Shakura & Sunyaev (1973) accretion disk SED with several improvements (general relativistic corrections and radiative transfer in the disk atmosphere; e.g., Slone & Netzer 2012), which results in Seyfert-like line ratios in standard line diagnostic diagrams (see Appendix A in Baron & Netzer 2019). Similarly to the HOLMES mixing sequence, the AGN SED is mixed with a young (2 Myr) stellar template by varying the relative contribution of each.

Using these different SEDs, in Appendix C.2, we confirm that in all of these cases, i.e., stellar age sequence from young to old; HOLMES mixing sequence with varying contribution of old versus young stars; and AGN+SF mixing sequence with varying contributions from the stars versus the AGN, the

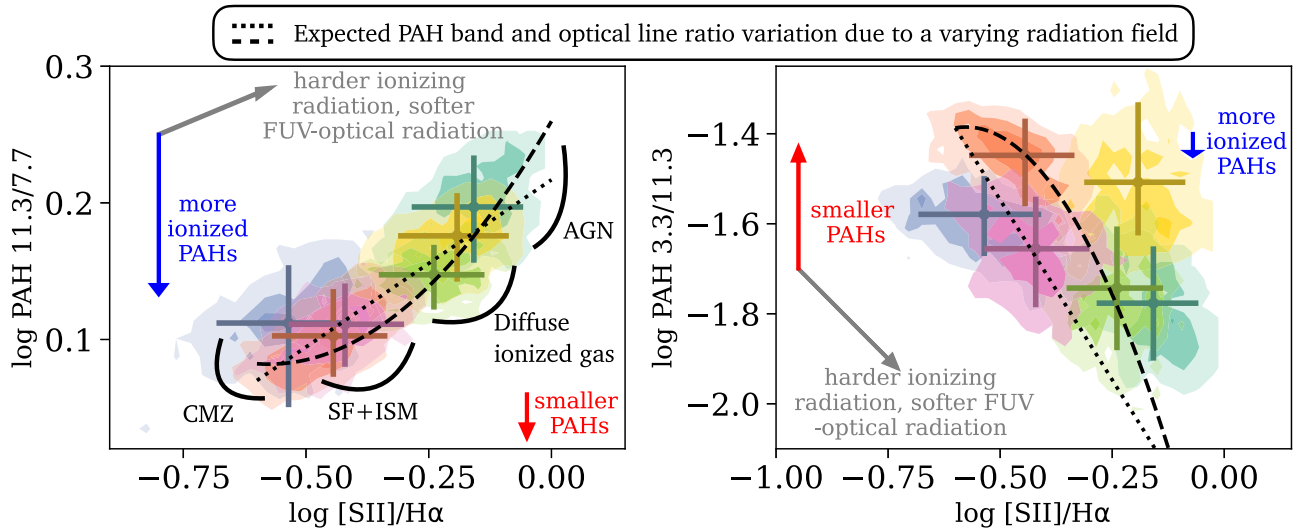

Figure 13. Interpreting the relation between PAH band ratios and log([S II]/Hα). Similar to Figure 12, but using the lower-ionization log([S II]/Hα) line ratio. The dotted and dashed lines are given by the following equations: Left panel (y = log PAH 11.3/7.7, x = log [S II]/Hα): $y = 0.22x + 0.21$, $y = 0.53x^2 + 0.61x + 0.26$. Right panel ($y$ = log PAH 3.3/11.3, x = log [S II]/Hα): $y = -1.6x - 2.34$, $y = -3.5x^2 - 4.02x - 2.54$.

general observed trend is a harder ionizing slope for a softer FUV-optical slope, matching the direction required to explain the relation in the left panel of Figure 12. Moreover, we find that a decrease of 2.5 dex in νL_ν(FUV)/νL_ν(optical) results in an increase of ~1 dex in $\nu L_\nu(O^{++})/\nu L_\nu(H^+)$, depending on the mixing sequence used. Therefore, the expected slope as indicated by the gray arrow is roughly consistent with the observed slope in the left panel of Figure 12.

The exact slope expected under the "common varying radiation field" interpretation also depends on the relation between the ionizing radiation and optical line emissivities. In particular, the optical line ratios we consider are measured with collisionally excited transitions whose line emissivity depends on the gas temperature (e.g., Osterbrock & Ferland 2006). Harder ionizing radiation leads to higher electron temperature (e.g., Garnett 1992; Byler et al. 2017). Both of these are expected to increase the expected optical line ratios, where the line emissivity depends linearly on ionic abundances (O^{++}, S^+, etc.) and exponentially on the electron temperature (e.g., Osterbrock & Ferland 2006). Indeed, we find equally strong correlations between 11.3/7.7 and the low-ionization transitions [N II]/Hα, [S II]/Hα, and [O I]/Hα (Figures 9 and 13), which can naturally be explained in a case where a change in temperature drives the change in the [O III], [N II], [S II], and [O I] line emissivities. These optical line ratios depend on the $\nu L_\nu(O^{++})/\nu L_\nu(H^+)$ ratio through the relation between the hardness of the ionizing radiation and the electron temperature. We intend to use detailed photoionization models to further investigate this scenario in a future work.

To summarize. The observed correlations between the PAH band 11.3/7.7 ratio and the optical line ratios [O III]/Hβ, [N II]/Hα, [S II]/Hα, and [O I]/Hα can be naturally explained in a scenario where the PAHs and ionized gas are exposed to different parts of the same spatially varying radiation field, without the need to invoke PAHs with different charge distributions. Since the PHANGS pixels trace regions with widely varying radiation SEDs that are a combination of young stars, old stars, and/or AGN, a variation of PAH band ratios due to the changing radiation field SED is *unavoidable*. We use the PAH models by Draine et al. (2021) and a wide range of assumptions about the variation of the incident radiation field to show that the expected slope of the relation is roughly consistent with that observed.

The main implications of this scenario. (I) The very small scatter (~0.03 dex) in the relation between 11.3/7.7 and [O III]/Hβ (or any other optical line ratio) implies that the ionized PAH fraction is quite uniform on scales of 150 pc across different environments in local galaxies. Since the fraction of ionized PAHs is set by the balance between ionization (by photons or collisions) and recombination, this uniformity suggests a strong self-regulation of the ISM that limits variations in gas temperature and electron density. In particular, the property $U\sqrt{T}/n_e$, where U is dimensionless intensity parameter, could, in principle, vary by a factor of a few (see Tielens 2005; Galliano et al. 2008; Draine et al. 2021). According to our interpretation, this property can vary by no more than ~100% across different environments (SF+ISM and the diffuse ionized gas). In a future study, we plan to use all 19 PHANGS galaxies to place constrains on $U\sqrt{T}/n_e$ in different environments. (II) The 11.3/7.7 PAH band ratio may potentially be used to trace the shape of the FUV-optical parts of the radiation field across nearby galaxies. The combination of the 11.3/7.7 and optical line ratios may therefore be used to constrain simultaneously the ionizing and nonionizing parts of the radiation field. (III) The varying radiation field is expected to impact PAH band ratios that are typically used as PAH size indicators (like the 3.3/11.3 band ratio). To constrain the PAH size distribution, the variation of the radiation SED must be accounted for.

Finally, it is worth noting that previous Spitzer-based studies found surprisingly high 11.3/7.7 PAH band ratios in AGN-dominated systems and in elliptical galaxies (e.g., Kaneda et al. 2005; Smith et al. 2007; Galliano et al. 2008; Kaneda et al. 2008; Diamond-Stanic & Rieke 2010; Vega et al. 2010), which, in some cases, were found outside the range of ratios predicted by models. There are two main differences between the variation of the 11.3/7.7 PAH band ratio found here and those reported by the studies mentioned above. First, while the Spitzer-based studies found a 11.3/7.7 PAH band ratio that may be larger by ~1 order of magnitude compared to the ratio

observed in normal environments, the variation we observe is much smaller, of the order of 0.1 dex. While a change of 1 dex cannot be explained solely by a change of the hardness of radiation field, a change of 0.1 dex can. Second, our work focuses on 150 pc sized regions, where the wealth of multiwavelength observations allows us to constrain separately the dominant source of ionizing radiation and the dominant source of FUV-optical photons. In our case, although we classify the first group as "AGN-photoionized gas," these pixels are very different from the AGN-dominated systems presented and discussed by, e.g., Smith et al. (2007), Diamond-Stanic & Rieke (2010), and Lai et al. (2022), since, in our case, old stars probably dominate the PAH heating.

4.2.2. Constraining the PAH Size Distribution

The right panels of Figures 12 and 13 show the PAH band ratio 3.3/11.3 versus the optical line ratios [O III]/Hβ and [S II]/Hα. Although the 3.3/11.3 had been used as a PAH size indicator (e.g., Lai et al. 2023; Ujjwal et al. 2024), it is also sensitive to the shape of the incident radiation. In particular, varying the stellar age from 2 Myr to 1 Gyr can change the ratio by 0.55 dex, similar to the entire range spanned by 3.3/11.3 in Figures 12 and 13. This has been pointed out by Chastenet et al. (2023b) when interpreting the PAH band ratios observed in the first three PHANGS-JWST galaxies, where, similarly to Dale et al. (2023), they argued that the hardness of the radiation field and PAH size may both affect the observed 3.3/11.3 ratio. In this section, we use the observed correlations between the PAH band and optical line ratios to break this degeneracy and disentangle the impact of the radiation field from that of the PAH size distribution on the observed 3.3/11.3 PAH band ratio.

Under the "common varying radiation field" interpretation, both 11.3/7.7 and 3.3/11.3 are expected to change with the optical line ratios, with the two slopes (11.3/7.7 versus optical line ratio), and (3.3/11.3 versus optical line ratio) connected to each other through the shape of the radiation field. In Section 4.2.1 above, we suggest that the relation between the 11.3/7.7 and the optical line ratios can be explained entirely using a varying radiation field, while keeping the PAH charge distribution fixed. We can use the Draine et al. (2021) models to propagate this relation to the expected relation between 3.3/11.3 and the optical line ratios.

In the left panels of Figures 12 and 13, we show two polynomial fits to the observed relation between 11.3/7.7 and the optical lines. Since the 11.3/7.7 is a function of the FUV-to-optical luminosity ratio of the radiation field, these best-fitting polynomials connect the FUV-to-optical luminosity ratio with the observed [O III]/Hβ or [S II]/Hα line ratios. Using the relation between 3.3/11.3 and the FUV-to-optical luminosity ratio, we can therefore calculate the expected relation between 3.3/11.3 and [O III]/Hβ or [S II]/Hα, under the "common varying radiation field" interpretation. We describe this process in Appendix C.1 and show these lines in the right panel of Figures 12 and 13.

The right panels of Figures 12 and 13 suggest that a significant part of the observed 3.3/11.3 variation is in fact due to a varying radiation field, rather than a varying PAH size distribution. Significant deviations from the expected slopes may be attributed to changes in the PAH size distribution. For example, among the two SF+ISM groups, the orange group is consistent with hosting smaller PAHs. Among the two diffuse ionized gas groups, the yellow group corresponds to regions with smaller PAHs. Finally, the figures suggest *smaller* PAHs in the AGN-photoionized gas group, since it is located *above* the propagated relation. Contrary to the left panels of Figures 12 and 13, the large deviations of individual groups from the propagated fitted lines suggest that a varying radiation field that affects the PAH temperature distribution, but not the PAH size distribution, cannot fully account for the observed band ratio variation. It may be possible that a more general model that includes the impact of the varying radiation field on the PAH size distribution can fully explain the observed trends. Constructing such a model is beyond the scope of this paper.

To further illustrate the impact of the different PAH heating interpretations on the derived PAH sizes, in Figure 14, we show the 3.3/11.3 PAH band ratio versus the Hα/CO ratio. The left panel shows the measured 3.3/11.3 ratio, and assuming that the radiation field does not vary, the ratio can be used directly to constrain PAH sizes. According to this interpretation, the 3.3/11.3 increases with the Hα/CO ratio, suggesting that regions with a larger fraction of ionized gas tend to host smaller PAHs. The positive trend in the left panel is surprising, given that smaller PAHs are believed to be more easily destroyed by ionizing radiation compared to larger PAHs (reviews by Tielens 2008; Li 2020). The right panel of Figure 14 shows Δ3.3/11.3, which is the expected 3.3/11.3 after accounting for the variation due to the varying illuminating SED. We calculate Δ3.3/11.3 by subtracting the black dotted line (linear fit) in the right panel of Figure 13 from the measured 3.3/11.3 band ratio.[48] According to this "common varying radiation field" interpretation, PAHs are generally larger with increasing Hα/CO, consistent with the idea that larger PAHs survive in regions with more ionizing radiation.

The right panel of Figure 14 shows a significant scatter that is dominated by the 3.3/11.3 ratios of the yellow and orange groups, both show significantly larger 3.3/11.3 ratios than those of their equivalent groups (the green and pink; diffuse ionized gas and SF+ISM respectively). Even if the "common varying radiation field" interpretation describes more accurately the observed PAH band ratios, the scatter in the diagram suggests that a simple picture of "larger PAHs in regions more dominated by ionizing radiation" is too simple to describe the observed variations in the 3.3/11.3 band in the PHANGS galaxies, and that additional factors may be at play. We plan to revisit this question in a future work, where we plan to apply this analysis to the full PHANGS+JWST sample of 19 galaxies.

4.2.3. PAH Bands versus Optical Line Ratios in Individual Groups

We find significant correlations between the PAH band and optical line ratios in a few groups where the dynamical range is large enough. In Figure 15, we show these relations with the [S II]/Hα ratio for the three groups: AGN-photoionized gas, and the two SF+ISM groups. The top row shows the 11.3/7.7 PAH band ratio. The observed slopes of the 11.3/7.7 versus [S II]/Hα relations observed for the different groups are comparable to each other (0.16, 0.13, and 0.15 for AGN, SF +ISM 1, and SF+ISM 2, respectively) and to the best-fitting slope found for all PHANGS pixels in Figure 13 (0.22).

[48] We reach the same conclusion if we use the relation with the [O III]/Hβ (Figure 12) instead.

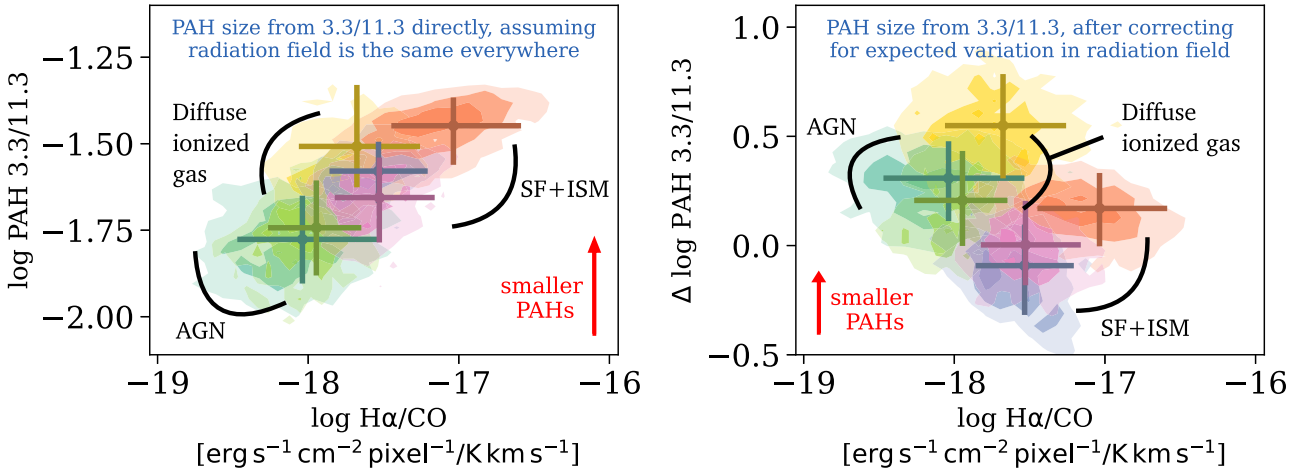

Figure 14. Different PAH heating scenarios lead to opposite interpretations regarding PAH sizes. The two panels show the variation of the 3.3/11.3 PAH band ratio vs. the Hα/CO ratio. The left panel shows the measured 3.3/11.3 PAH band ratio. Assuming the "nonvarying radiation field" interpretation, the 3.3/11.3 can be used directly to constrain the PAH size distribution. The left panel therefore suggests that regions with higher fractions of ionized gas (traced by larger Hα/CO ratios) are associated with smaller PAHs. This seems to be in tension with the common picture that smaller PAHs are more efficiently destroyed by ionizing radiation than larger PAHs. The right panel shows Δ3.3/11.3, which is the measured 3.3/11.3 PAH band ratio after correcting for the expected variation due to changing SED (subtracting the black dotted line in Figure 13 from the measured 3.3/11.3 values). Contrary to the left panel, the right panel suggests that regions with higher fractions of ionized gas host larger PAHs, in line with the common picture that larger PAHs survive in harsher environments. The yellow and orange clusters, which predominantly come from NGC 7496, are above this relation and show significantly smaller PAH size distributions (see Section 4.2.2 for additional details).

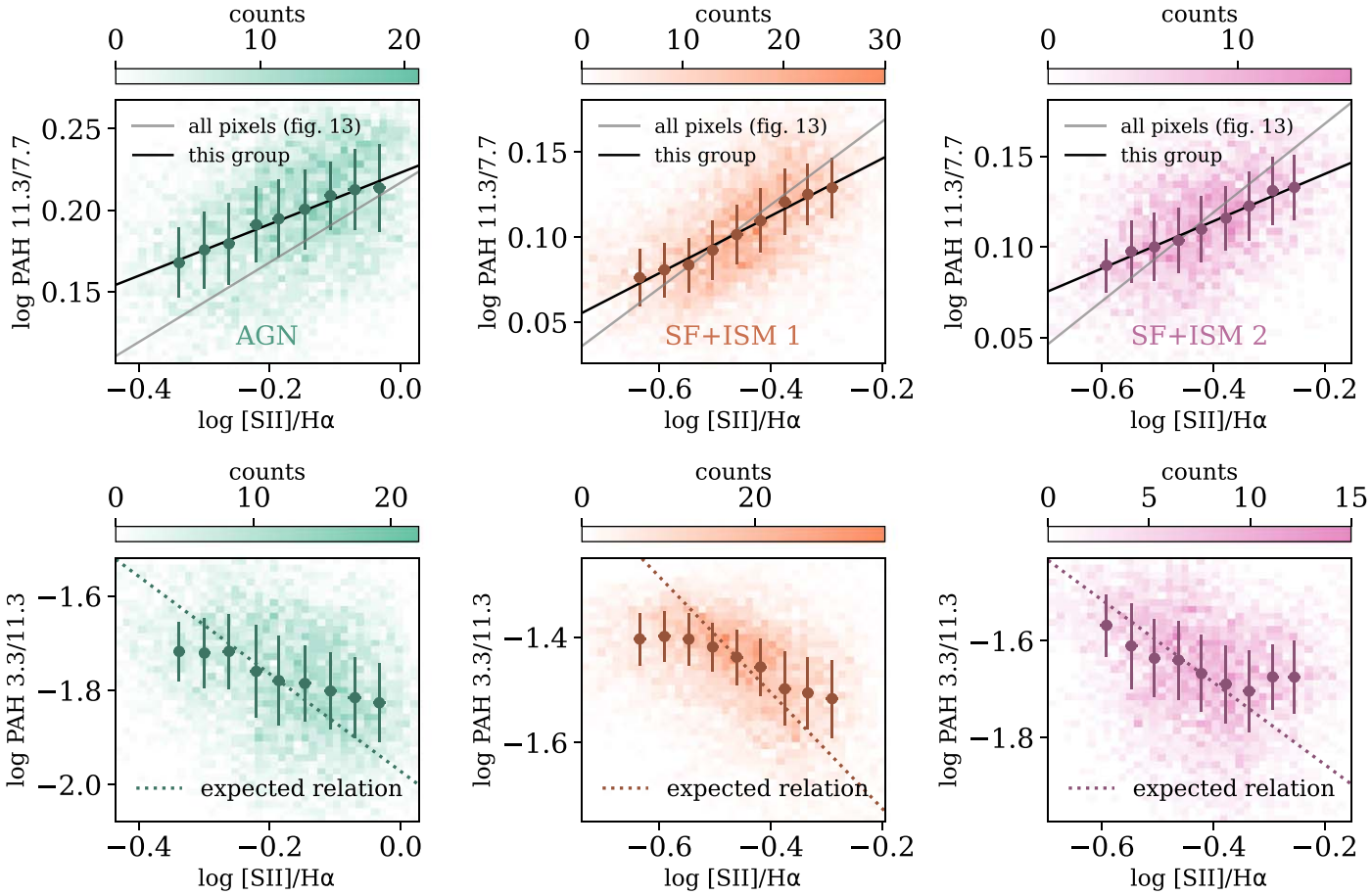

Figure 15. PAH band ratios vs. log([S II]/Hα) for individual groups. The figure shows three groups that show significant correlations between their PAH band and optical line ratios: AGN-photoionized gas (left), SF+ISM with smaller PAHs (middle), and SF+ISM with larger PAHs (right). The colors represent two-dimensional histogram of the pixels, and the points and error bars represent the median and median absolute deviation values in bins of log([S II]/Hα). The top panels show the PAH band ratio 11.3/7.7, and the bottom 3.3/11.3. The black lines at the top row represent the best linear fits to the observed relations. The light gray lines represent the best linear fit for all the groups together (Figure 13). The dashed lines at the bottom row represent the expected relation between 3.3/11.3 and [S II]/Hα, propagated from the best-fit relation between 11.3/7.7 and [S II]/Hα, using the PAH models by Draine et al. (2021), and assuming that the PAHs and gas are exposed to different parts of the same varying radiation. The expected relations are steeper than the observed relations, which may suggest variations in PAH size distributions within each group.

We find that the observed trends between the 11.3/7.7 and [S II]/Hα ratio are all in line with the "common varying radiation field" interpretation, with different types of radiation dominating in each of the groups. For the AGN-ionized gas group, the radiation may be a sequence of varying contribution of AGN and stellar light. Importantly, while the AGN may dominate the ionizing radiation in this group, resulting in the Seyfert-like line ratios we observe in Figure 7, it probably does not dominate the FUV-optical part of the radiation field. This is because these regions are at distances of a few kiloparsecs from the center, where the old stellar population probably dominates the PAH heating.

For the SF+ISM groups, the variation can be driven by a sequence in stellar ages (or equivalently, a mixing of younger and slightly older stellar populations). Since these clusters are consistent with H II ionization in the BPT diagrams, the relevant stellar ages are 1–10 Myr. For gas metallicity of $\log(Z/Z_{\odot}) \sim -0.3$, which is roughly the metallicity in the three PHANGS galaxies we consider (e.g., Williams et al. 2022), a variation of the stellar age from 1 to 10 Myr results in a variation of gas electron temperature from 12,000 to 7000 K (Byler et al. 2017), which can explain the varying [S II] emission. In particular, the radiation of a 1 Myr old star has a harder ionizing radiation and softer FUV-optical slope. The harder ionizing radiation results in higher electron temperature and thus brighter [S II] emission (Byler et al. 2017). At the same time, the softer FUV-optical slope results in a larger 11.3/7.7 PAH band ratio (Draine et al. 2021), which explains the positive relation between 11.3/7.7 and [S II]/Hα.

In the above discussion, we assume that the optical line ratios in the two SF+ISM groups vary due to a varying sequence in stellar ages. We disfavor the interpretation that the optical line ratios are driven by a varying ionization parameter, a property that is typically invoked to explain the optical line ratios in H II regions (see review by Kewley et al. 2019). For a varying ionization parameter, we expect a negative correlation between [O III]/Hβ and [S II]/Hα, while for a varying stellar age we expect a positive one (e.g., Blanc et al. 2015; Byler et al. 2017; Kewley et al. 2019). Since we observe that [O III]/Hβ increases with increasing [S II]/Hα, we favor the varying stellar age sequence.

The bottom panel of Figure 15 shows the 3.3/11.3 PAH band ratio versus [S II]/Hα. Since the 3.3/11.3 band ratio is also sensitive to variations in the incident radiation, we use the best-fitting relation between 11.3/7.7 and [S II]/Hα and propagate it to the expected relation between 3.3/11.3 and [S II]/Hα. In all the three groups, Figure 15 shows that the observed relation between $\log(3.3/11.3)$ and $\log([\text{S II}]/\text{H}\alpha)$ is shallower than that expected. This may suggest small variations in the PAH size distribution within each of the clusters, which may be related to the unaccounted impact of the radiation field on the PAH size distribution. We intend to further investigate it in future works.

If instead we interpret these relations in the context of intrinsic variations in PAH properties while keeping the radiation field fixed, then the figure suggests that PAHs are smaller and more ionized in regions dominated by older (10 Myr) stars and are larger and more neutral in regions dominated by younger (1 Myr) stars. Testing this interpretation requires a more general model that accounts for radiative transfer effects on PAH properties for a radiation field that varies in shape and intensity.

5. Extensions and Generalizations of the Methodology

5.1. Omitted Observations: Rationale, Implications, and Future Considerations

For this pilot study, we choose to concentrate on a limited set of properties derived from the PHANGS observations, leaving out many potential properties of interest that could be included in future works. The choice to limit the number of considered properties simplifies the problem significantly but also limits the discovery space to that spanned by the properties we focus on. In this section, we briefly discuss some of the omitted properties and the reasons for excluding them, and describe additional properties that may be included in future analyses of this type.

The current analysis is based on MUSE, JWST, and ALMA observations, and it does not use HST photometry. In particular, the PHANGS-HST galaxies have been observed with the filters F275W, F336W, and F438W, covering UV and blue optical wavelengths not observed by MUSE (Lee et al. 2022). Including these wavelengths may add information related to the young stellar populations, burstiness of star formation, star formation histories, and reddening toward the stars.

For the ALMA data, we use the integrated CO intensity obtained through the "broad" mask and leave out other properties, such as the width of the line. The decision to use only the integrated CO intensity, and not the line width, was motivated by our attempt to include as many pixels as possible in the analysis. Since the CO line width is available only in the "strict" mask products, it would have required us to exclude ~90% of the pixels. Even with the "broad" mask, the CO detection requirement (along with the 3.3 μm PAH detection) excludes ~30% of the pixels, restricting our analysis to regions hosting molecular gas masses equivalent to those of giant molecular clouds within 1″. In a future work, we plan to explore the use of the "flat" CO intensity maps presented by Leroy et al. (2023) and used by Belfiore et al. (2023), where CO flux measurement is available in every pixel.

For the MUSE data, we use derived properties that are based on high-S/N observables such as the integrated stellar continuum and strong emission lines. Additional properties that can be derived from the MUSE observations include the gas metallicity, ionization parameter, electron density, and temperature in the warm ionized phase; and the neutral atomic column density, traced by NaID absorption. Some of these properties require the detection of weaker lines, making maps derived with them incomplete.[49] In addition, the derivation of some of these properties is often based on physical models that may not be general enough to describe the variety of gas conditions in the PHANGS galaxies. For example, the metallicity is estimated using calibrations that are valid in H II regions, but not in the diffuse or AGN-photoionized gas (e.g., Pilyugin & Grebel 2016; Kreckel et al. 2020; Williams et al. 2022).

For the JWST data, our selected features trace primarily PAH emission, and we do not include features that trace the continuum emission from hot and larger dust grains (21 μm). Several PHANGS-JWST studies show that the F2100W filter behaves differently from the JWST filters we consider, in that it

[49] Alternatively, the spaxels can be binned until the weaker lines are measured with sufficient S/Ns. In that case, the maps will be complete but may have a much coarser resolution.

traces both gas column density and heating, showing connection to the Hα and CO emission (e.g., Belfiore et al. 2023; Hassani et al. 2023; Leroy et al. 2023; Pathak et al. 2024). In addition, by excluding the F2100W filter, we also exclude R_{PAH}, a feature that traces the PAH-to-total dust mass, and shows galaxy-wide variations that are related to the ionization parameter and metallicity (e.g., Chastenet et al. 2023a; Egorov et al. 2023; J. Sutter et al. 2024, in preparation). In the first three PHANGS-JWST galaxies studied here, the F2100W filter is saturated in the centers of two (NGC 1365 and NGC 7496), showing significant diffraction spikes on kiloparsec scales. Including this feature in the analysis would have required us to mask-out these pixels, leaving out the two galaxy centers that host AGN. We plan to revisit this choice in the future as more galaxies are included in the analysis.

Finally, it is worth noting that unsupervised machine-learning algorithms can, in principle, be applied directly to the raw data (illustrated in Figure 1). In the PHANGS survey case, these include the MUSE spectra, the photometric bands by HST and JWST, and the clean CO spectra reconstructed from the ALMA interferometric observations. Working with the raw data allows a more general analysis that is not limited by our prior physical knowledge, and may have a bigger potential for unexpected discoveries. It also provides a clear way to account for nondetections. However, working with the raw data also presents some challenges, including (i) treatment of highly correlated features, (ii) different features carrying different amounts of information, requiring some physically motivated weighting scheme, (iii) presence of catastrophic outliers due to problems in observations and reduction, and, most importantly, (iv) the challenge of interpreting the output of the unsupervised learning algorithms when applied to highly complex data sets (see discussion in B19). These challenges are also present when applying machine-learning algorithms to derived features, but the usage of raw data can make them considerably more difficult to address.

5.2. Treatment of Missing Measurements

In astronomy, as well as in other fields, it is typical to find missing values in data sets. The values could be missing due to different reasons. Values may be missing if observations were not conducted. This can occur, for example, when certain astronomical objects are included in one survey but not in another. Values may be derived from fitting observations with theoretical models, where missing values might occur if the fitting failed, or the observations are outside of the parameter space allowed by the theoretical models. Perhaps the most common type of missing values in astronomical data sets is nondetections, where observations were performed, but the astronomical object was too faint to be detected. These are also referred to as upper limits, and their statistical properties are significantly different from those of features missing because observations were not conducted. It has been known for several decades that excluding objects with upper limits from a data set may lead to significant biases in estimating summary statistics such as the mean, median, or standard deviation of a distribution, as well as biases in the output of regression and/or correlation analyses (e.g., Feigelson & Nelson 1985; Isobe et al. 1986).

In the machine-learning and data-science literature, excluding objects with missing values has also been shown to lead to significant biases in estimated parameters and to an increasing prediction error. Interestingly, studies suggest that using a noisier but more complete data set leads to smaller prediction errors compared to using cleaner but less complete data sets (e.g., Niederhut 2018). This has been shown specifically in the context of supervised machine learning, where a model is trained to predict a certain variable, and the prediction can be compared to some *ground truth* value (regression or classification task). In unsupervised machine-learning tasks, such as dimensionality reduction or clustering, the effect of excluding objects with missing values is more ambiguous, as there is typically no definitive reference point for comparison. Nevertheless, we choose to construct a noisier and more complete data set by using the CO intensity derived using the "broad" mask.

Several approaches have been suggested to handle missing values in a data set (e.g., Little & Rubin 2002; Schafer & Graham 2002; Newman 2014; Niederhut 2018). The simplest one is to replace the missing values by the mean value within a given feature. However, this simple approach of imputing the mean has been shown to lead to biased results and to underestimate the prediction variance (Niederhut 2018). Additional approaches include more sophisticated imputations based on KNN search, random forest regressor, and multiple imputation (e.g., Hruschka et al. 2003; Jonsson & Wohlin 2004; Stekhoven & Bühlmann 2011; van Buuren & Groothuis-Oudshoorn 2011; Shah et al. 2014; Niederhut 2018), or approaches that are not based on imputation but rather on estimating distances between features with missing values, and then applying the algorithm to the distance matrix directly (e.g., Eirola et al. 2013 and references therein). All of these approaches assume that the missing features are "missing at random," which means that the likelihood of a particular measurement being missing is independent of the value it could have had. This in turn assumes that any object with a missing value is well represented in the data set, and by fitting a model to the features of other objects, the missing feature value can be predicted. Once the missing values are replaced by predicted values, standard machine-learning algorithms can be applied to the data without the need to exclude objects with missing features.

In Section 3.1, we examine two different methods for missing value imputation: KNN search and random forest regression, both assume that the features are "missing at random." The missing features are optical line ratios that are based on strong emission lines: [N II], [O I], and [O III]. Our manual inspection of the pixels shows that when one of the ratios is missing the others are measured and can be used to predict the value of the missing ratio. In this particular case, the assumption of "missing at random" approximately holds, since we are working with line ratios that have been well measured throughout most of field of view.

In many cases, astronomical upper limits cannot be considered as "missing at random," since their missingness directly depends on their (low) value. Furthermore, objects with missing values are not necessarily well represented by objects that were detected, as the two may represent different populations of objects (e.g., star-forming versus passive galaxies that show significantly different correlations between their properties). In a future work where we plan to include all 19 Cycle 1 PHANGS-JWST galaxies, we intend to examine whether the assumption of "missing at random" may hold approximately. The Cycle 1 PHANGS-JWST galaxies were

observed to the same depth, but they span a distance range of 5–20 Mpc (see Table 1 in Lee et al. 2023). Therefore, convolving all the maps to the same physical scale of 150 pc, and working in units of luminosity per physical scale of 150 pc, would result in lower noise levels for the closer galaxies. Then, the CO and 3.3 μm PAH luminosities in brighter regions corresponding to the closer galaxies, along with other measured features, may be used as an approximate model for the undetected CO and 3.3 μm PAH luminosities in farther away galaxies. It may therefore be possible to build a regression model that predicts the undetected CO and 3.3 μm luminosities from the detected Hα, 7.7, 11.3, and 10 μm luminosities using pixels where all the features are measured. This, of course, assumes that the CO luminosity distributions are similar, and that the physical properties of pixels with undetected CO and 3.3 μm can be well represented by pixels with detections.

An alternative solution for including upper limits in the data set is to represent all features as probability distribution functions (PDFs). Features that are not missing may be represented by a normal distribution whose mean is the measured value, and its standard deviation is the measurement uncertainty. Upper limits may be represented by a step or a box function whose limits are defined by the survey depth and other physical considerations (e.g., flux is nonnegative). Then, instead of estimating pairwise distances between numbers, distances can be estimated using metrics that estimate distances between PDFs, such as the Wasserstein distance (e.g., Rubner et al. 2000; Ramdas et al. 2015) or the Kullback & Leibler (1951) divergence. With the pairwise distances estimated, traditional unsupervised learning algorithms can be used without excluding objects with missing values. We have developed and tested such an approach for the random forest algorithm (Reis et al. 2019), although additional tests must be performed in the case of unsupervised learning. We intend to compare different solutions for including upper limits for the 19 PHANGS-JWST galaxies in a future work.

6. Summary

The PHANGS survey has been making high-resolution observations of nearby galaxies across the electromagnetic spectrum (Figure 1). This complex multiwavelength data set offers the opportunity to explore the interplay between different processes operating on scales as small as the molecular cloud scale. This interplay is what controls the baryon life cycle in galaxies as well as their star formation, and thus is key to their overall evolution. In this work, we use unsupervised machine-learning algorithms to dissect the complex high-dimensional space spanned by the PHANGS multiwavelength observations. With these tools, we identify groups and previously unknown trends in the data, allowing us to form data-driven hypotheses from this rich data set.

We extract properties of interest from the ALMA, MUSE, JWST NIRCam, and JWST MIRI observations of the three PHANGS galaxies: NGC 0628, NGC 1365, and NGC 7496 (Section 2), from ~24,000 pixels that are half of the 150 pc resolution. These properties pertain to the stellar populations; warm ionized and cold molecular gas properties; and PAH and dust conditions (Section 3.1 and Table 2); on scales of 150 pc. We apply the dimensionality reduction algorithm UMAP to embed the high-dimensional data set into a two-dimensional plane (Section 3.2). We then use the hierarchical clustering algorithm to divide the pixels into groups (Section 3.3), each group showing distinct properties. In the process, we identify novel galaxy-wide correlations across the different regions, and crucially, use our defined groups to interpret them. Our results and their broader implications are summarized below.

(I) The identified groups have distinct multiphase gas and PAH properties (Section 4.1 and Figure 6). The 150 pc sized regions are divided into six groups, where each maps to large-scale (~kpc) and coherent structures within the galaxies. Using optical line ratios such as [O III]/Hβ, [N II]/Hα, [S II]/Hα, and [O I]/Hα, the Hα-to-CO flux ratio, the PAH band ratios 11.3/7.7 and 3.3/11.3, and additional properties, we interpret the groups, finding that they mostly differ in their multiphase gas and PAH properties. The different groups trace gas photoionized by the AGN; by standard star formation in spiral arms; extreme star formation fueled by galactic bars; and diffuse gas ionized by a combination of radiation leaking from H II regions with harder radiation from hot and evolved stars. The different groups also show different PAH properties, with some showing significantly smaller PAH size distributions than others.

(II) There is a close connection between the heating of PAHs and the ionization of the warm ionized gas (Section 4.2). We identify significant and tight correlations between different PAH band and optical line ratios (Figures 9 and 10). These correlations are seen across the entire data set, covering the star-forming regions and the ISM, the diffuse ionized gas, and the AGN-photoionized gas. They suggest a strong connection between the heating of PAHs and the ionization of the warm ionized gas on 150 pc scales.

1. The observed correlations between the PAH band 11.3/7.7 ratio and the optical line ratios [O III]/Hβ, [N II]/Hα, [S II]/Hα, and [O I]/Hα can be naturally explained in a scenario where the PAHs and ionized gas are exposed to different parts of the same spatially varying radiation field, without the need to invoke PAHs with different charge distributions (Section 4.2.1 and Figure 12). Since the PHANGS pixels trace regions with widely varying radiation fields that are a combination of young stars, old stars, and/or AGN, a variation of PAH band ratios due to the changing radiation SED is *unavoidable*. We combine PAH models with a wide range of assumed radiation fields to show that the observed slope of the relation is roughly consistent with that observed.
2. The scatter in the PAH 11.3/7.7—optical lines relations is small, ~0.03 dex, and suggests that the fraction of ionized PAHs is quite uniform on 150 pc scales in nearby galaxies. It suggests a significant self-regulation in the ISM across different regions.
3. The 11.3/7.7 PAH band ratio may potentially be used to trace the shape of the non-ionizing FUV to optical parts of the radiation field. Combining it with optical line ratios such as [O III]/Hβ and [S II]/Hα may therefore serve as a powerful diagnostic of the local radiation field, including its ionizing and nonionizing parts simultaneously.
4. The varying radiation field is expected to also impact PAH band ratios that are typically used as PAH size diagnostics (Section 4.2.2 and Figure 12). We show that the 3.3/11.3 PAH band ratio is strongly impacted by the varying radiation, and we combine empirical fits of the 11.3/7.7 versus optical line ratios' relations with PAH models to correct the 3.3/11.3 band ratio for the varying radiation field. Once the varying radiation field is accounted for, we find that PAHs tend to be smaller in

regions with low Hα/CO ratios, and larger in regions with high Hα/CO ratios. This is in line with the picture that smaller PAHs are more easily destroyed by ionizing radiation than larger PAHs. We also show that using the 3.3/11.3 band ratio directly to infer the PAH size distribution, which implicitly assumes that the incident radiation field is constant throughout, leads to completely different conclusions regarding the PAH size distribution in different regions in the galaxies (Figure 14).

5. We identify the same correlations between PAH band and optical line ratios in individual groups where the dynamical range is large enough (Section 4.2.3). The observed slopes of the correlations are comparable between the groups and comparable to the slope we find in the galaxy-wide correlations. Our analysis suggests that the variation in the 11.3/7.7 PAH band can be completely attributed to a variation in the radiation field. On the other hand, the 3.3/11.3 PAH band ratio requires an additional component, for example, a small variation of the PAH size distribution within each group.
6. The above conclusions are based on our simplified assumption that intrinsic PAH properties such as size and charge distribution do not vary with the spatially varying radiation field. In practice, however, these properties may depend on the radiation field, adding a secondary dependence of the PAH band ratios on the varying radiation field. This secondary dependence may explain the observed deviations of certain groups from the 3.3/11.3 PAH versus optical line ratio relations expected under the varying radiation field interpretation. To test this, it is necessary to construct a general radiation transfer model that accounts self-consistently for variations of PAH intrinsic properties (size and charge) as well as PAH heating due to a varying radiation field.

This analysis is based on the first three galaxies in the PHANGS-JWST survey, covering ∼24,000 75 pc sized pixels. We plan to apply a similar analysis to all the 19 PHANGS galaxies observed by JWST in Cycle 1, and to combine it with an extensive set of photoionization models to better explore the PAH–ionized gas correlations.

In this era of big data in astronomy, the abundance of multiwavelength opportunities leads to complex and high-dimensional data sets, where millions of objects are observed across the electromagnetic spectrum, sometimes also as a function of time. In many cases, we have only scratched the surface of what is possible to learn from these information-rich data sets. When the information content of the data is so large, the data itself can be used to form new hypotheses, an approach that is at at the heart of data science. This work illustrates that unsupervised machine-learning algorithms can be used to mine for novel information in complex multiwavelength data sets such as that of the PHANGS survey. To expedite the discovery process, we used simple, yet powerful, machine-learning algorithms and applied them to a set of physically motivated features. This allowed us to quickly interpret the results and focus on the scientific implications of the newly discovered correlations. On the other hand, this decision limited the discovery space to that spanned by the features we chose to consider. In future works, we plan to examine more sophisticated algorithms that can be applied directly to the raw data. While these may be more challenging to interpret, they can also potentially lead to new unexpected discoveries.

Acknowledgments

We are grateful to our reviewer, Alexandros Maragkoudakis, for helpful and constructive suggestions that helped improve this article.

D.B. is grateful to L. Armus, T. Lai, and G. Rudie, for insightful discussions regarding the interpretation of the clusters and correlations presented in this work. D.B. is supported by the Carnegie-Princeton fellowship. This project started during the GALEVO-23 program: Building a Physical Understanding of Galaxy Evolution with Data-driven Astronomy. This research was supported in part by grant NSF PHY-1748958 to the Kavli Institute for Theoretical Physics (KITP).

K.S. acknowledges funding support from JWST-GO-02107.006-A. M.B. acknowledges support by the ANID BASAL project FB210003 and by the French government through the France 2030 investment plan managed by the National Research Agency (ANR), as part of the Initiative of Excellence of Université Côte d'Azur under reference No. ANR-15-IDEX-01. G.A.B. acknowledges the support from the ANID Basal project FB210003. T.B. acknowledges support from the National Research Council of Canada via the Plaskett Fellowship of the Dominion Astrophysical Observatory. K.G. is supported by the Australian Research Council through the Discovery Early Career Researcher Award (DECRA) Fellowship (project No. DE220100766) funded by the Australian Government. E.W.K. acknowledges support from the Smithsonian Institution as a Submillimeter Array (SMA) Fellow and the Natural Sciences and Engineering Research Council of Canada.

Some of the data presented in this paper were obtained from the Mikulski Archive for Space Telescopes (MAST) at the Space Telescope Science Institute. The specific observations analyzed can be accessed via doi:10.17909/ew88-jt15. STScI is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS5–26555. Support to MAST for these data is provided by the NASA Office of Space Science via grant NAG5–7584 and by other grants and contracts.

This paper makes use of the following ALMA data: ADS/JAO.ALMA#2012.1.00650.S, 2013.1.01161.S, 2015.1.00925.S, 2017.1.00392.S, and 2017.1.00886.L. ALMA is a partnership of ESO (representing its member states), NSF (USA), and NINS (Japan), together with NRC (Canada), NSC and ASIAA (Taiwan), and KASI (Republic of Korea), in cooperation with the Republic of Chile. The Joint ALMA Observatory is operated by ESO, AUI/NRAO, and NAOJ. The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc.

Some of the data used in this work are based on observations collected at the European Southern Observatory under ESO programmes 094.C-0623 (PI: Kreckel), 095.C-0473, 098.C-0484 (PI: Blanc), 1100.B-0651 (PHANGS-MUSE; PI: Schinnerer), as well as 094.B-0321 (MAGNUM; PI: Marconi), 099.B-0242, 0100.B-0116, 098.B-0551 (MAD; PI: Carollo), and 097.B-0640 (TIMER; PI: Gadotti).

Software: Astropy (Astropy Collaboration et al. 2013, 2018, 2022), IPython (Pérez & Granger 2007), scikit-learn (Pedregosa et al. 2011), SciPy (Jones et al. 2001), matplotlib (Hunter 2007), reproject (Robitaille et al. 2020), UMAP (McInnes et al. 2018).

Appendix A
UMAP Hyperparameter Exploration

We examine the two-dimensional embeddings by UMAP for a range of distance metrics and `n_neighbors` parameters. For the distance metric, we examine 10 metrics from the list of proposed metrics in the UMAP python library[50]: Euclidean, Manhattan, Chebyshev, Minkowski, Canberra, Braycurtis, Mahalanobis, WMinkowski, Cosine, and Correlation distance. In addition to these, we estimate the unsupervised random forest (URF) distance matrix, used in Baron & Poznanski (2017) to perform anomaly detection on galaxy spectra. The motivation to examine the URF-based distance is that the random forest algorithm can be generalized to handle missing values, upper limits, and measurement uncertainties (Reis et al. 2019), and can thus be used to include objects with missing features in our future work of the full PHANGS-JWST sample.

Figure A1 shows the two-dimensional embeddings obtained with UMAP for different distance metrics, and using `n_neighbors=25` and `min_dist=0`. These are compared to the embedding adopted in this paper (hyperparameters: `metric=correlation`, `n_neighbors=10`, and `min_dist=0`). The points are color-coded according to their location in our adopted embedding, using rectangles we defined manually (top left panel). The figure shows that objects that are within a given neighborhood in one of the two-dimensional embeddings are also in the same neighborhood in another. This suggests that the global structure of the data does not depend significantly on the distance metric, and that the clusters that would have been identified using different metrics should be roughly similar to those we identify using our adopted embedding.

Figure A2 compares the two-dimensional embeddings by UMAP for different `n_neighbors` parameters ranging from 10 to 500. The points are color-coded according to their designated clusters we defined manually in Figure A1 (top left panel). As expected, the general structure in the two-dimensional embedding becomes more connected and less clustered with increasing `n_neighbors`. However, similarly to the previous figure, the figure shows that objects remain in the same rough neighborhood for different choices of `n_neighbors`, suggesting that roughly the same clusters would have been identified using different values of `n_neighbors`.

Dimensionality reduction algorithms such as UMAP attempt to populate points in a low-dimensional space such that their distance distribution with respect to their neighbors will be as close as possible to the distance distribution in the high-dimensional space the observations originally span. As a result, the two-dimensional embedding can be treated as invariant under affine transformations of the data, including flipping, rotation, and translation (see, e.g., McInnes et al. 2018). For example, the top left and bottom right panels in Figure A1 are flipped one with respect to the other, but showing very similar neighborhoods and structures. These representations are considered very close to one another.

[50] https://umap-learn.readthedocs.io/en/latest/parameters.html#metric

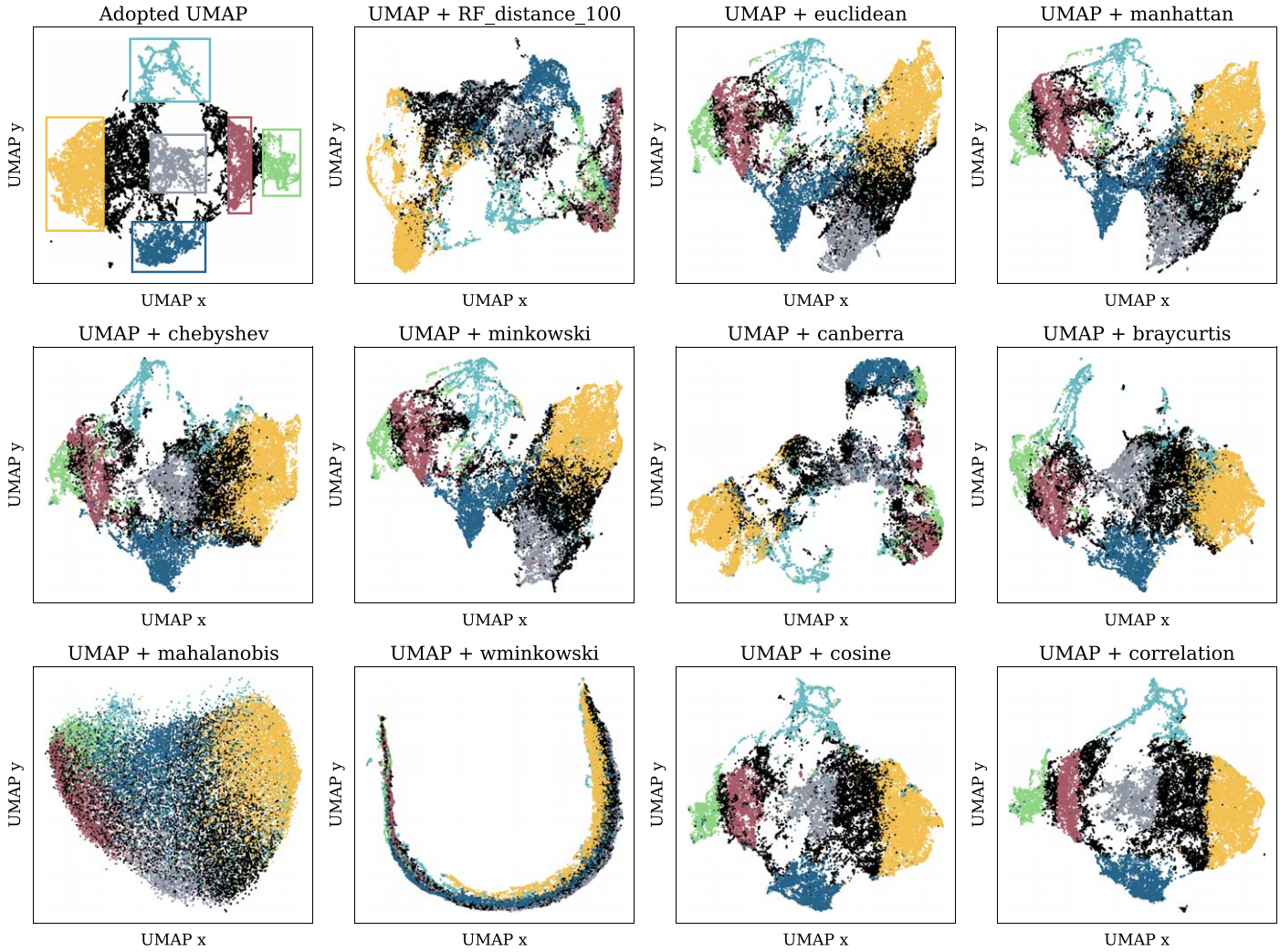

Figure A1. Comparison of UMAP two-dimensional embeddings for different distance metrics. The top left panel shows the UMAP two-dimensional embedding adopted in this study, which was obtained using the following hyperparameters: `metric=correlation`, `n_neighbors=10`, and `min_dist=0`. The six rectangles identify neighborhoods in the two-dimensional space that might be considered as different clusters. The other panels show the two-dimensional embedding obtained when varying the distance metric, assuming `n_neighbors=25` and `min_dist=0`. The points are color-coded according to the designated clusters of the pixels in our adopted two-dimensional embedding on the top left. These panels show that objects that were identified to be in the same cluster using one metric will also be identified within the same cluster using another metric. This shows that the global structure of the data in the two-dimensional embedding does not depend significantly on the assumed metric.

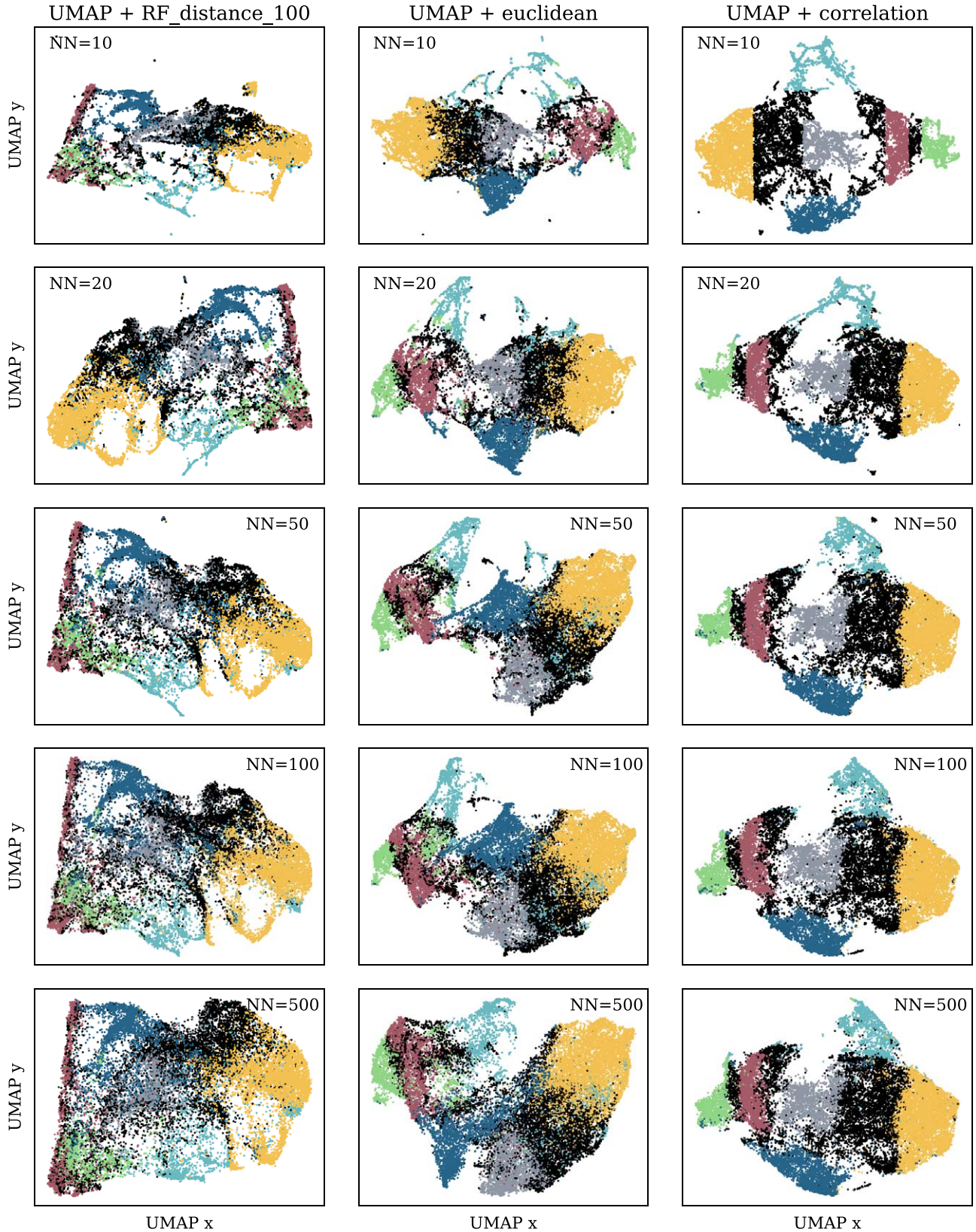

Figure A2. Comparison of UMAP two-dimensional embeddings for different `n_neighbors` parameters. The panels show the resulting two-dimensional embedding assuming different metrics and different values of `n_neighbors`. In each panel, the points are colored according to the designated clusters of the pixels in our adopted two-dimensional embedding (top left in Figure A1). For a given distance metric (given column), increasing the `n_neighbors` parameter from 10 to 500 (top to bottom) does not change the general structure of the data in the two-dimensional space significantly, suggesting that comparable clusters would have been identified for different choices of hyperparameters.

Appendix B
UMAP and Clustering Interpretation

Figure B1 shows the Pearson correlation coefficient between different pairs of features. It includes the correlations calculated considering all the 24,007 PHANGS pixels, as well as those obtained when considering pixels from individual clusters. The correlation coefficients are listed in the figure only if their associated p-value is smaller than 0.01. The figure shows significant correlations between different PAH band and optical line ratios. In Figures B2, B3, B4, B5, and B6, we show the distribution of the clusters in some of the features we considered in our analysis. We used these figures to interpret the clusters we identified.

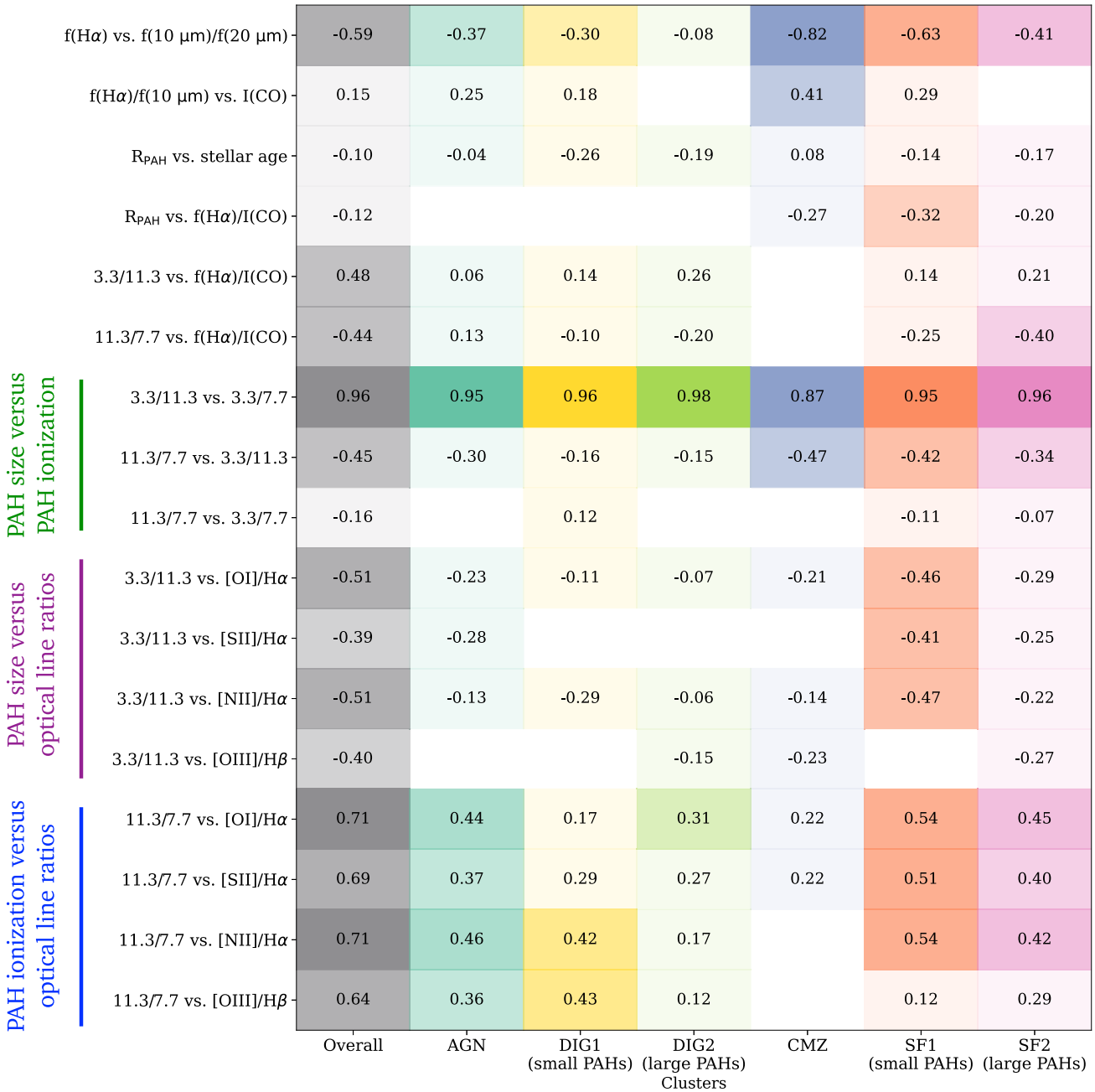

Figure B1. Correlations between various features considered in our analysis. The matrix shows the Pearson correlation coefficient between pairs of features as indicated on the left, where empty cells correspond to correlations with p-values equal or larger than 0.01. Each row corresponds to a pair of features we consider. Each column marks the set of pixels considered for the correlation estimation, where the first column shows the correlations using all the 24,007 pixels we considered, and the rest correspond to pixels in the different clusters we identified. High correlation coefficients are marked with brighter colors, and low coefficients with fainter colors. We identify significant and tight correlations between PAH band and optical line ratios, suggesting a strong connection between them.

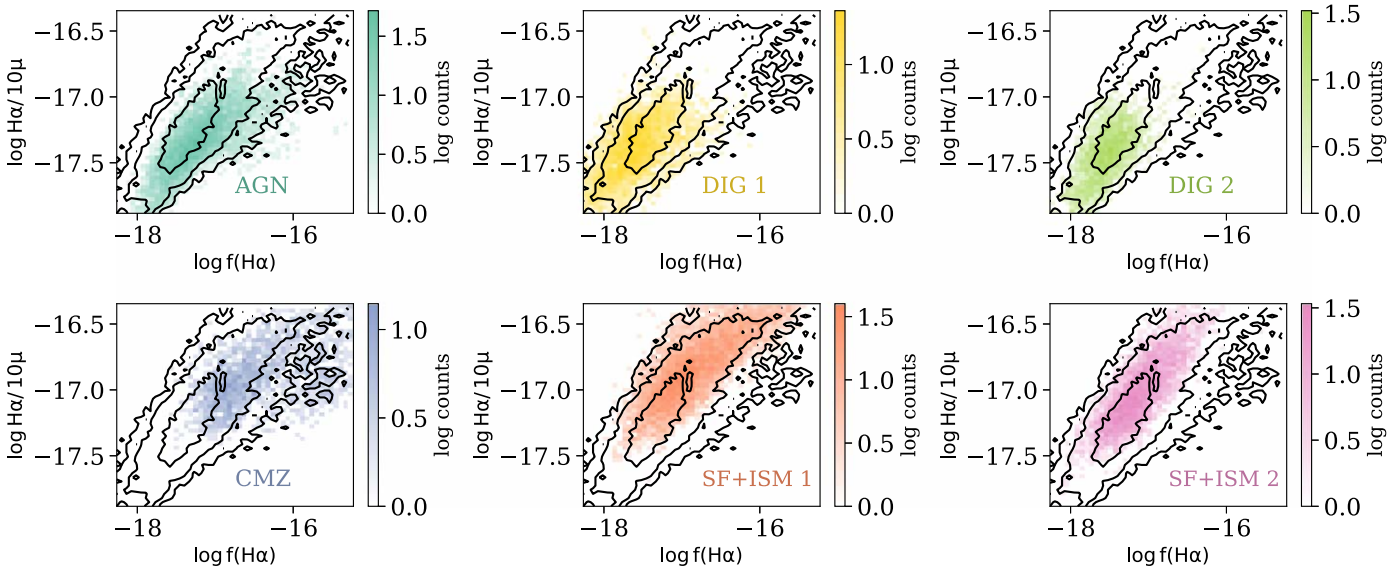

Figure B2. logHα/10 μm vs. logf(Hα) for the identified clusters. The black contours represent the two-dimensional distribution of all the pixels in the data set, and the colored colormaps represent the distribution in each cluster.

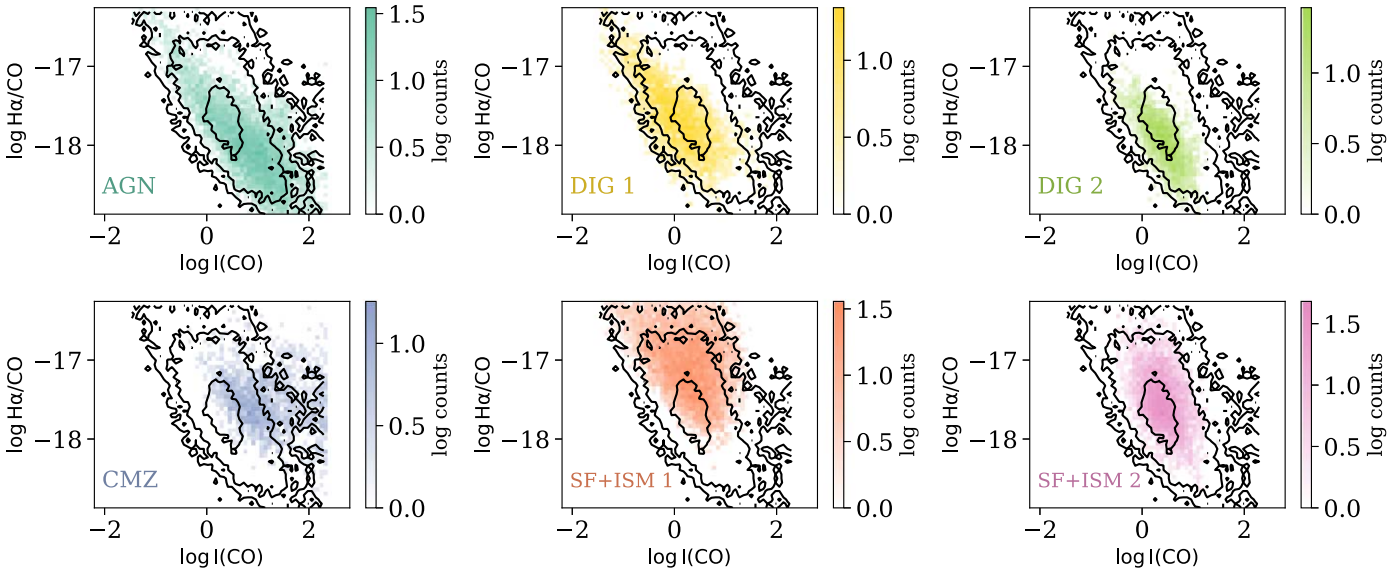

Figure B3. logHα/CO vs. logI(CO) for the identified clusters. The black contours represent the two-dimensional distribution of all the pixels in the data set, and the colored colormaps represent the distribution in each cluster.

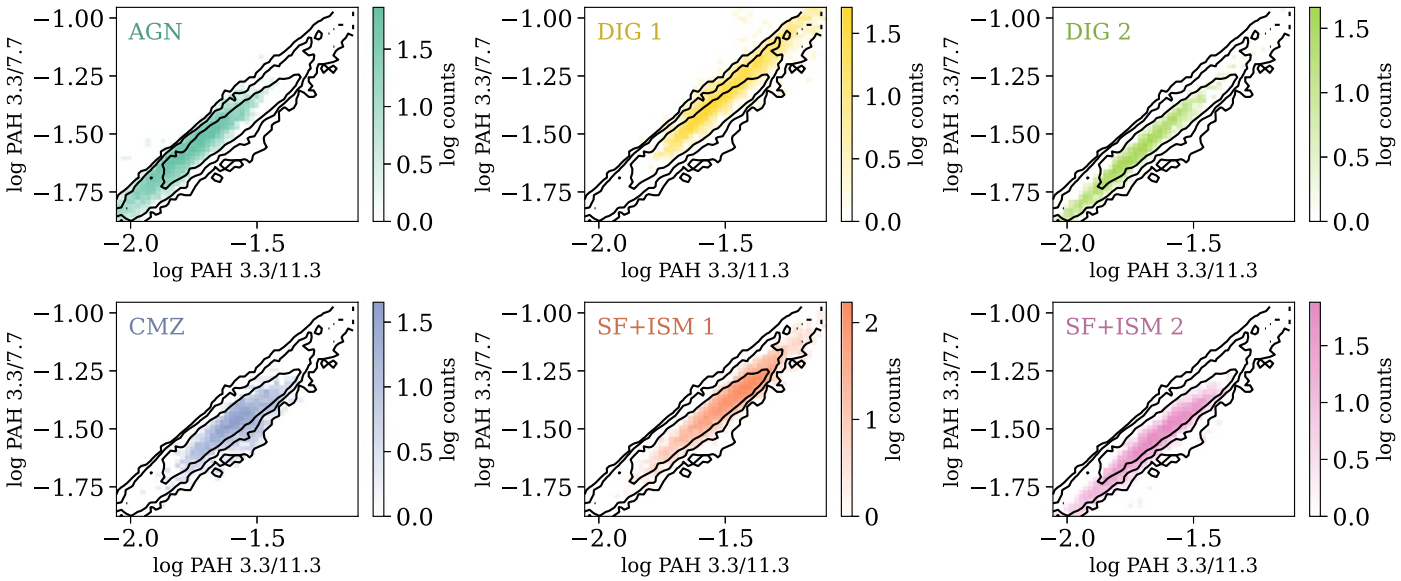

Figure B4. PAH band 3.3/7.7 μm vs. 3.3/11.3 μm ratios for the identified clusters. The black contours represent the two-dimensional distribution of all the pixels in the data set, and the colored colormaps represent the distribution in each cluster.

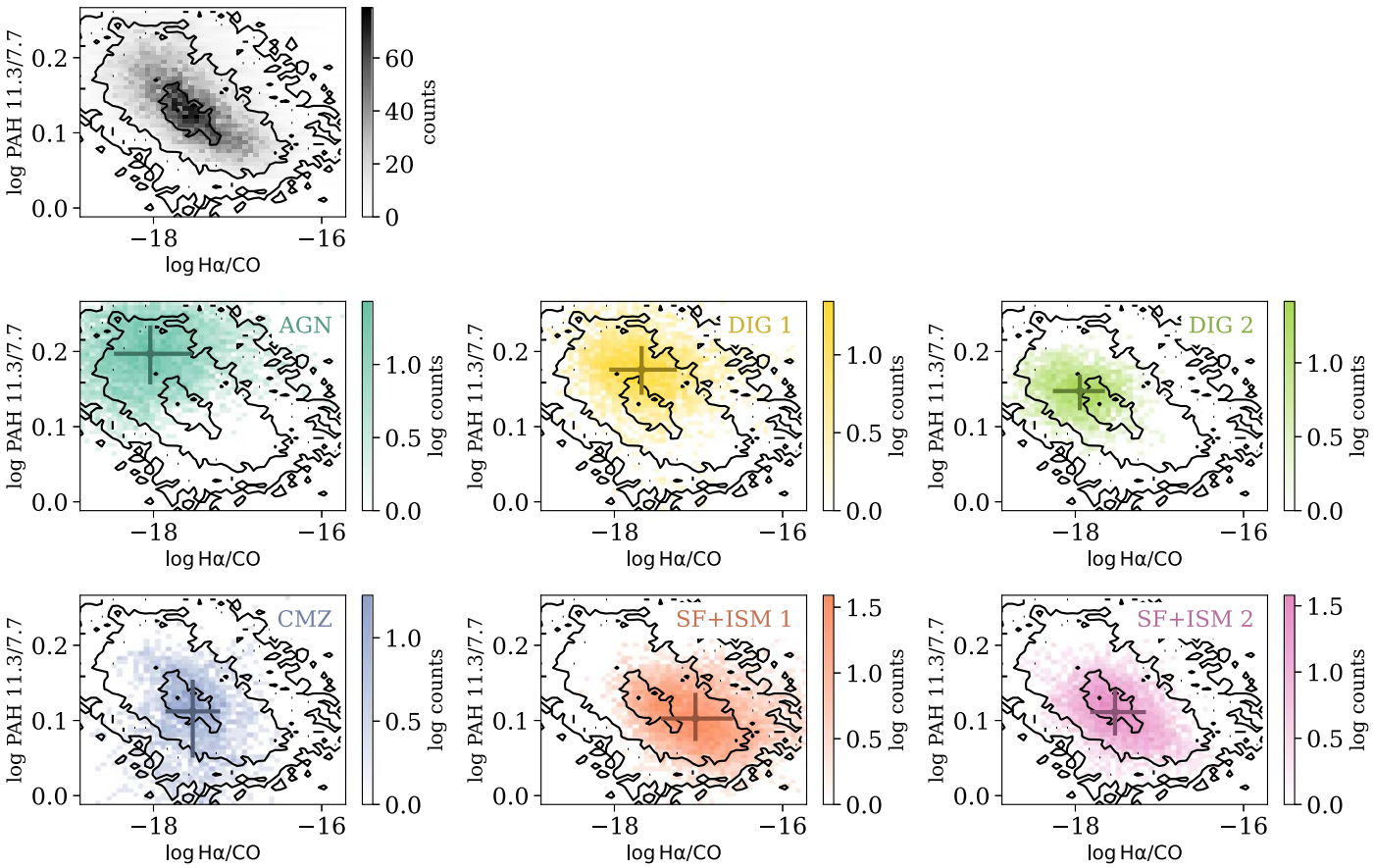

Figure B5. PAH band ratio 11.3/7.7 μm vs. the Hα/CO ratio. The top left panel shows the distribution for all the PHANGS pixels, and the rest show the distribution for each individual cluster. The black contours represent the distribution of all the pixels in the data set.

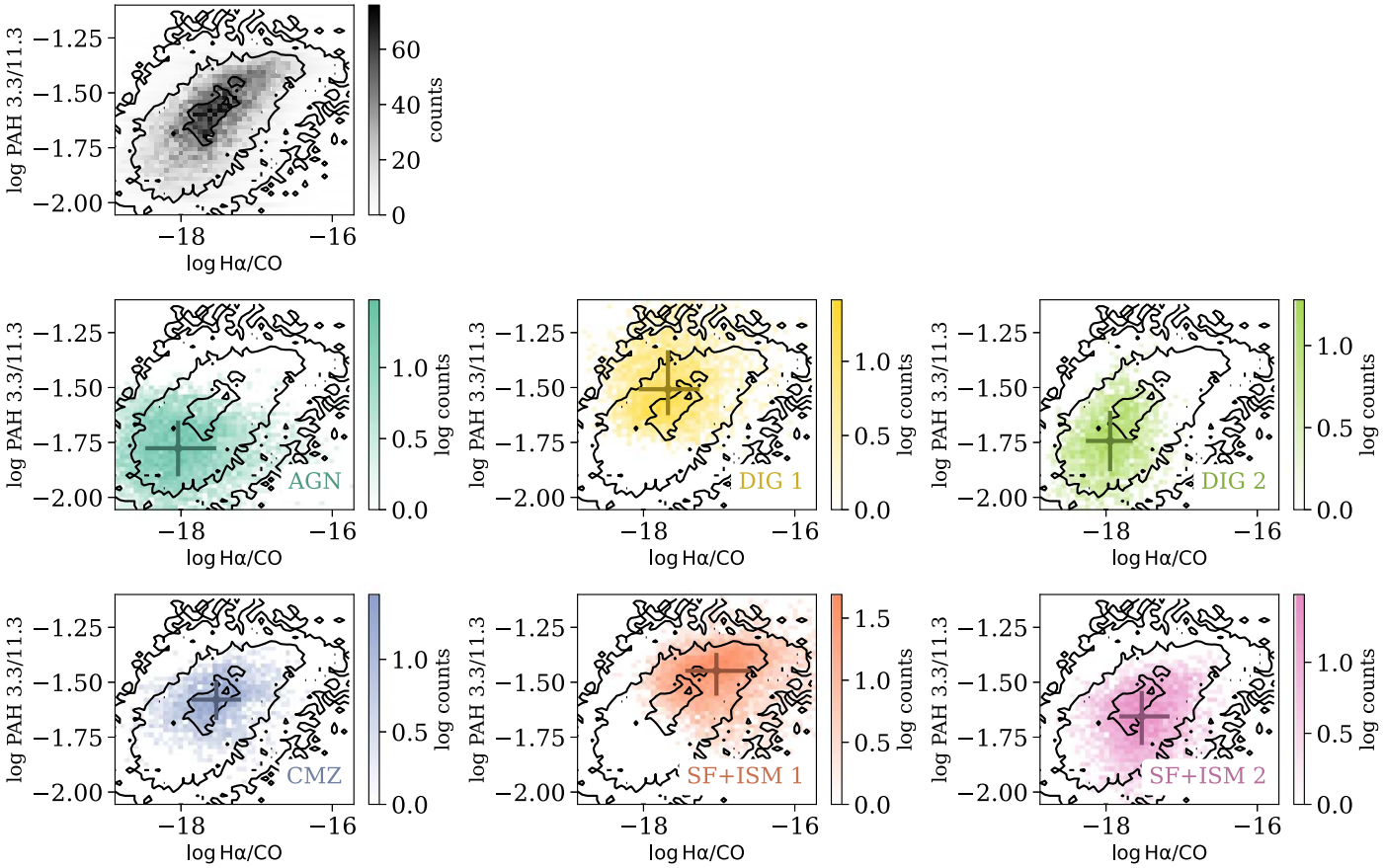

Figure B6. PAH band ratio 3.3/11.3 μm vs. the Hα/CO ratio. The top left panel shows the distribution for all the PHANGS pixels, and the rest show the distribution for each individual cluster. The black contours represent the distribution of all the pixels in the data set.

Appendix C
Models Used to Interpret the PAH–ionized Gas Connection

C.1. PAH Models

We make use of the models presented in Draine et al. (2021). These models predict the infrared emission from dust under a wide range of assumptions, including incident radiation with varying SEDs, PAH size distributions, and PAH charge distributions. We use models with solar metallicity. For the SED, we use two sets of single-age stellar population models: BC03 (Bruzual & Charlot 2003) and BPASS (Eldridge et al. 2017), with stellar ages of 3, 10, 100, 300 Myr, and 1 Gyr. We also consider the models calculated with the `mMMP` and `m31bulge` SEDs (see Draine et al. 2021 for additional details). Therefore, we consider 12 different SEDs in our analysis. We consider the three different options for the PAH size distribution, small, standard, and large; and the three options for the PAH charge distribution, low, standard, and high.

To estimate the 11.3/7.7 and 3.3/11.3 PAH band ratios, we follow the procedure outlined in Draine et al. (2021). We parameterize the radiation field SEDs using their FUV-to-optical luminosity ratio, where the FUV luminosity was defined between $\lambda = 1350\,\mathring{A}$ and $\lambda = 1780$ Å, and the optical was defined between $\lambda = 3000$ Å and $\lambda = 4000$ Å. Figure C1 shows how the PAH band ratios 11.3/7.7 and 3.3/11.3 change with the FUV-to-optical luminosity ratio, the PAH size distribution, and the PAH charge distribution.

We estimate the expected change in the PAH band ratios, $\Delta\log(11.3/7.7)$ and $\Delta\log(3.3/11.3)$, for varying PAH size and charge distributions. These are represented as red and blue arrows in Figures 12 and 13 in the main text.

For the variation with the PAH charge distribution, we use models with the standard PAH size distribution, and for each FUV-to-optical ratio, we calculate $\Delta\log(11.3/7.7)$ and $\Delta\log(3.3/11.3)$ when changing the PAH ionization from low to standard and from standard to high. The resulting $\Delta\log(11.3/7.7)$ and $\Delta\log(3.3/11.3)$ do not vary significantly for different FUV-optical slopes and when considering low–standard versus standard–high changes. Therefore, we take the median over these values, with the final adopted values $\Delta\log(11.3/7.7) = -0.11$ dex and $\Delta\log(3.3/11.3) = -0.04$ dex. These represent the expected changes in the PAH band ratios when changing the PAH charge distribution from low to standard or from standard to high, and they are marked with blue arrows in Figures 12 and 13. A change of the PAH charge distribution from low to high would result in twice the change.

For the dependence on the PAH size distribution, we use models calculated with the standard PAH ionization distribution, and estimate $\Delta\log(11.3/7.7)$ and $\Delta\log(3.3/11.3)$ when changing the PAH size distribution from large to standard and from standard to small. These values also do not vary significantly with the FUV-to-optical luminosity ratio and when considering large to standard versus standard to small changes. The final adopted values, which are marked with red arrows in Figures 12 and 13, are $\Delta\log(11.3/7.7) = -0.022$ dex and $\Delta(3.3/11.3)\log = 0.24$ dex. They represent the change in the

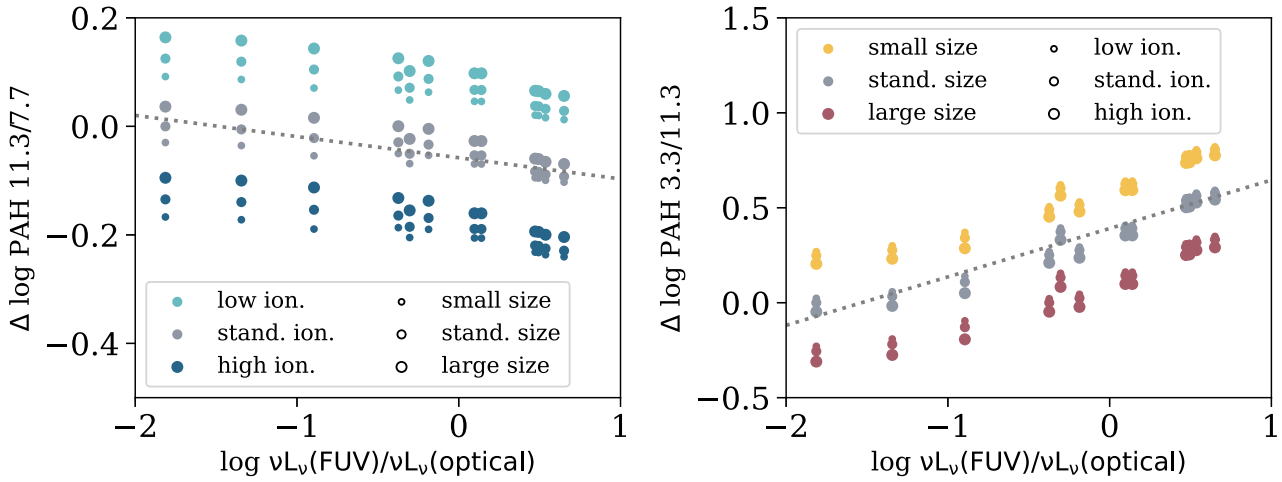

Figure C1. Relation between PAH band ratios and the FUV-optical slope for the Draine et al. (2021) models. The panels show the PAH band ratios 11.3/7.7 μm and 3.3/11.3 μm vs. the FUV-to-optical luminosity ratio of the SEDs the PAHs were exposed to. The figures show different PAH size and charge distributions. The 11.3/7.7 μm band ratio is primarily sensitive to the PAH charge distribution, and increases for more neutral PAHs. It is also sensitive to the FUV-optical slope, and decreases for softer FUV-optical slopes. The ratio is only weakly related to the PAH size distribution, and can change by ∼0.03 dex when the PAH sizes change from standard to small, or from large to standard. The 3.3/11.3 μm ratio is primarily sensitive to the PAH size distribution, and increases for smaller PAHs. It is also sensitive to the FUV-optical slope, and increases for softer SEDs. It only weakly depends on the PAH size distribution, and can change by ∼0.05 dex when changing the PAH charge distribution from low to standard, or from standard to high. The gray dashed lines represent the best-fitting linear relations between the 11.3/7.7 μm and 3.3/11.3 μm band ratios and the FUV-to-optical luminosity ratio, as described in Appendix C.1.

PAH band ratios when changing the PAH size distribution from large to standard and from standard to small. A change of the PAH size distribution from large to small would result in twice the change.

The dashed gray lines in Figure C1 are the best-fitting linear relations between the PAH band ratios log(11.3/7.7) and log(3.3/11.3) and the FUV-to-optical luminosity ratio. These lines are used to translate the observed relation between log(11.3/7.7) and log([O III]/Hβ) to the expected relation between log(3.3/11.3) and log([O III]/Hβ), assuming that the ratios vary only due to varying SED shape. While it is clear from Figure C1 that the relations have some curvature, we choose to fit linear relations for simplicity. The relations are as follows:

$$\log \mathrm{PAH}\frac{11.3}{7.7} = -0.039 \times \log \frac{\nu \mathrm{L}_\nu(\mathrm{FUV})}{\nu \mathrm{L}_\nu(\mathrm{optical})} - 0.057, \quad \text{(C1)}$$

and

$$\log \mathrm{PAH}\frac{3.3}{11.3} = 0.25 \times \log \frac{\nu \mathrm{L}_\nu(\mathrm{FUV})}{\nu \mathrm{L}_\nu(\mathrm{optical})} + 0.39. \quad \text{(C2)}$$

To calculate the dashed and dotted lines in the right panels of Figures 12 and 13, we first fit the relations between log(11.3/7.7) and log([O III]/Hβ) using 2° and 3° polynomials. The relations are given by the following:

$$y = 0.15x + 0.19, \quad \text{(C3)}$$

and

$$y = -0.087x^2 + 0.084x + 0.18, \quad \text{(C4)}$$

where $y = \log \mathrm{PAH}\frac{11.3}{7.7}$, and $x = \log([\mathrm{O\ III}]/\mathrm{H}\beta)$. Similarly, we fit relations between log(11.3/7.7) and log([S II]/Hα) using 2° and 3° polynomials:

$$y = 0.22x + 0.21, \quad \text{(C5)}$$

and

$$y = 0.53x^2 + 0.61x + 0.26, \quad \text{(C6)}$$

where $y = \log \mathrm{PAH}\frac{11.3}{7.7}$, and $x = \log([\mathrm{S\ II}]/\mathrm{H}\alpha)$.

To obtain the dotted (dashed) line in right panel of Figure 12, we combine Equations (C3)–(C4) with Equation (C1) to obtain a relation between log([O III]/Hβ) and the FUV-to-optical ratio. We then populate this relation into Equation (C2) to obtain the expected relation between log(3.3/11.3) and log([O III]/Hβ). The resulting relations are as follows:

$$y = -x - 2.14, \quad \text{(C7)}$$

and

$$y = 0.57x^2 - 0.55x - 2.12, \quad \text{(C8)}$$

where $y = \log \mathrm{PAH}\frac{3.3}{11.3}$, and $x = \log([\mathrm{O\ III}]/\mathrm{H}\beta)$.

To obtain the dotted (dashed) line in the right panel of Figure 13, we combine Equation (C5)–(C6) with Equation (C1) to obtain a relation between log([S II]/Hα) and the FUV-to-optical ratio. We then populate this relation into Equation (C2) to obtain the expected relation between log(3.3/11.3) and log([S II]/Hα). The resulting relations are as follows:

$$y = -1.6x - 2.34, \quad \text{(C9)}$$

and

$$y = -3.5x^2 - 4.02x - 2.54, \quad \text{(C10)}$$

where $y = \log \mathrm{PAH}\frac{3.3}{11.3}$, and $x = \log([\mathrm{S\ II}]/\mathrm{H}\alpha)$.

C.2. Spectral Energy Distributions

To examine the relation between the FUV-optical slope and the hardness of the ionizing radiation, we considered several stellar population models, all of which are shown in Figure C2 below. They include the stellar populations used by Draine et al. (2021) for the PAH models described above. We used the flexible stellar population synthesis (FSPS) code by Conroy

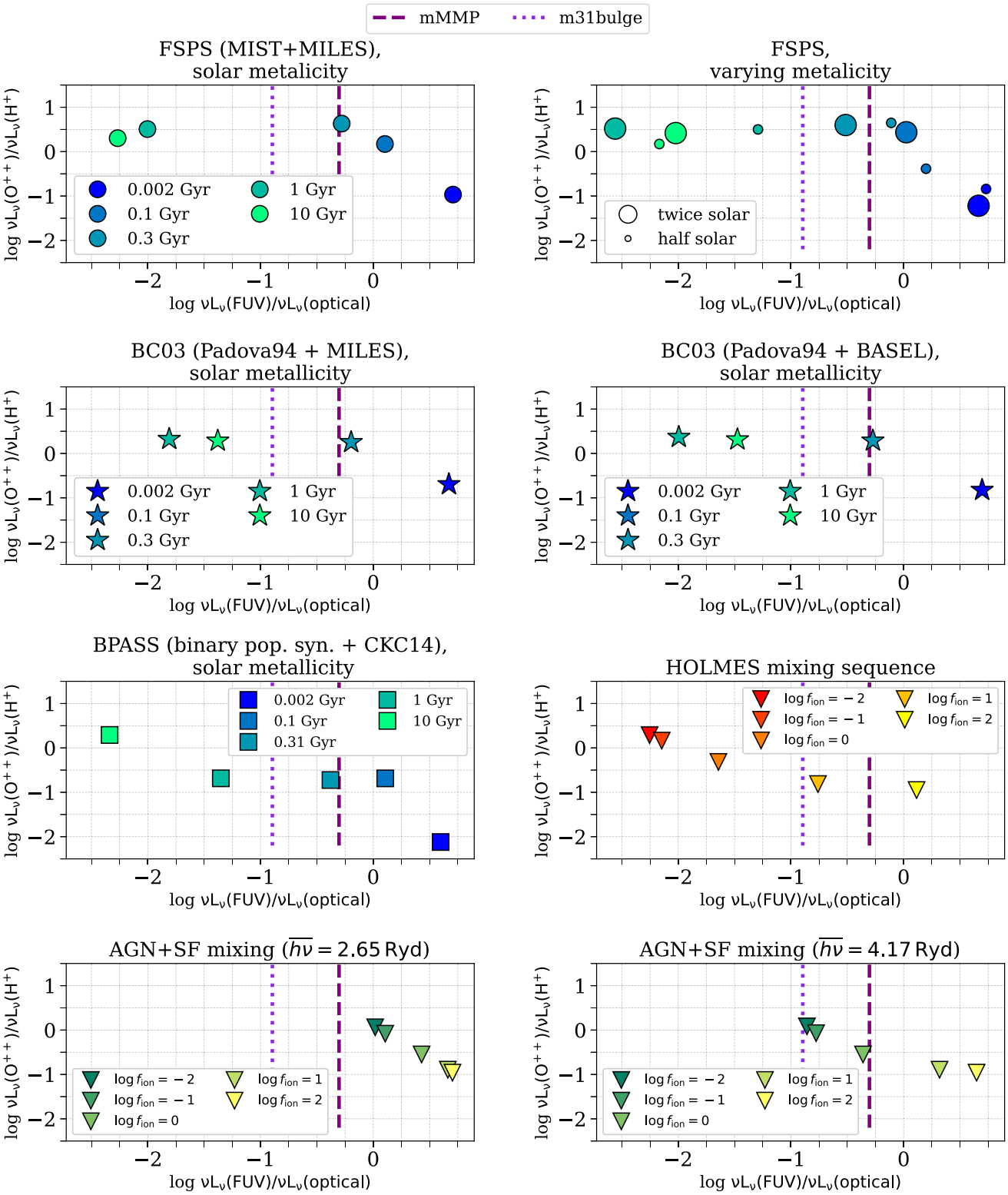

Figure C2. $\nu L_\nu(O^{++})/\nu L_\nu(H^+)$ vs. the FUV-to-optical luminosity ratio for a range of different assumed SEDs. Each panel shows the ratios for different sets of SEDs, where we examined single stellar populations estimated with different codes, isochrones, and stellar libraries, as well as assuming different metallicities and stellar ages. For the half/twice solar metallicity, we examined the same ages as those presented in the other solar metallicity panels. We also considered SEDs that are a combination of young and old stellar populations (HOLMES mixing sequence), which are used to interpret the optical line ratios seen in the diffuse ionized gas, and SEDs that are a combination of stellar and AGN radiation. For comparison, we show the UV-to-optical luminosity ratios of the Draine et al. (2021) `mMMP` and `m31bulge` SEDs as dashed and dotted purple lines. The figure shows that harder ionizing radiation (increasing y-axis) is typically associated with softer FUV-optical slope (decreasing x-axis), for a wide range of assumptions of the SED.

et al. (2009) to simulate single stellar populations. We used the MIST isochrones (Choi et al. 2016; Dotter 2016) and the MILES stellar library (Vazdekis et al. 2010). We considered stellar population ages of 2, 100, 300 Myr, 1, and 10 Gyr, and considered solar metallicity, as well as half and twice solar. We also considered the Bruzual & Charlot (2003) models,

calculated with the Padova 1994 tracks, using two different stellar libraries: MILES and BaSeL (see Bruzual & Charlot 2003 for details). We considered stellar population ages of 2, 100, 300 Myr, 1, and 10 Gyr, assuming solar metallicity. We also considered the BPASS models that include a treatment of binary stars (e.g., Eldridge et al. 2017), considering the same stellar ages and a solar metallicity.

We constructed the HOLMES mixing sequence described by Belfiore et al. (2022), which was used to interpret the optical line ratios seen in the diffuse ionized gas in PHANGS galaxies. The spectra are a combination of a young stellar spectrum (2 Myr), which represents the radiation leaking from H II regions, with an old stellar spectrum (10 Gyr), which represents the radiation of hot and evolved stars. The mixing between these spectra is determined through $f_{\rm ion}$, which represents the ratio of the ionizing fluxes from the young and old stars. For example, $\log f_{ion} = 1$ represents a case where the ionizing flux of the young stars is 10 times larger than the ionizing flux from the old stars. $\log f_{\rm ion} = 0$ represents the case that the ionizing flux from the young stars equals to the ionizing radiation from the old stars. To construct this sequence, we used the FSPS models with ages 2 Myr and 10 Gyr, assuming solar metallicity, and considering mixing fractions in the range $\log f_{\rm ion}$ between -2 and 2.

Finally, we constructed two AGN+SF mixing sequences. For the AGN accretion disk, the SED consists of a combination of an optical-UV continuum emitted by an optically thick geometrically thin accretion disk, and an additional X-ray power-law source that extends to 50 keV with a photon index of $\Gamma = 1.9$. The normalization of the UV (2500 Å) to X-ray (2 keV) flux is defined by $\alpha_{\rm OX}$, which we take to be 1.37. We consider two models that differ in the mean energy of an ionizing photon (2.65 and 4.17 Ryd), which correspond to different different black hole masses, accretion rates, and spins (see Table A1 in Baron & Netzer 2019). We then construct a mixing sequence with a young stellar spectrum (2 Myr), considering mixing fractions in the range $\log f_{\rm ion}$ between -2 and 2.

For each of these SEDs, we estimate the FUV-to-optical luminosity ratio by integrating over the spectra in the ranges 1350–1780 Å (FUV) and 3000–4000 Å (optical). To probe the hardness of the ionizing SED, we estimate νL_ν at 353 and 912 Å, which correspond to photon energies that are required to ionize oxygen twice (probed by the [O III] line) and hydrogen (probed by the Hα line). Figure C2 shows the luminosity ratio $\nu L_\nu(O^{++})/\nu L_\nu(H^+)$, which traces the slope of the ionizing radiation, versus the FUV-to-optical luminosity ratio, which traces the hardness of the FUV-optical part of the SED. The figure includes all the models described in this section. The figure shows that a harder ionizing radiation is typically associated with a softer FUV-optical slope for a wide range of SEDs, including single stellar populations, mixing of young and old stars, and mixing of stellar and AGN SEDs. The slopes of the gray arrow in Figures 12 and 13 were estimated by combining the observed slopes in Figures C1 and C2.

ORCID iDs

Dalya Baron https://orcid.org/0000-0003-4974-3481
Karin M. Sandstrom https://orcid.org/0000-0002-4378-8534
Erik Rosolowsky https://orcid.org/0000-0002-5204-2259
Oleg V. Egorov https://orcid.org/0000-0002-4755-118X
Ralf S. Klessen https://orcid.org/0000-0002-0560-3172
Adam K. Leroy https://orcid.org/0000-0002-2545-1700
Médéric Boquien https://orcid.org/0000-0003-0946-6176
Eva Schinnerer https://orcid.org/0000-0002-3933-7677
Francesco Belfiore https://orcid.org/0000-0002-2545-5752
Brent Groves https://orcid.org/0000-0002-9768-0246
Jérémy Chastenet https://orcid.org/0000-0002-5235-5589
Daniel A. Dale https://orcid.org/0000-0002-5782-9093
Guillermo A. Blanc https://orcid.org/0000-0003-4218-3944
José E. Méndez-Delgado https://orcid.org/0000-0002-6972-6411
Eric W. Koch https://orcid.org/0000-0001-9605-780X
Kathryn Grasha https://orcid.org/0000-0002-3247-5321
Mélanie Chevance https://orcid.org/0000-0002-5635-5180
David A. Thilker https://orcid.org/0000-0002-8528-7340
Dario Colombo https://orcid.org/0000-0001-6498-2945
Thomas G. Williams https://orcid.org/0000-0002-0012-2142
Debosmita Pathak https://orcid.org/0000-0003-2721-487X
Jessica Sutter https://orcid.org/0000-0002-9183-8102
Toby Brown https://orcid.org/0000-0003-1845-0934
John F. Wu https://orcid.org/0000-0002-5077-881X
Josh E. G. Peek https://orcid.org/0000-0003-4797-7030
Eric Emsellem https://orcid.org/0000-0002-6155-7166
Kirsten L. Larson https://orcid.org/0000-0003-3917-6460
Justus Neumann https://orcid.org/0000-0002-3289-8914

References

Abdurro'uf, Accetta, K., Aerts, C., et al. 2022, ApJS, 259, 35
Ali, M., Jones, M. W., Xie, X., & Williams, M. 2019, Vis. Comput., 35, 1013
Allamandola, L. J., Hudgins, D. M., & Sandford, S. A. 1999, ApJL, 511, L115
Allen, M. G., Groves, B. A., Dopita, M. A., Sutherland, R. S., & Kewley, L. J. 2008, ApJS, 178, 20
Aniano, G., Draine, B. T., Gordon, K. D., & Sandstrom, K. 2011, PASP, 123, 1218
Astropy Collaboration, Price-Whelan, A. M., Lim, P. L., et al. 2022, ApJ, 935, 167
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
Baldwin, J. A., Phillips, M. M., & Terlevich, R. 1981, PASP, 93, 5
Baron, D. 2019, arXiv:1904.07248
Baron, D., & Netzer, H. 2019, MNRAS, 486, 4290
Baron, D., & Poznanski, D. 2017, MNRAS, 465, 4530
Becht, E., McInnes, L., Healy, J., et al. 2018, NatBi, 37, 38
Belfiore, F., Leroy, A. K., Williams, T. G., et al. 2023, A&A, 678, A129
Belfiore, F., Santoro, F., Groves, B., et al. 2022, A&A, 659, A26
Bellm, E. 2014, Proc. Third Hot-wiring the Transient Universe Workshop, ed. P. R. Wozniak et al., 27, http://www.slac.stanford.edu/econf/C131113.1/
Blanc, G. A., Kewley, L., Vogt, F. P. A., & Dopita, M. A. 2015, ApJ, 798, 99
Boersma, C., Bregman, J., & Allamandola, L. J. 2016, ApJ, 832, 51
Boersma, C., Bregman, J., & Allamandola, L. J. 2018, ApJ, 858, 67
Boselli, A., Boissier, S., Heinis, S., et al. 2011, A&A, 528, A107
Boselli, A., Fossati, M., Ferrarese, L., et al. 2018, A&A, 614, A56
Brown, T., Wilson, C. D., Zabel, N., et al. 2021, ApJS, 257, 21
Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000
Bundy, K., Bershady, M. A., Law, D. R., et al. 2015, ApJ, 798, 7
Byler, N., Dalcanton, J. J., Conroy, C., et al. 2019, AJ, 158, 2
Byler, N., Dalcanton, J. J., Conroy, C., & Johnson, B. D. 2017, ApJ, 840, 44
Cao, J., Spielmann, M., Qiu, X., et al. 2019, Natur, 566, 496
Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, ApJ, 345, 245
Carter, S., Armstrong, Z., Schubert, L., Johnson, I., & Olah, C. 2019, Distill,
Chastenet, J., Sandstrom, K., Chiang, I.-D., et al. 2019, ApJ, 876, 62
Chastenet, J., Sutter, J., Sandstrom, K., et al. 2023a, ApJL, 944, L11
Chastenet, J., Sutter, J., Sandstrom, K., et al. 2023b, ApJL, 944, L12
Choi, J., Dotter, A., Conroy, C., et al. 2016, ApJ, 823, 102
Chown, R., Sidhu, A., Peeters, E., et al. 2023, arXiv:2308.16733
Cid Fernandes, R., Stasińska, G., Schlickmann, M. S., et al. 2010, MNRAS, 403, 1036

Conroy, C., Gunn, J. E., & White, M. 2009, ApJ, 699, 486
Côté, P., Blakeslee, J. P., Ferrarese, L., et al. 2004, ApJS, 153, 223
Croiset, B. A., Candian, A., Berné, O., & Tielens, A. G. G. M. 2016, A&A, 590, A26
Dale, D. A., Boquien, M., Barnes, A. T., et al. 2023, ApJL, 944, L23
Dale, D. A., Cook, D. O., Roussel, H., et al. 2017, ApJ, 837, 90
Dale, D. A., Gil de Paz, A., Gordon, K. D., et al. 2007, ApJ, 655, 863
Dawson, K. S., Schlegel, D. J., Ahn, C. P., et al. 2013, AJ, 145, 10
DESI Collaboration, Adame, A. G., Aguilar, J., et al. 2023, arXiv:2306.06308
DESI Collaboration, Aghamousa, A., Aguilar, J., et al. 2016, arXiv:1611.00036
Dewdney, P. E., Hall, P. J., Schilizzi, R. T., & Lazio, T. J. L. W. 2009, IEEEP, 97, 1482
Diamond-Stanic, A. M., & Rieke, G. H. 2010, ApJ, 724, 140
Donnelly, G. P., Smith, J. D. T., Draine, B. T., et al. 2024, ApJ, 965, 75
Dotter, A. 2016, ApJS, 222, 8
Draine, B. T. 2011, Physics of the Interstellar and Intergalactic Medium (Princeton, NJ: Princeton Univ. Press)
Draine, B. T., & Li, A. 2001, ApJ, 551, 807
Draine, B. T., Li, A., Hensley, B. S., et al. 2021, ApJ, 917, 3
Egorov, O. V., Kreckel, K., Sandstrom, K. M., et al. 2023, ApJL, 944, L16
Eirola, E., Doquire, G., Verleysen, M., & Lendasse, A. 2013, Inf. Sci., 240, 115
Eisenstein, D. J., Weinberg, D. H., Agol, E., et al. 2011, AJ, 142, 72
Eldridge, J. J., Stanway, E. R., Xiao, L., et al. 2017, PASA, 34, e058
Emsellem, E., Schinnerer, E., Santoro, F., et al. 2022, A&A, 659, A191
Euclid Collaboration, Scaramella, R., Amiaux, J., et al. 2022, A&A, 662, A112
Feigelson, E. D., & Nelson, P. I. 1985, ApJ, 293, 192
Ferguson, H. C., Dickinson, M., & Williams, R. 2000, ARA&A, 38, 667
Fluke, C. J., & Jacobs, C. 2020, WDMKD, 10, e1349
Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, A&A, 595, A1
Gaia Collaboration, Vallenari, A., Brown, A. G. A., et al. 2023, A&A, 674, A1
Galliano, F., Madden, S. C., Tielens, A. G. G. M., Peeters, E., & Jones, A. P. 2008, ApJ, 679, 310
Garnett, D. R. 1992, AJ, 103, 1330
Giavalisco, M., Ferguson, H. C., Koekemoer, A. M., et al. 2004, ApJL, 600, L93
Gordon, K. D., Engelbracht, C. W., Rieke, G. H., et al. 2008, ApJ, 682, 336
Hassani, H., Rosolowsky, E., Leroy, A. K., et al. 2023, ApJL, 944, L21
Henshaw, J. D., Barnes, A. T., Battersby, C., et al. 2023, in ASP Conf. Ser. 534, Protostars and Planets VII, ed. S. Inutsuka et al. (San Francisco, CA: ASP), 83
Hony, S., Van Kerckhoven, C., Peeters, E., et al. 2001, A&A, 370, 1030
Hruschka, E., Hruschka, E., & Ebecken, N. 2003, in AI 2003: Advances in Artificial Intelligence (New York: Springer), 723
Hunter, J. D. 2007, CSE, 9, 90
Isobe, T., Feigelson, E. D., & Nelson, P. I. 1986, ApJ, 306, 490
Jones, E., Oliphant, T., Peterson, P., et al. 2001, SciPy: Open source scientific tools for Python, http://www.scipy.org/
Jonsson, P., & Wohlin, C. 2004, in Proc. 10th Int. Software Metrics Symp. (Piscataway, NJ: IEEE), 108
Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, ApJ, 873, 111
Kaiser, N., Burgett, W., Chambers, K., et al. 2010, Proc. SPIE, 7733, 77330E
Kaneda, H., Onaka, T., & Sakon, I. 2005, ApJL, 632, L83
Kaneda, H., Onaka, T., Sakon, I., et al. 2008, ApJ, 684, 270
Kauffmann, G., Heckman, T. M., Tremonti, C., et al. 2003, MNRAS, 346, 1055
Kennicutt, R. C., Calzetti, D., Aniano, G., et al. 2011, PASP, 123, 1347
Kennicutt, R. C. J., Armus, L., Bendo, G., et al. 2003, PASP, 115, 928
Kewley, L. J., Dopita, M. A., Sutherland, R. S., Heisler, C. A., & Trevena, J. 2001, ApJ, 556, 121
Kewley, L. J., Groves, B., Kauffmann, G., & Heckman, T. 2006, MNRAS, 372, 961
Kewley, L. J., Nicholls, D. C., & Sutherland, R. S. 2019, ARA&A, 57, 511
Knight, C., Peeters, E., Stock, D. J., Vacca, W. D., & Tielens, A. G. G. M. 2021, ApJ, 918, 8
Kollmeier, J. A., Zasowski, G., Rix, H.-W., et al. 2017, arXiv:1711.03234
Kreckel, K., Ho, I. T., Blanc, G. A., et al. 2020, MNRAS, 499, 193
Kullback, S., & Leibler, R. A. 1951, Ann. Math. Statist., 22, 79
Lai, T. S. Y., Armus, L., Bianchin, M., et al. 2023, ApJL, 957, L26
Lai, T. S. Y., Armus, L., U, V., et al. 2022, ApJL, 941, L36
Lee, J. C., Sandstrom, K. M., Leroy, A. K., et al. 2023, ApJL, 944, L17
Lee, J. C., Whitmore, B. C., Thilker, D. A., et al. 2022, ApJS, 258, 10
Leroy, A. K., Hughes, A., Liu, D., et al. 2021b, ApJS, 255, 19
Leroy, A. K., Sandstrom, K., Rosolowsky, E., et al. 2023, ApJL, 944, L9
Leroy, A. K., Schinnerer, E., Hughes, A., et al. 2021a, ApJS, 257, 43
Li, A. 2020, NatAs, 4, 339
Little, J. A., & Rubin, D. B. 2002, Statistical Analysis with Missing Data (2nd ed.; New York: Interscience (Wiley-Interscience))
Majewski, S. R., Schiavon, R. P., Frinchaboy, P. M., et al. 2017, AJ, 154, 94
Maragkoudakis, A., Boersma, C., Temi, P., Bregman, J. D., & Allamandola, L. J. 2022, ApJ, 931, 38
Maragkoudakis, A., Peeters, E., & Ricca, A. 2020, MNRAS, 494, 642
McInnes, L., Healy, J., & Melville, J. 2018, arXiv:1802.03426
Morganti, R., Tsvetanov, Z. I., Gallimore, J., & Allen, M. G. 1999, A&AS, 137, 457
Moustakas, J., Kennicutt, R. C. J., Tremonti, C. A., et al. 2010, ApJS, 190, 233
Newman, D. 2014, Organ. Res. Methods, 17, 372
Niederhut, D. 2018, Proc. 17th Python in Science Conf. (SciPy 2018), ed. F. Akici et al. (Austin, TX: SciPy), 56
O'Halloran, B., Satyapal, S., & Dudik, R. P. 2006, ApJ, 641, 795
Osterbrock, D. E., & Ferland, G. J. 2006, Astrophysics of Gaseous Nebulae and Active Galactic Nuclei (Sausalito, CA: Univ. Science Books)
Packer, J. S., Zhu, Q., Huynh, C., et al. 2019, Sci, 365, eaax1971
Pasquini, S., Peeters, E., Schefter, B., et al. 2023, arXiv:2311.01163
Pathak, D., Leroy, A. K., Thompson, T. A., et al. 2024, AJ, 167, 39
Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, JMLR, 12, 2825
Peeters, E., Bauschlicher, C. W., Jr., & Allamandola, J. 2017, ApJ, 836, 198
Peeters, E., Habart, E., Berne, O., et al. 2023, arXiv:2310.08720
Pérez, F., & Granger, B. E. 2007, CSE, 9, 21
Pessa, I., Schinnerer, E., Sanchez-Blazquez, P., et al. 2023, A&A, 673, A147
Pilyugin, L. S., & Grebel, E. K. 2016, MNRAS, 457, 3678
Ramdas, A., Garcia, N., & Cuturi, M. 2015, arXiv:1509.02237
Reis, I., Baron, D., & Shahaf, S. 2019, AJ, 157, 16
Rich, J. A., Kewley, L. J., & Dopita, M. A. 2011, ApJ, 734, 87
Rich, J. A., Kewley, L. J., & Dopita, M. A. 2015, ApJS, 221, 28
Rigopoulou, D., Barale, M., Clary, D. C., et al. 2021, MNRAS, 504, 5287
Robitaille, T., Deil, C., & Ginsburg, A., 2020 reproject: Python-based astronomical image reprojection, Astrophysics Source Code Library, ascl:2011.023
Rubner, Y., Tomasi, C., & Guibas, L. J. 2000, Int. J. Comput. Vis., 40, 99
Saintonge, A., Catinella, B., Tacconi, L. J., et al. 2017, ApJS, 233, 22
Saintonge, A., Kauffmann, G., Kramer, C., et al. 2011, MNRAS, 415, 32
Sandstrom, K. M., Chastenet, J., Sutter, J., et al. 2023a, ApJL, 944, L7
Sandstrom, K. M., Koch, E. W., Leroy, A. K., et al. 2023b, ApJL, 944, L8
Schafer, J. L., & Graham, J. W. 2002, Psychol. Methods, 7, 147
Schinnerer, E., Emsellem, E., Henshaw, J. D., et al. 2023, ApJL, 944, L15
Scoville, N., Aussel, H., Brusa, M., et al. 2007, ApJS, 172, 1
Shah, A. D., Bartlett, J. W., Carpenter, J., Nicholas, O., & Hemingway, H. 2014, Am. J. Epidemiol., 179, 764
Shakura, N. I., & Sunyaev, R. A. 1973, A&A, 24, 337
Slone, O., & Netzer, H. 2012, MNRAS, 426, 656
Smith, J. D. T., Draine, B. T., Dale, D. A., et al. 2007, ApJ, 656, 770
Stekhoven, D. J., & Bühlmann, P. 2011, Bioinformatics, 28, 112
Sun, J., Leroy, A. K., Rosolowsky, E., et al. 2022, AJ, 164, 43
Tielens, A. G. G. M. 2005, The Physics and Chemistry of the Interstellar Medium (Cambridge: Cambridge Univ. Press)
Tielens, A. G. G. M. 2008, ARA&A, 46, 289
Ujjwal, K., Kartha, S. S., Krishna, R. A., et al. 2024, A&A, 684, 8
van Buuren, S., & Groothuis-Oudshoorn, K. 2011, J. Stat. Softw., 45, 1
Vazdekis, A., Sánchez-Blázquez, P., Falcón-Barroso, J., et al. 2010, MNRAS, 404, 1639
Vega, O., Bressan, A., Panuzzo, P., et al. 2010, ApJ, 721, 1090
Veilleux, S., & Osterbrock, D. E. 1987, ApJS, 63, 295
Walter, F., Brinks, E., de Blok, W. J. G., et al. 2008, AJ, 136, 2563
Williams, T. G., Kreckel, K., Belfiore, F., et al. 2022, MNRAS, 509, 1303
Williams, T. G., Lee, J. C., Larson, K. L., et al. 2024, arXiv:2401.15142
Xu, D., & Tian, Y. 2015, AnDS, 2, 165
York, D. G., Adelman, J., Anderson, J. E., Jr, et al. 2000, AJ, 120, 1579