

A geostatistical analysis of multiscale metallicity variations in galaxies – II. Predicting the metallicities of H II and diffuse ionized gas regions via universal kriging

Benjamin Metha¹,^{1,2}★ Michele Trenti,^{1,2} Tingjin Chu³ and Andrew Battisti^{1,2,4}

¹*School of Physics, The University of Melbourne, VIC 3010, Australia*

²*Australian Research Council Centre of Excellence for All-Sky Astrophysics in 3-Dimensions, Canberra, ACT 2611, Australia*

³*School of Mathematics and Statistics, The University of Melbourne, VIC 3010, Australia*

⁴*Research School of Astronomy and Astrophysics, Australian National University, Cotter Road, Weston Creek, ACT 2611, Australia*

Accepted 2022 May 25. Received 2022 May 25; in original form 2022 March 1

ABSTRACT

The metallicity of diffuse ionized gas (DIG) cannot be determined using strong emission line diagnostics, which are calibrated to calculate the metallicity of H II regions. Because of this, resolved metallicity maps from integral field spectroscopy (IFS) data remain largely incomplete. In this paper (the second of a series), we introduce the geostatistical technique of universal kriging, which allows the complete 2D metallicity distribution of a galaxy to be reconstructed from metallicities measured at H II regions, accounting for spatial correlations between nearby data points. We apply this method to construct high-fidelity metallicity maps of the local spiral galaxy NGC 5236 using data from the TYPHOON/PrISM survey. We find significant correlation in the metallicity of H II regions separated by up to 0.4–1.2 kpc. Predictions constructed using this method were tested using cross-validation in H II regions, and we show that they outperform significantly interpolation based on metallicity gradients. Furthermore, we apply kriging to predict the metallicities in regions dominated by DIG emission, considering seven additional spiral galaxies with high resolution ($\lesssim 100$ pc) metallicity maps. We compare kriging maps to DIG metallicities computed with novel ionization corrections, and find that such corrections introduce a systematic offset of up to ± 0.1 dex for any individual galaxy, with a scatter of 0.02–0.07 dex for the sample. Overall we recommend universal kriging, together with a calibrated geostatistical model, as the superior method for inferring the metallicities of DIG-dominated regions in local spiral galaxies, demonstrating further the potential of applying geostatistical methods to spatially resolved galaxy observations.

Key words: methods: statistical – ISM: abundances – galaxies: abundances – galaxies: ISM.

1 INTRODUCTION

Recent advances in integral field spectroscopy (IFS) have produced datacubes containing an unprecedented amount of information about nearby galaxies, revealing how stellar masses, ages, metallicities, and luminosities, gas metallicities and kinematics, and dust extinction vary spatially. These maps offer increasingly high resolution, but are also potentially affected by missing data and/or regions where the signal to noise ratio is degraded. In such big and complex data frameworks, traditional astronomical analysis techniques are not necessarily suited to fully extract the wealth of information that these data products contain.

One particularly information-rich data-product that can be extracted from IFS surveys is the distribution of metals (elements heavier than helium) throughout the interstellar medium (ISM) of a galaxy. How the metallicity varies spatially throughout a galaxy is a product of many competing factors, including (i) continuous metal enrichment throughout the star formation history of a galaxy (e.g. Edmunds & Greenhow 1995; Freeman & Bland-Hawthorn 2002),

(ii) turbulent mixing driven by thermal or gravitational instabilities (de Avillez & Mac Low 2002; Scalo & Elmegreen 2004; Pan & Scannapieco 2010; Krumholz & Ting 2018) or large structures such as spiral arms or bars (Di Matteo et al. 2013; Grand et al. 2016; Ho et al. 2017), (iii) galactic outflows caused by supernovae or galactic winds (Heckman, Armus & Miley 1990; Hopkins et al. 2014; Christensen et al. 2018), and (iv) inflows of pristine gas (Sánchez Almeida et al. 2014) or previously ejected materials (galactic fountains, Kim et al. 2020). Understanding the relative strengths, spatial scales, and time-scales of these processes is a major challenge in modelling galaxy formation and evolution (Naab & Ostriker 2017).

Geostatistics is a subfield of spatial statistics that focuses on understanding stochastic processes that occur over a continuous spatial domain. As both star formation and the turbulent processes that mix the products of star formation into the surrounding ISM are stochastic processes, and galaxies in the local Universe are viewed as continuous extended objects, this branch of mathematics can be naturally applied to the field of extragalactic astronomy, offering a novel perspective on galaxy evolution.

Already, these techniques have been successfully applied to improve traditional data analysis recipes and extract interesting

* E-mail: methab@student.unimelb.edu.au

information from IFS data. González-Gaitán et al. (2019) showed that geostatistical methods hold the potential to offer improvements over binning techniques (such as the popular Voronoi tessellation, used by e.g. Cappellari & Copin 2003) for accurately predicting the age and metallicity of stellar populations in IFS data without sacrificing spatial resolution. In order to capture non-symmetric deviations around a mean metallicity gradient, Clark et al. (2019) fit a geostatistical model using Gaussian process regressions to accurately map the dust mass-to-light ratio throughout two nearby galaxies. Widely used in machine learning algorithms (Rasmussen & Williams 2006), these Gaussian process regressions attempt to reconstruct a random field by creating a model of the covariance between nearby data points, allowing for complex structures to be captured with only a few parameters. These techniques were further employed by Williams et al. (2022), who found that 12 out of 19 galaxies observed as a part of the PHANGS-MUSE survey showed significant 2D metallicity variations that could not be captured by a simple metallicity gradient model. Metha, Trenti & Chu (2021) (hereafter M21) used a semivariogram¹ analysis to distinguish these small-scale metallicity fluctuations from measurement error, allowing the predictions of an analytical metal-mixing model to be directly compared to IFS data.

Kriging refers to a family of statistically optimal, unbiased algorithms that can be used to fit a model of a Gaussian process to data while minimizing prediction variance (Cressie 1990). Since their inception in the late 1960s (Huijbregts & Matheron 1971), these methods have formed a cornerstone of the geostatistical paradigm, seeing applications in a variety of fields, from environmental science and meteorology to economics and population statistics (Wikle, Zammit-Mangion & Cressie 2019).

In this study (the second in a series), we continue the exploration on how geostatistical techniques may be used to improve the analysis of high-resolution IFS data, which started in M21. First, we construct a geostatistical hierarchical model of the multiscale metallicity structure of the warm ISM for NGC 5236, using mock-IFS data from the TYPHOON/PrISM survey. Then, using the method of universal kriging, we derive robust predictions, with uncertainty, about the metallicity in regions dominated by diffuse ionized gas (DIG), where the metallicity cannot be measured directly. Finally, we compare these DIG metallicity predictions to metallicities determined using the empirical correction factors of Kumari et al. (2019), in order to offer an independent test of these new diagnostics.

This paper is organized as follows. In Section 2, we review the main challenges in estimating DIG metallicity, before describing the TYPHOON survey in Section 3. In Section 4, we introduce hierarchical geostatistical models and our model fitting methods before presenting the algorithm used for universal kriging. In Section 5.1, we discuss in detail the results of our geostatistical model fitting procedure for the local spiral galaxy NGC 5236 (also known as M83), and confirm using 10-fold cross validation that our universal kriging method gives robust predictions for the metallicity of H II regions in Section 5.2. After validating our methodology, we extend this analysis to DIG-dominated regions in Section 6, providing an independent test of the DIG-corrected metallicity diagnostics of Kumari et al. (2019). Strengths, weaknesses, and future applications of the universal kriging approach are discussed in Section 7, and our main results are summarized in Section 8.

¹A *semivariogram* is a geostatistical function that shows how the variance between data points in a sample depends on their spatial separation.

2 DIFFUSE IONIZED GAS METALLICITY DETERMINATION

There are several methods to measure metallicities within a galaxy's ISM (see e.g. Kewley, Nicholls & Sutherland 2019; Maiolino & Mannucci 2019 for comprehensive reviews). The abundance of an element can be measured directly from emission-line spectroscopy, if the temperature and density of the gas it resides inside is known (Aller 1984). However, the auroral lines required for accurate temperature measurements are often too faint to be seen in many systems of interest, as well as in IFS data (e.g. Andrews & Martini 2013). In these cases, strong emission-line calibrations from photoionization models (e.g. Kewley & Dopita 2002; Dopita et al. 2016), or empirical diagnostics (e.g. Pilyugin & Thuan 2005; Pilyugin & Grebel 2016) may also be used.

One limitation of all of these methods is that they are only valid for determining metallicities within H II regions, where the ISM has been completely ionized by local young, bright O/B stars. Narrow-band imaging of local galaxies targeting the H α line shows that approximately half of the H α flux emitted from the ISM originates from a thick disc of low density plasma that extends beyond the galactic mid-plane known as diffuse ionized gas (DIG; Zurita, Rozas & Beckman 2000; Oey et al. 2007). Emission-line spectra of the DIG are complex, revealing that it is ~ 2000 K hotter than H II regions, ionized by a harder spectrum (Hoopes & Walterbos 2003; Haffner et al. 2009), and contains enhanced forbidden line ratios for [O II]/H α and [N II]/H α . Many different sources for the ionization of the DIG have been proposed, including radiation escaping from H II regions in the discs of galaxies (Zurita et al. 2000, 2002), radiation from evolved, low-mass stars (Stasińska et al. 2008), supernova-driven shocks (Dopita & Sutherland 1995), excitation within mixing layers between regions of hot and cool gas (Binette et al. 2009), or any combination of these phenomena (Sánchez 2020; Mannucci et al. 2021).

Despite the limited understanding of the ionization sources within the DIG, it may still be possible to determine the metallicity within this gas using empirical means. Using data from the MUSE Atlas of Disks (MAD) survey (Erroz-Ferrer et al. 2019), Kumari et al. (2019) measured the difference in strong emission line ratios between H II regions and nearby DIG-dominated spaxels. Under the assumption that the ISM is homogeneous on spatial scales smaller than 0.5 kpc, Kumari et al. (2019) searched for systematic offsets in the emission line ratios from pairs of H II and DIG-dominated spaxels and found that these offsets were correlated with the [O III]/H β line ratio. Using this observation, two corrections to popular emission-line diagnostics of Curti et al. (2017) were constructed which may be used to more accurately infer the metallicity of DIG-dominated regions.

While the results of this approach are promising, there are two potential caveats that should be taken into account. First, a recent geostatistical analysis revealed that the ISM of local spiral galaxies contain significant metallicity fluctuations on scales smaller than ~ 1 kpc (M21), inconsistent with the assumption that the ISM is homogeneous on spatial scales smaller than 0.5 kpc (although see Kreckel et al. 2020 and Williams et al. 2022 for different conclusions on the characteristic mixing scale). Secondly, each of the aforementioned sources that ionize the DIG have their own spectral properties, and the relative contribution of these sources is not expected to be the same in different galaxies. For this reason, it is unlikely that a single emission-line based correction factor exists that will be able to correct the DIG contamination stemming from all of these different components in all galaxies.

To overcome these issues we present here an alternate opportunity. The approach to estimating the metallicity of DIG-dominated regions

Table 1. Astrometric and morphological properties of the eight galaxies investigated in this study. Ra, Dec. are taken from the NASA/IPAC Extragalactic Database. Morphological classification, position angles, and inclination data are taken from HyperLEDA. Effective radii are taken to be the half-light radii in the B band as determined by the Carnegie-Irvine Galaxy Survey (Ho et al. 2011). References for distances: (a) Tully et al. (2008); (b) Tully, Courtois & Sorce (2016); (c) Tully et al. (2013); (d) Tully & Fisher (1988).

Name	Type	RA	Dec.	PA	i	D (Mpc)	R_e (arcmin)
NGC 1068	Sb	02:42:40.71	−00:00:47.8	72.7	34.7	12.3 ^a	0.61
NGC 1365	Sb	03:33:36.37	−36:08:25.5	23.4	62.6	16.98 ^b	3.05
NGC 1566	SABb	04:20:00.42	−54:56:16.1	44.2	49.1	6.61 ^c	1.30
NGC 2835	Sc	09:17:52.91	−22:21:16.8	1.3	56.2	8.75 ^b	1.51
NGC 2997	SABc	09:45:38.79	−31:11:27.9	98.9	53.7	11.3 ^a	1.68
NGC 5068	Sc	13:18:54.81	−21:02:20.8	152	27.3	6.70 ^d	1.75
NGC 5236	Sc	13:39:55.96	−29:51:55.5	54.0	15.3	4.66 ^c	2.93
NGC 7793	Scd	23:57:49.83	−32:35:27.7	94.4	63.5	3.60 ^c	2.21

presented in this study makes no assumptions about the ionization sources of the DIG, depending instead on the spatial locations of the DIG/H II spaxels, and a model of metal mixing throughout galaxies. Under the assumption that the processes that govern turbulence and gas-mixing are constant throughout all parts of the ISM that are in the same phase, the predictions made by a geostatistical model trained on H II region data will be able to produce accurate predictions about the metallicity of all components of the warm interstellar medium.

3 DATA

In this paper, we use data from the TYPHOON survey (Seibert et al., in preparation). Below we briefly summarize the process used to generate these data products – for further details, see Poetrodjojo et al. (2019).

Based on observations with the 2.5 m du Pont Telescope at Las Campanas Observatory, datacubes were constructed for a sample of large, nearby spiral galaxies with the Progressive Integral Step Method (PrISM). The process builds a 3D datacube (two spatial and one spectral dimensions) through a series of long slit observations for each galaxy in a scanning mode, i.e. shifting each time the long-slit orthogonally to its extension by a constant angular offset. Specifically, a long slit of size 18 arcmin \times 1.65 arcsec was used, with native angular resolution along the slit of 0.484 arcsec (rebinned to 1.65 arcsec to construct the datacubes), and spectral resolution $R \approx 800$ over the wavelength range 3650–8150 Å. For each galaxy considered in this paper, 200 long-slit observations have been acquired, creating a datacube covering approximately 650 \times 200 pixels.

Targets were chosen from parent sample of 11HUGS and Local Volume Legacy Survey galaxies (Dale et al. 2009). Galaxies were selected based on their declination ($\leq +10^\circ$), angular diameter (< 18 arcmin, chosen to correspond to the length of the slit), inclination (face-on), and surface brightness (brighter than 21.1 AB mag/arcsec² in the B band).

This work focuses on the local bright galaxy NGC 5236, also known as M83. Lying at a distance of 4.66 Mpc (Tully et al. 2013) and with a stellar mass of $10^{10.55} M_\odot$ (Bresolin et al. 2016) and a star formation rate of $4.2 M_\odot \text{ yr}^{-1}$ (Leroy et al. 2019), it is one of the closest face-on grand-design spiral galaxies, and is therefore a popular subject for studies on internal metallicity structure (e.g. Bresolin & Kennicutt 2002; Bresolin et al. 2009, 2016; Boettcher, Gallagher & Zweibel 2017; Poetrodjojo et al. 2019).

To ensure our results are valid for a variety of local spiral galaxies, we supplement our study with a sample of seven other galaxies from the TYPHOON/PrISM survey for which the emission line

maps have been constructed at time of writing, and the number of identifiable, resolvable H II regions is > 300 . Summary properties of these galaxies are listed in Table 1, and results for these galaxies are described briefly in Appendix A. The median resolution of each spaxel in this supplementary sample is 68.4 pc, which is more than sufficient for the analysis of small-scale metallicity fluctuations using the geostatistical framework introduced in M21.

For NGC 5236 (and the other seven supplementary galaxies), the strengths of eight different emission lines and their errors were computed for each spaxel using LZIFU. We imposed several data quality cuts to ensure that the signal-to-noise ratio, S/N, is greater than 3 for all relevant lines, except [O II] $\lambda\lambda 3627, 29$, for which we only imposed that $S/N > 1$.² Dust extinction was corrected using the model of Cardelli, Clayton & Mathis (1989), before DIG-dominated regions were separated from H II regions using three different diagnostics detailed in B2, yielding three maps of H II regions and DIG-dominated regions with high-quality spectral data per galaxy. For each of these maps, the metallicity was determined for each spaxel using four diagnostics: $N_2S_2H\alpha$ (Dopita et al. 2016), N_2O_2 (Dopita et al. 2013), O_3N_2 (Curti et al. 2017), and RS_{32} (Curti et al. 2020a). Details on the strengths, weaknesses, and calibration process behind each of these diagnostics are given in Appendix B1.

4 GEOSTATISTICAL METHODS

In M21, a framework for geostatistical hierarchical modelling was introduced. Furthermore, the semivariogram was discussed as a tool for geostatistical analysis. In this paper, we extend on these ideas, and introduce the concept of universal kriging as a technique that can be used to interpolate metallicities in DIG-dominated spaxels from H II region metallicity maps.

We briefly review the M21 hierarchical geostatistical metallicity model here. The observed metallicity Z_{obs} measured at each location \vec{x} is taken to be a combination of the true metallicity at that location, $Z(\vec{x})$, plus some measurement error $\epsilon(\vec{x})$:

$$Z_{\text{obs}}(\vec{x}) = Z(\vec{x}) + \epsilon(\vec{x}). \quad (1)$$

The measurement error is modelled as a stationary Gaussian process with zero mean: $E(\epsilon(\vec{x})) = 0$. It can be completely described by the covariance between any two pairs of data points, $\text{Cov}(\epsilon(\vec{x}), \epsilon(\vec{y}))$, which in turn depends on the uncertainty of each

²At the wavelength of this emission line, the du Pont Telescope is not very sensitive. Imposing an S/N cut of 3 for this line was found to exclude practically all of the data.

metallicity measurement, and how spatially correlated measurement errors are expected to be. The measurement error of each emission line for each spaxel is computed for the TYPHOON/PrISM survey using the `id1` package LZIFU.³

For CCD data, the dominant sources of noise come from dark current, read noise, and shot noise arising from Poisson processes associated with the generation of the detected photons. All these sources will generate white noise that is not spatially correlated between CCD pixels. Furthermore, given the atmospheric conditions at Las Campanas Observatory, the median seeing is 0.6 arcsec (Persson, Carr & Jacobs 1990), which is several times smaller than the size of each spaxel (1.65 arcsec). Therefore, the signal itself is also uncorrelated between spaxels. Because of this, we are justified in modelling all errors in TYPHOON data as being uncorrelated between spaxels.

The true metallicity distribution $Z(\vec{x})$ is modelled as a non-random process mean $\mu(\vec{x})$, and a spatially varying random component $\eta(\vec{x})$ with zero mean:

$$Z(\vec{x}) = \mu(\vec{x}) + \eta(\vec{x}). \quad (2)$$

In this paper, as in M21, $\mu(\vec{x})$ is treated as a basic metallicity gradient – that is,

$$\mu(\vec{x}) = Z_c + \langle \nabla Z \rangle \cdot r(\vec{x}), \quad (3)$$

where $r(\vec{x})$ is the distance between a galaxy’s centre and a given location \vec{x} .

The purpose of the function $\eta(\vec{x})$ is to capture the idea that (i) the true metallicities of spaxels are not expected to lie exactly on the mean radial metallicity trend, and (ii) nearby spaxels should have metallicity values that positively correlate with each other. In M21, $\eta(\vec{x})$ was assumed to take a form such that its covariance matrix matched the five-parameter model developed by Krumholz & Ting (2018). In this paper, we use a simpler, more general model for $\eta(\vec{x})$. We model the covariance of the metallicity between two data points separated by a distance of $h = |\vec{x} - \vec{y}|$ using an exponential function:

$$\text{Cov}(Z(\vec{x}), Z(\vec{y})) = \sigma^2 \exp\left(-\frac{h}{\phi}\right). \quad (4)$$

This equation has two free parameters: σ^2 , which describes the intrinsic variance within the data caused by random fluctuations around a mean trend (not to be confused with measurement error); and ϕ , which sets the range over which local variations in metallicity remain correlated. Such a model was chosen for two reasons: first, it is simple, with only two parameters, both of which have physical interpretations in the context of gas-phase physics; yet it remains general enough to capture the overall behaviour of the local metallicity fluctuations. Second, this function is well-studied in the geostatistical literature. It is a special case of the Matérn family (Matérn 1960), and is known to be positive definite, which is a necessary condition to ensure the covariance matrix between any set of observed data points has no negative entries. We encourage other

researchers familiar with this field to construct their own covariance functions using models of ISM turbulence.⁴

The semivariogram, among other applications, can be applied to distinguish the small-scale metallicity fluctuations, η , from measurement error, ϵ , as discussed in M21. Formally, the semivariogram $\gamma(h)$ is defined to be half the variance between data points separated by a distance of h . When the covariance function depends only on the distance between data points (that is, there exists a $C(r)$ such that if $\|\vec{x} - \vec{y}\| = r$ then $\text{Cov}(Z(\vec{x}), Z(\vec{y})) = C(r)$), the semivariogram can be calculated using the formula $\gamma(h) = C(0) - C(h)$. Because of this, equation (4) induces the following form for the semivariogram:

$$\gamma(h) = \sigma^2 (1 - \exp(-h/\phi)). \quad (5)$$

For spatially correlated data for which there is no spatially correlated measurement error, $\gamma(0) = 0$. As h increases, $\gamma(h)$ increases, until the data are so separated that they can be considered uncorrelated, at which point $\gamma(h)$ approaches the intrinsic variance of the correlated data. For this model, σ^2 sets the height of the semivariogram, and ϕ determines the physical scale h at which it flattens out. Additional uncorrelated noise (e.g. from measurement error) adds a constant factor to the measured semivariogram at all separations, and can thus be identified by computing $\gamma(0)$. Further details on the semivariogram and its role in geostatistical data analysis can be found in Wikle et al. (2019).

4.1 Fitting parameters for a geostatistical hierarchical model

Overall, our model for the 2D internal metallicity structure of the ISM of local galaxies has four parameters; two of which describe the global trend of metallicity with radius (Z_c , $\langle \nabla Z \rangle$), and two of which describe the small-scale metallicity fluctuations associated with stochastic star formation and diffusion of metals through the ISM (σ^2 , ϕ).

In M21, Z_c and $\langle \nabla Z \rangle$ were fit using a weighted least-squares approach, taking into account only the measurement error $\epsilon(\vec{x})$ of each spaxel, and not accounting for the small-scale deviations of the metallicity around the median value, $\eta(\vec{x})$. In hindsight, a semivariogram analysis showed that these local spiral galaxies are far from well-mixed, and the magnitude of these small-scale fluctuations was of the same order as the size of the measurement errors. Therefore, in order to properly model the metallicity gradient of these galaxies, the size and structure of small-scale metallicity fluctuations must be accounted for.

Such an approach is conceptually similar to the framework employed by Clark et al. (2019) and Williams et al. (2022) (described in e.g. Hogg, Bovy & Lang 2010) to fit the mean radial trend of their geostatistical models. This approach employs an additional parameter to describe the intrinsic scatter around the linear trend, and fits this parameter together with the linear fit using a maximum-likelihood method. However, this method only accounts for the additional variance caused by small-scale fluctuations around the mean trend, without accounting for correlations between nearby data points.

In this paper, we demonstrate an improved, statistically robust approach, wherein all four parameters required to build a geostatistical model for a galaxy are fit simultaneously using maximum-likelihood estimation (ML), following the procedure described in Diggle & Riberio (2007). By fitting all of these parameters at the same time, the

³LZIFU is an IFS data fitting tool that uses a penalized pixel-fitting method (pPXF; Cappellari & Emsellem 2004; Cappellari 2017) to model stellar continuum and Gaussian components using the Levenberg–Marquardt least-squares method to fit emission lines. The stellar continuum is modelled using the MIUSCAT simple stellar population models (Vazdekis et al. 2012) with 13 ages and 3 metallicities. For further details, see Ho et al. (2016).

⁴For a list of known positive-definite functions that can be applied in order to model spatial covariance, see e.g. Gneiting (1997).

uncertainty in their estimated values is reduced (Mardia & Marshall 1984). We detail this method below.

Let \mathbf{Z}_{obs} be a vector of N observed metallicity values over a spatial domain, and let \mathbf{C}_ϵ be the $N \times N$ covariance matrix that completely describes the structure of correlated measurement errors between data points. For the case of TYPHOON galaxies, all measurement errors are modelled as uncorrelated and so \mathbf{C}_ϵ is a diagonal matrix, with elements in the diagonal equal to the variance in each metallicity measurement. These metallicity measurement uncertainties are computed from the uncertainty in emission line flux using linear error propagation [described in Section B1; see equation (B1)], and are fixed for each galaxy.

Let \mathbf{D} be the $N \times 2$ matrix with every element in the first column equal to 1 and the second column given by the deprojected distance from each observation point to the galaxy centre. Let $\boldsymbol{\beta} = [Z_c, \langle \nabla Z \rangle]^T$ – where T is the transpose operator – so that $\mathbf{D}\boldsymbol{\beta} = \boldsymbol{\mu}$, the metallicity of each spaxel predicted by the linear metallicity gradient model.⁵ Let $\mathbf{C}_Z(\sigma^2, \phi)$ be the covariance matrix associated with local metallicity fluctuations $\eta(\vec{x})$. Because the deprojected distance between each pair of spaxels is known, the value of each entry in this matrix can be calculated using equation (4).

Using our geostatistical framework, we model the vector \mathbf{Z}_{obs} as being drawn from a N -dimensional Gaussian distribution, with mean given by $\boldsymbol{\mu} = \mathbf{D}\boldsymbol{\beta}$, and total covariance matrix given by $\mathbf{C}_\epsilon + \mathbf{C}_Z(\sigma^2, \phi)$. The natural logarithm of the likelihood of drawing \mathbf{Z}_{obs} from this distribution is then an analytical function of our four parameters:

$$\begin{aligned} \mathcal{L}(\boldsymbol{\beta}, \sigma^2, \phi) = & -\frac{1}{2} \left[N \log(2\pi) + \log(|\mathbf{C}_\epsilon + \mathbf{C}_Z(\sigma^2, \phi)|) \right. \\ & + (\mathbf{Z}_{\text{obs}} - \mathbf{D}\boldsymbol{\beta})^T (\mathbf{C}_\epsilon + \mathbf{C}_Z(\sigma^2, \phi))^{-1} \\ & \left. \times (\mathbf{Z}_{\text{obs}} - \mathbf{D}\boldsymbol{\beta}) \right]. \end{aligned} \quad (6)$$

This function can be maximize using many numerical methods. To simplify the problem and speed up computation, we use a profile log-likelihood approach. For any specified values of σ^2 and ϕ , the most likely values for $\boldsymbol{\beta}$ are given by the following linear equation:

$$\hat{\boldsymbol{\beta}}(\sigma^2, \phi) = (\mathbf{D}^T \mathbf{V}(\sigma^2, \phi)^{-1} \mathbf{D})^{-1} \mathbf{D}^T \mathbf{V}(\sigma^2, \phi)^{-1} \mathbf{Z}_{\text{obs}}, \quad (7)$$

where $\mathbf{V}(\sigma^2, \phi) = \mathbf{C}_\epsilon + \mathbf{C}_Z(\sigma^2, \phi)$. This equation can be substituted back into equation (6), in order to give a function of σ^2 and ϕ that, when maximize, gives their maximum likelihood estimates. This approach effectively reduces the complexity of this task from a 4D optimization problem into a 2D one.

The uncertainty in the best-fitting parameters for a kriging model can be estimated in many ways, including using robust empirical methods such as bootstrapping, or by taking a Bayesian approach and analysing the posterior distribution of the parameters. We estimate the uncertainty of σ^2 and ϕ by computing the inverse of the Hessian matrix, which for exponentially correlated data provides a good approximation of the covariance matrix for these parameters (Mardia & Marshall 1984). To find the uncertainty in Z_c and $\langle \nabla Z \rangle$, we use the formula for the covariance matrix for a generalized

least-squares fit, using the maximum likelihood estimates of σ^2 and ϕ :

$$\text{Cov}(\hat{\boldsymbol{\beta}}) = (\mathbf{D}^T \mathbf{V}(\sigma^2, \phi)^{-1} \mathbf{D})^{-1}. \quad (8)$$

4.2 Universal Kriging

When the structure of spatial correlations is known for a process, measured data points can be used to predict nearby unknown values. *Universal kriging* is a method by which the true metallicity can be estimated at any point within a galaxy, accounting for large-scale trends using a Gaussian additive model fit via a generalized least-squares regression (GLS), and modelling small-scale variation using a known covariance function (Matheron 1969; Huijbrechts & Matheron 1971).

To define universal kriging, we follow Wikle et al. (2019). Using the same notation as in Section 4.1, our observations \mathbf{Z}_{obs} , their errors \mathbf{C}_ϵ , and the best-fitting geostatistical model may be used to estimate Z at any unknown data point, \vec{x}_0 . Let \mathbf{c}_0 be the vector with i th element given by $\text{Cov}(Z(\vec{x}_0), Z(\vec{x}_i))$ – that is, \mathbf{c}_0 gives the modelled covariance between this vector and each observed data point. Finally, let \mathbf{d}_0 be the valuation of each covariate function at this point – for the simple metallicity gradient model, $\mathbf{d}_0 = [1, r(\vec{x}_0)]^T$. Then we may estimate the value of Z at this point and the uncertainty of this estimate using the following pair of equations:

$$\hat{Z}(\vec{x}_0) = \mathbf{d}_0 \boldsymbol{\beta} + \mathbf{c}_0^T \mathbf{V}(\sigma^2, \phi)^{-1} (\mathbf{Z} - \mathbf{D}\boldsymbol{\beta}), \quad (9)$$

$$\sigma_{Z,uk}^2 = \sigma^2 - \mathbf{c}_0^T \mathbf{V}(\sigma^2, \phi)^{-1} \mathbf{c}_0 + k. \quad (10)$$

Here, $\mathbf{c}_0^T \mathbf{V}(\sigma^2, \phi)^{-1} \mathbf{c}_0$ represents a factor by which the variance associated with this data point is reduced after taking into account its correlation with known (but uncertain) data points, and k is the additional uncertainty that comes about from estimating $\boldsymbol{\beta}$ values using the GLS method:

$$\begin{aligned} k = & (\mathbf{d}_0 - \mathbf{D}^T \mathbf{V}(\sigma^2, \phi)^{-1} \mathbf{c}_0)^T (\mathbf{D}^T \mathbf{V}(\sigma^2, \phi)^{-1} \mathbf{D})^{-1} \\ & \times (\mathbf{d}_0 - \mathbf{D}^T \mathbf{V}(\sigma^2, \phi)^{-1} \mathbf{c}_0). \end{aligned} \quad (11)$$

Because this mathematics is powered by linear algebra, it is simple to extend these definitions to work for multiple points at a time (see e.g. Cressie 1993). This allows metallicity estimates to be computed rapidly for a grid of data points, leading to highly detailed maps of the predicted internal metallicity structure of galaxies, or for a list of data points where the metallicity cannot be measured directly, such as DIG-dominated spaxels. In this paper, we will show examples of both of these constructions.

5 RESULTS

5.1 The geostatistical model for H II regions

In this section, we present the results of geostatistical model fitting for NGC 5236. Using the [S II]/H α diagnostic of Kaplan et al. (2016) described in Section B2.1, 3664 H II spaxels and 19607 spaxels with significant DIG contamination were identified. The number of spaxels with metallicities determined after imposing S/N cuts for the emission lines used in each diagnostic are listed in Table 2. Using the H II regions, the four free parameters of the geostatistical model described in Section 4 were fit, using the ML method outlined in Section 4.1.

In Fig. 1, we show the most likely large-scale metallicity trend, $\mu(\vec{x})$, according to this model, for NGC 5236. All H II spaxels with metallicities determined for each diagnostic are shown, with error

⁵More generally, \mathbf{D} may be a matrix containing the values of any number of covariates for each data point. Each covariate is a variable that is expected to influence a galaxy's local metallicity, such as the local star formation rate of each spaxel, or whether or not each spaxel is contained within a spiral arm. In this general case, $\boldsymbol{\beta}$ would be a vector containing the parameters describing the global trend, such that $\mathbf{D}\boldsymbol{\beta}$ is still the predicted mean metallicity for all data points, $\boldsymbol{\mu}$.

Table 2. Number of H II spaxels and spaxels with significant DIG contamination in NGC 5236 for which metallicities have been determined with each diagnostic. Here, DIG-dominated regions are separated from H II regions using the methodology of Kaplan et al. (2016) described in Appendix B2.1. Further details on the metallicity diagnostics presented here are provided in Appendix B1.

Diagnostic	# of H II regions	# of DIG regions
$N_2S_2H\alpha$ (Dopita et al. 2016)	3664	19 602
N_2O_2 (Dopita et al. 2013)	623	629
O_3N_2 (Curti et al. 2020a)	2459	6625
RS_{32} (Curti et al. 2020a)	2482	6628

bars reflecting the uncertainty in the metallicity obtained by linear error propagation. For O_3N_2 , RS_{32} , and $N_2S_2H\alpha$, the error associated with each spaxel is small, with median values of 0.014, 0.013, and 0.025 dex, respectively. Conversely, for N_2O_2 the median error associated with each spaxel is 0.058 dex. This large uncertainty stems from the lack of sensitivity of the Wide Field CCD instrument on the

du Pont telescope at the wavelength of the $[O II]\lambda\lambda 3726, 29$ emission line, and our subsequent decision to use a less stringent S/N cut for this line.

The dearth in metallicities computed with the $N_2S_2H\alpha$ diagnostic between $\log(O/H) + 12 = 9.05$ and 9.1 stems from the fact that this diagnostic contains a correction term, for metallicities greater than 9.05. This introduces a discontinuity into the metallicity diagnostic that can be seen in Fig. 1. The uncertainties due to this correction factor are completely accounted for in this analysis, as described in Appendix B1.

In M21, local star-forming galaxies were found to have significant variations in metallicity on small scales due to inefficient mixing. The geostatistical method we present in this work accounts for these small-scale metallicity fluctuations when fitting the global trend in a self-consistent way, by simultaneously modelling the local covariance structure and using maximum-likelihood estimation. In Fig. 1, we compare the metallicity gradient models obtained using this method to the metallicity gradients one would obtain using a simple weighted least-squares (WLS) approach. The WLS method assumes (i) that there is no correlated variance between local data points beyond what is described by the metallicity gradient, (ii) that all data points should, in reality, lie exactly on the line-of-best-fit found by this method, and (iii) that any deviations in

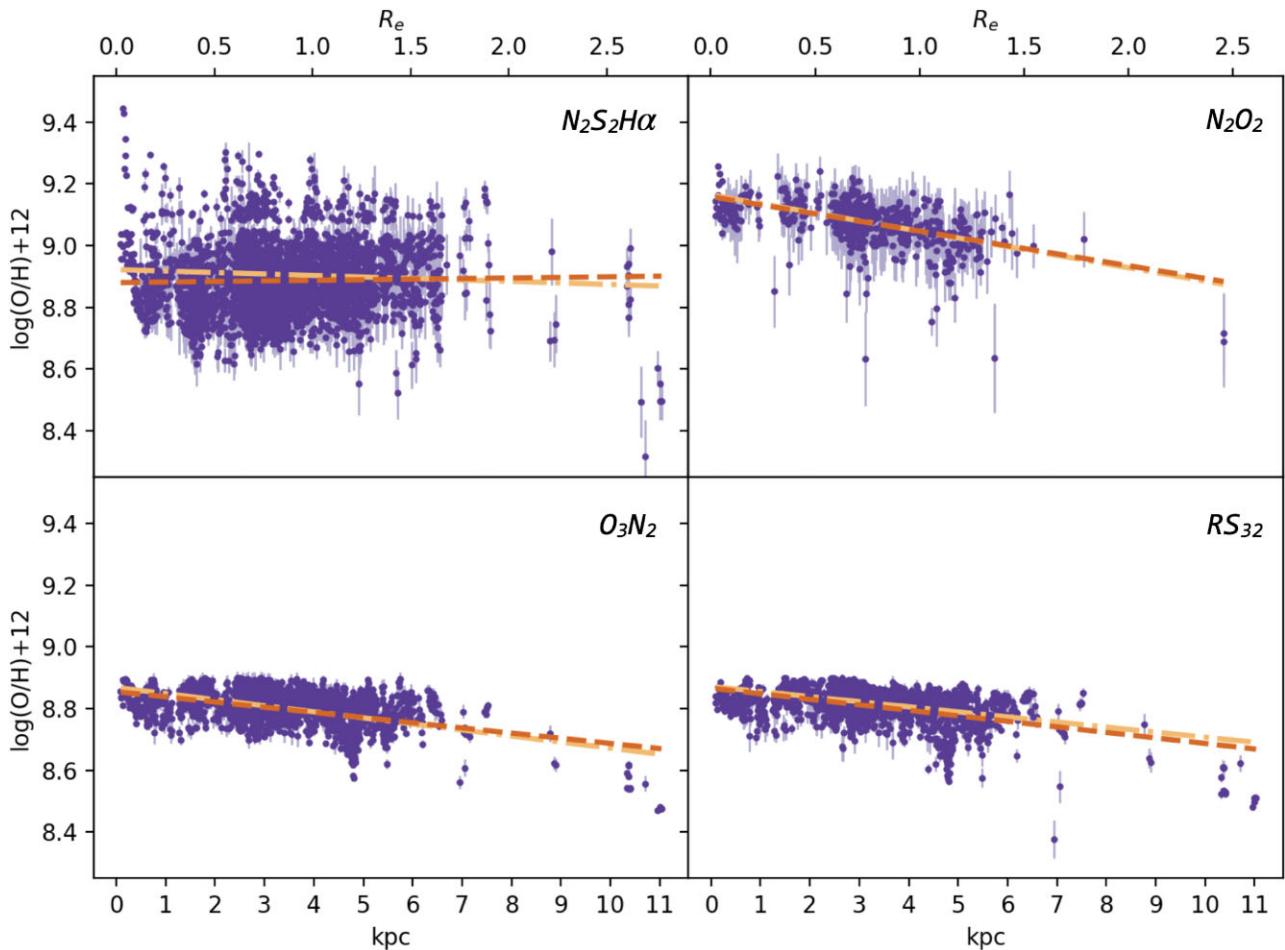


Figure 1. Radial metallicity profiles of H II regions in the galaxy NGC 5236, computed using the four different metallicity diagnostics described in Section B1. Uncertainties in the metallicity of each H II spaxel are computed using linear error propagation, and are shown as error bars. For each diagnostic, the best-fitting metallicity gradient model computed using an ML method accounting for small-scale correlations is shown with an orange dashed line. These are consistent with the best-fitting metallicity gradients computing using the WLS method, shown as the golden dashed-dotted line.

Table 3. Results of ML model fitting for the geostatistical model, compared to a WLS approach to compute the metallicity gradient using each metallicity diagnostic for NGC 5236.

Diagnostic	Fitting method	Z_c	$\langle \nabla Z \rangle$ (dex kpc $^{-1}$)	$\langle \nabla Z \rangle$ (dex R_e^{-1})	σ^2	ϕ (kpc)
$N_2S_2H\alpha$	ML model	8.88 ± 0.012	0.0019 ± 0.012	0.0076 ± 0.012	0.0100 ± 0.0013	0.1329 ± 0.015
	WLS	8.92 ± 0.004	-0.0048 ± 0.001	-0.0192 ± 0.005	–	–
N_2O_2	ML model	9.16 ± 0.009	-0.0265 ± 0.009	-0.1053 ± 0.010	0.0013 ± 0.0002	0.1331 ± 0.091
	WLS	9.16 ± 0.005	-0.0279 ± 0.002	-0.1110 ± 0.006	–	–
O_3N_2	ML model	8.85 ± 0.009	-0.0168 ± 0.009	-0.0665 ± 0.008	0.0022 ± 0.0024	0.2872 ± 0.386
	WLS	8.87 ± 0.002	-0.0196 ± 0.007	-0.0780 ± 0.003	–	–
RS_{32}	ML model	8.86 ± 0.011	-0.0179 ± 0.011	-0.0712 ± 0.009	0.0021 ± 0.0003	0.3980 ± 0.053
	WLS	8.91 ± 0.002	-0.0106 ± 0.006	-0.0420 ± 0.002	–	–

the metallicity of individual spaxels about this line is caused by measurement error (Hogg et al. 2010). For this galaxy, the radial metallicity profiles determined using this methodology agree with the popular WLS method for all diagnostics except $N_2S_2H\alpha$, where the geostatistical ML method favours a slightly positive metallicity gradient, as compared to the slightly negative metallicity gradient that is determined with WLS. However, in this case both models agree that the metallicity gradient is very close to 0, and are consistent within one standard deviation.

The metallicity gradients returned by each of these methods for this galaxy under each diagnostic are listed in Table 3. We note that the uncertainties in the values of Z_c and $\langle \nabla Z \rangle$ computed using the ML method are larger than those determined using the WLS method, which seemingly implies that the WLS method produces more precise metallicity gradient estimates. We stress that this is not the case. The smaller uncertainty in the WLS fitted parameters come from the erroneous assumption that there is no scatter about the radial metallicity trend besides measurement error, which would imply that the data points are very reliable tracers of the metallicity gradient, and only a small number of data points are needed in order to recover this global trend. On the other hand, under the geostatistical model, data are not expected to perfectly trace the average radial trend, leading to more realistic uncertainties in the recovered metallicity gradient parameters.

In Fig. 2, the small-scale structure of the metallicity fluctuations determined under each diagnostic are illustrated with semivariograms (see M21 for a thorough explanation of semivariograms and their role in extragalactic astronomy). Solid purple lines show the empirical semivariograms, computed using bins of width 0.1 kpc, up to a separation of 5 kpc. With all diagnostics, significant structure is seen in these semivariograms, revealing correlations between spaxels separated by up to 1–2 kpc. Similar behaviour has also been seen for a population of local star-forming galaxies observed by the PHANGS collaboration (Kreckel et al. 2020, M21). For $N_2S_2H\alpha$, O_3N_2 , and RS_{32} , the true variance about the mean metallicity at each radius (σ^2 in our model) is ~ 1 order of magnitude larger than the median variance in metallicity caused by error measurements. For O_3N_2 , N_2O_2 , and RS_{32} , the height of these semivariograms is broadly consistent with the height of the semivariograms computed using PHANGS data in M21. For $N_2S_2H\alpha$, the height of the semivariogram is much larger. This can be understood by examining Fig. 1: under this diagnostic, there is a much larger scatter around the mean metallicity at any radius than there is when any other diagnostic is used.

Dashed orange lines show the semivariograms associated with the most likely exponential model of $\eta(\bar{x})$ for each diagnostic, with maximum likelihood parameters listed in Table 3. For all diagnostics,

this model adequately captures the most important features of the small-scale metallicity structure; namely, the spatial extent of these local metallicity fluctuations (ϕ), and the total amount of correlated variance (σ^2). However, improvements could be made in modelling the general shape of this correlation function, as correlations between nearby datapoints seem to decrease with distance less rapidly than is predicted by an exponential covariance function. To properly model this behaviour, a physically motivated positive-definite covariance function may need to be constructed, using theoretical ideas about turbulence in the ISMs of local galaxies (in the style of e.g. Krumholz & Ting 2018). In this respect, semivariograms such as those constructed for this galaxy may be used as a tool to assist theoretical astronomers in the development of testable models of local ISM enrichment and metal transport.

Shaded regions in Fig. 2 show the 1σ uncertainty on the theoretical semivariogram produced by the fitted parameters, computed using equation (8).⁶ For the model based on the O_3N_2 diagnostic, the uncertainties in the best-fitting values of ϕ and σ^2 reported in Table 3 are very high. While the maximum likelihood geostatistical model for the metallicity distribution under this diagnostic is broadly consistent with the models obtained using other diagnostics, the size of the uncertainty in the fit implies that this data is also consistent with no additional uncertainty. We can understand the size of these parameter uncertainties by looking at the semivariogram associated with this data, shown in Fig. 2. Clearly, the data do exhibit spatial correlation. However, it is difficult to determine both the range over which data become uncorrelated and the limiting value of the semivariance as the separation tends towards infinity. This indicates that the exponential correlation model, chosen for its simplicity, may not be sophisticated enough to capture the small-scale metallicity trends associated with this diagnostic. A similarly high uncertainty is present in the ML estimate for ϕ when the N_2O_2 diagnostic is used, which further hints that the simple exponential model with a single correlation scale for η may not be sufficient for capturing all of the details of the metallicity structure of the ISM.

Using the geostatistical models calibrated to HII region data, predictions of the metallicity can be made at unmeasured locations using the technique of universal kriging, described in Section 4.2. We illustrate this process using the RS_{32} diagnostic for NGC 5236

⁶The uncertainties in the estimate of the measured semivariance at each separation (purple line) is very small, and is not shown. This is because the uncertainty in the estimate of the variance at each separation is proportional to the number of pairs of data points separated by that distance, and this number is very large (>990 for N_2O_2 and >9700 for all other diagnostics) at all separations shown.

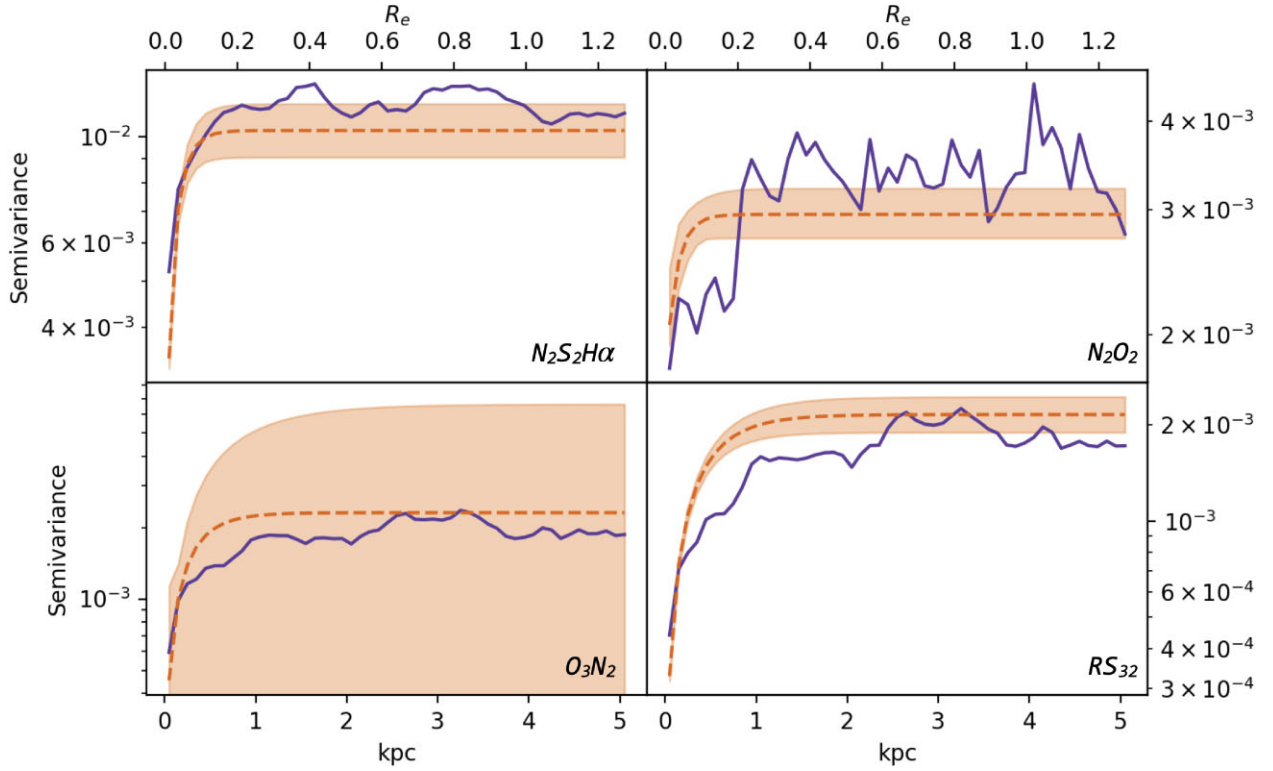


Figure 2. Semivariograms showing the small-scale metallicity structure of NGC 5236, using a suite of metallicity diagnostics. Purple solid lines show the empirical semivariogram. Orange dashed lines show the semivariogram predicted from the model of best fit. Orange shaded regions show the 1σ error in the fitted curve (errors on the semivariance estimation from the data at each separation shown are negligible). In all cases, the behaviour of the model shows good agreement with the data, although for the O_3N_2 diagnostic, the uncertainties on the parameters for the most likely model for small-scale structure are very large.

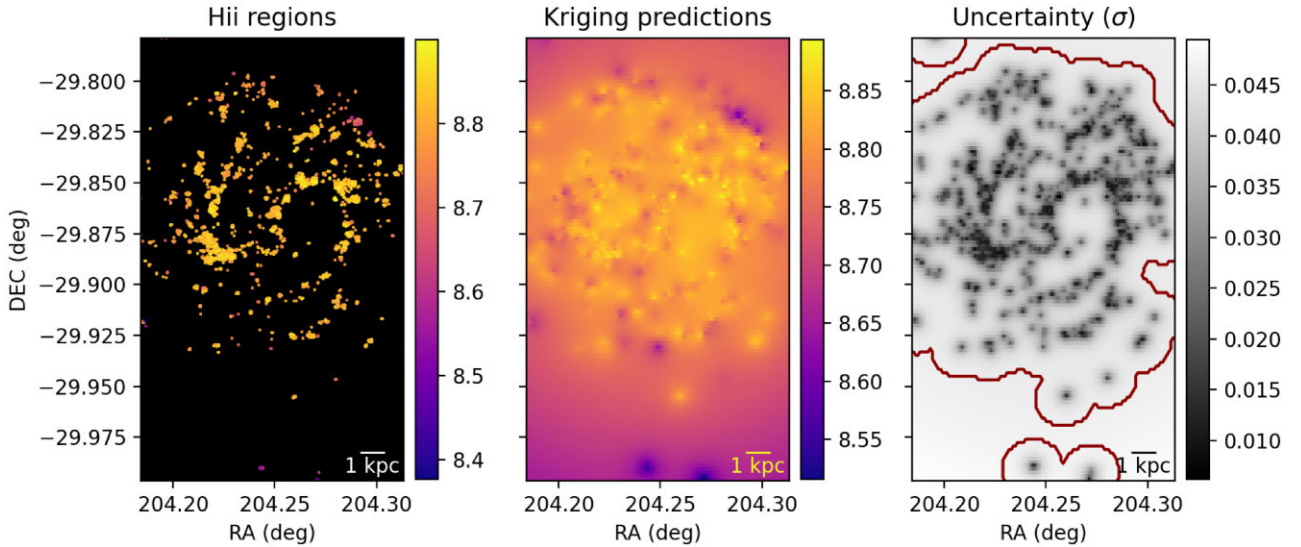


Figure 3. Left: Measured metallicity values, in units of $\log([O/H]) + 12$, for H II regions in NGC 5236, determined using the RS_{32} diagnostic. Centre: An interpolated metallicity map, computed using the technique of universal kriging, modelling small-scale metallicity fluctuations with a Gaussian model with an exponential correlation function. Right: Uncertainties (standard deviation) of the kriging predictions at each location. At distances greater than 3ϕ from any H II spaxel (red contour), the uncertainty of the metallicity estimated for any point is approximately the uncertainty given by a metallicity gradient model.

in Fig. 3. This diagnostic was chosen for visualization because it contains both a significant metallicity gradient, and substantial small-scale metallicity fluctuations. Using the World Coordinate System (WCS) data of the PrISM datacube for this galaxy, a grid of 128×128 sky locations was constructed, spanning the FOV of the constructed

metallicity map. At each of these locations, the metallicity was predicted using equation (9), with uncertainty given by equation (10).

In the left-hand panel of Fig. 3, we show the metallicity measurements for all H II spaxels. Together with our geostatistical model, these metallicities and their uncertainties are used to predict the

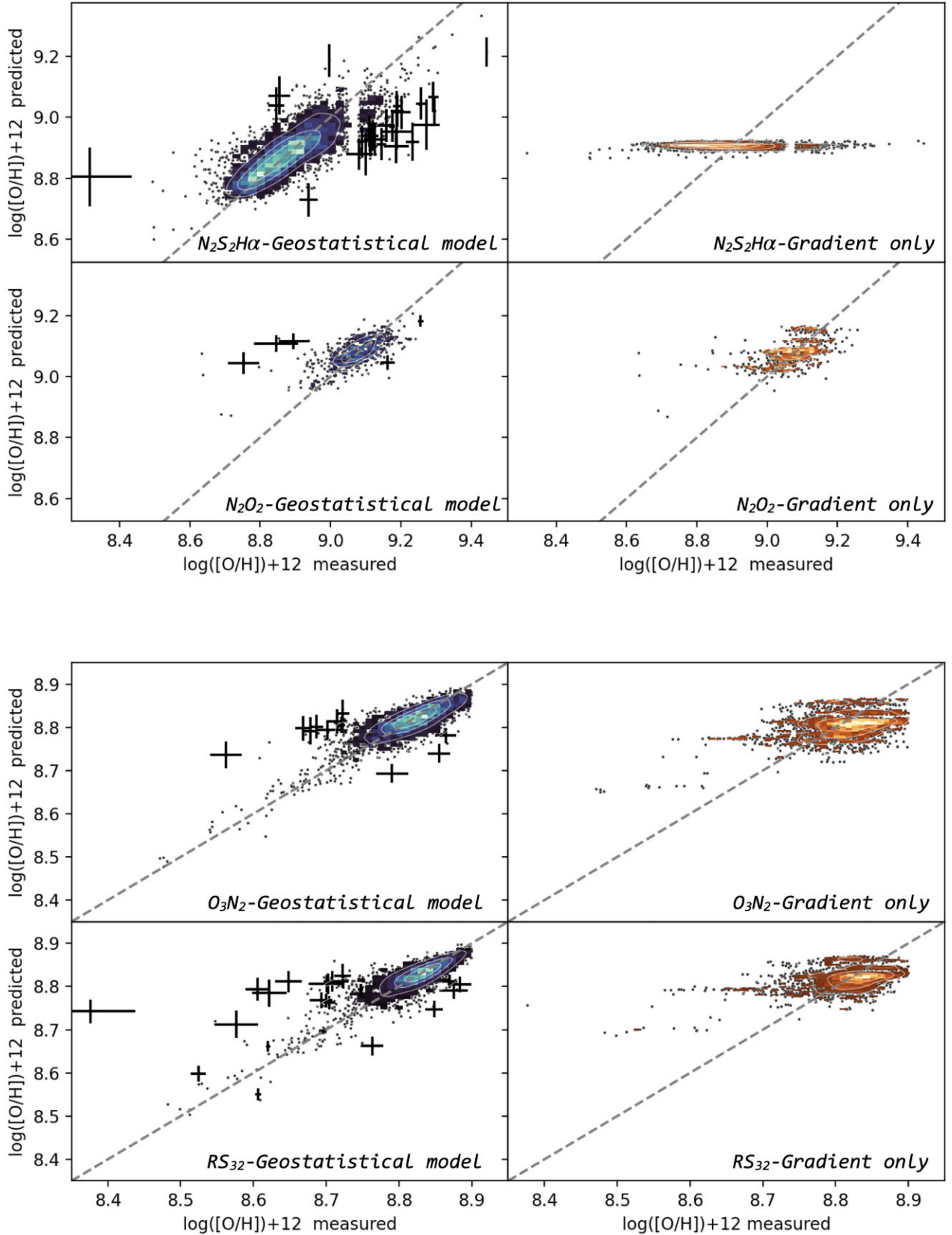


Figure 4. Scatter plots comparing the metallicities predicted for H II regions using 10-fold cross validation to the true measured metallicities of these points. Left: Comparing metallicities predicted with universal kriging to the measured metallicity of each spaxel, computed using the $N_2S_2H\alpha$ (top) and N_2O_2 (bottom) diagnostics. 3σ outliers are shown with error bars. Right: The same, but using predictions from a simple metallicity gradient model computed using weighted least-squares, without accounting for small-scale metallicity fluctuations. As this method does not return the uncertainty associated with each metallicity prediction, outliers are not explicitly shown.

metallicity at each location in NGC 5236. In the middle panel, we show the predicted metallicity map. Within this predicted map, metallicity decreases as distance from the galactic centre decreases – but not smoothly or uniformly. Instead, small-scale stochastic metallicity fluctuations can be seen, with a characteristic scale of $\phi \approx 400$ pc, revealing locations with local chemical enrichment and depletion. Interestingly, there appears to be some enrichment along the main spiral arms of NGC 5236 in the predicted 2D metallicity distribution in concordance with the chemical carousel model of Ho et al. (2017), even though such enrichment along spiral arms was not modelled explicitly. A quantitative exploration of this effect with a wider sample of galaxies, using geostatistical methods to improve completeness of data in the interarm regions of spiral galaxies, is an avenue for future research.

In the right-hand panel of Fig. 3, the uncertainty associated with each metallicity prediction is plotted. When predictions are made close to an observation point, the metallicity is expected to be highly correlated with the metallicity measured at the H II region, and so the uncertainty of the prediction is reduced to ~ 0.01 dex, similar to the uncertainty in metallicity measurements at that data point. As the distance from a H II region increases, the uncertainty likewise increases, in a manner that is consistent with the semivariogram for this diagnostic shown in Fig. 2. When the distance from any observed data point is larger than ~ 1.2 kpc ($\sim 3\phi$, shown by the red contour in the Fig. 3), the error in the predicted metallicity is equal to the true scatter around the metallicity gradient, plus the uncertainty associated with the metallicity gradient computed (see equation 10).

5.2 Validation

To verify the accuracy of this model, we predict the metallicity of each H II region using a model trained on a subset of the data. We then compare metallicities predicted by the geostatistical model to those measured directly using strong emission line diagnostics.

Formally, this was done by performing a 10-fold cross-validation (James et al. 2013). Each data point was randomly assigned to one of 10 groups. For each of these groups, a geostatistical model of the form described in Section 4 was fit using the ML method detailed in Section 4.1, calibrated using only the H II spaxels in the other 9 groups (the *training set*). Then, the value of the metallicity at each spaxel in the selected group (the *testing set*) was estimated using universal kriging, combining knowledge of the large-scale metallicity trend, $\mu(\vec{x})$ with correlations to nearby H II regions in the training set under our model of local metallicity fluctuations $\eta(\vec{x})$.

Fig. 4 shows the results of this analysis. In the left-hand panels, 2D histograms show the distribution of metallicities predicted using this method of universal kriging, as compared to the observed metallicities of each spaxel. All 3σ outliers are shown with error bars in the predicted and measured metallicity. In the right-hand panels, we repeat this cross-validation analysis, but instead predicting the metallicity of each H II region using a simple metallicity gradient model computed using WLS, without accounting for any correlations in the metallicities of nearby H II regions. Because this method does not produce uncertainties in the predicted metallicities of each data point, we do not show outliers explicitly.

Universal kriging shows the greatest improvement over the WLS method for NGC 5236 when the $\text{N}_2\text{S}_2\text{H}\alpha$ diagnostic is used. Using this diagnostic, the metallicity gradient of this galaxy is very close to zero, and so metallicity prediction methods that rely solely on the metallicity gradient will predict the metallicity of all data points to be approximately equal. Clearly, this is not the case. $\text{N}_2\text{S}_2\text{H}\alpha$

metallicities of H II spaxels in this galaxy have been measured to span a range of $\log([\text{O}/\text{H}]) + 12 = 8.6\text{--}9.2$. This large variance in metallicity can, however, be modelled using local metallicity fluctuations. When universal kriging is used, a large majority (76.20 per cent) of predicted metallicities lie within 1 standard deviation of the measured metallicity value, and 95.63 per cent agree to 2σ . Furthermore, for all diagnostics tested, less than 5 per cent of predicted metallicities differ from the true metallicities by more than 2σ , and the proportion of 3σ outliers is less than 1 per cent. These proportions are in line with the theoretical numbers of predictors within 1σ , 2σ , and 3σ expected from the 68–95–99.7 per cent heuristic, under the Gaussian assumption of the kriging predictor.

For all other diagnostics, the improvement in predictions gained from the use of universal kriging is more mild, but still significant. In particular, when a simple Z -gradient model is used, the metallicities of the most metal-rich spaxels is underpredicted, and the metallicities of the most metal-poor spaxels are overpredicted. This demonstrates that universal kriging is a powerful tool for capturing the detailed behaviour of all spaxels in the galaxy, including departures from a linear gradient trend.

To make the improvement in the accuracy of predictions gained by creating a geostatistical model clear, in Fig. 5 we plot histograms showing the difference between the metallicities predicted by this model (Z_{pred} , measured in units of $\log([\text{O}/\text{H}]) + 12$) and the metallicities measured (Z_{meas} , reported in the same units) for each of the 4 studied metallicity diagnostics. Two features are clearly visible from this graph. First, for all diagnostics, the distribution of $Z_{\text{meas}} - Z_{\text{pred}}$ is narrower for predictions obtained using universal kriging than those obtained using a metallicity gradient (although this improvement is only marginal for N_2O_2), reflecting the increased accuracy of the geostatistical model. Secondly, for $\text{N}_2\text{S}_2\text{H}\alpha$, O_3N_2 and RS_{32} , $Z_{\text{meas}} - Z_{\text{pred}}$ is slightly biased for the WLS model, with peak values of $Z_{\text{meas}} - Z_{\text{pred}}$ of -0.08 , $+0.02$, and $+0.02$ dex, respectively. Conversely, the distributions gained from universal kriging are unbiased. This is because universal kriging is the optimal unbiased linear estimator for Gaussian processes where the global trend is a priori unknown, and correlation between nearby data points cannot be assumed to be negligible (Wikle et al. 2019), whereas the WLS method is only unbiased under the assumption that the only deviation of data points from a linear trend line is due to uncorrelated Gaussian error (Hogg et al. 2010), which is not a valid assumption for this kind of data.

6 PREDICTING METALLICITIES IN DIG-DOMINATED REGIONS

In Section 5.2, we showed that the geostatistical models constructed via the methods described in Section 4.1 may be used to produce accurate predictions of the metallicities of H II regions. We now use the method of universal kriging to predict the metallicities of DIG-dominated regions, and compare these predictions to the metallicities that are measured using the H II region calibrated metallicity diagnostics listed in Section B1. In Fig. 6, the metallicities predicted by this model (Z_{pred}) are compared to the metallicities measured by applying the strong emission-line calibrations to the spectra of the DIG-dominated regions (Z_{meas}). In short, we find that generally they do not agree. For all diagnostics, values of Z_{pred} are almost always greater than Z_{meas} . This implies that either (i) the metallicity of the DIG is intrinsically lower than that of H II regions, (ii) our geostatistical model is not accurate, or (iii) that traditional strong emission line diagnostics underpredict the metallicity of the

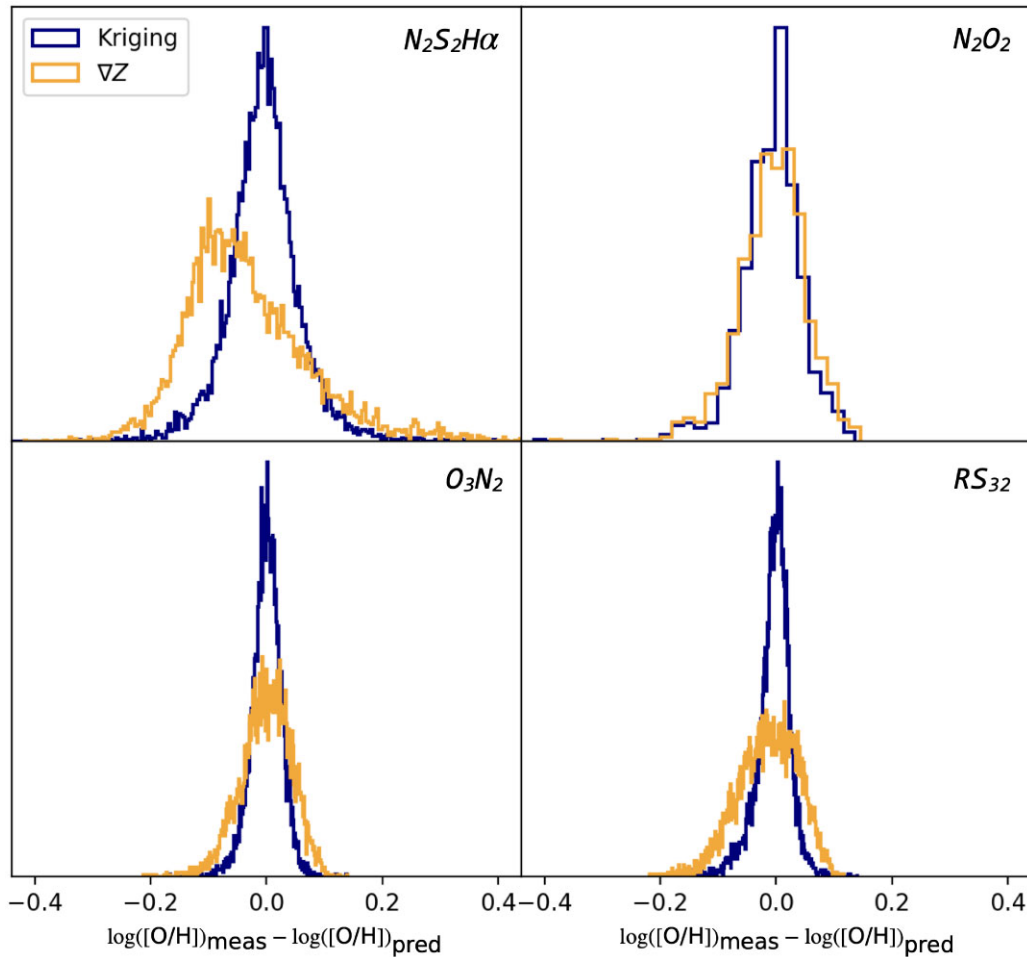


Figure 5. Histograms of the difference between the metallicities measured for each H II spaxel Z_{meas} , and the predicted metallicities using 10-fold cross validation Z_{pred} , using universal kriging (dark blue lines) and a simple metallicity gradient approach (gold lines). For all diagnostics, the metallicities predicted via universal kriging show less bias and are less scattered than predictions that do not account for local spatial trends.

DIG by a factor of ~ 0.2 dex. We believe that this final option is the most likely explanation, although we cannot rule out the first option. Depending on the efficiency of metal mixing throughout the ISM, H II regions may be more likely to be metal-rich than DIG-dominated regions, due to star formation being more likely to trigger in higher metallicity regions (e.g. Wolfire et al. 2003), and the recent release of metals from the deaths of massive stars creating local enrichment around the sites of the stars’ birth (Vogt et al. 2017).

The offset between Z_{meas} and Z_{pred} is revealed more clearly in Fig. 7. Dark blue lines show the same histograms presented in Fig. 5 for the geostatistical model. Maroon lines show the offset between Z_{meas} and Z_{pred} for DIG-dominated regions. For all of these diagnostics, the distribution of $Z_{\text{meas}} - Z_{\text{pred}}$ for DIG regions is (i) wider than for the H II counterpart, (ii) negatively biased, with modes ranging from -0.02 (N_2O_2) to -0.14 ($N_2S_2H\alpha$), and (iii) skewed negatively, with longer tails in the negative direction. This is likely due to the varying amounts of DIG contamination in each spaxel. By our selection methods described in Appendix B2, all spaxels with less than 90 per cent of their $H\alpha$ emission originating from H II regions are classified as being ‘DIG-dominated’; and metallicities determined for spaxels with higher fractions of DIG emission will show greater errors when measured with H II-calibrated Z-diagnostics.

Of the four diagnostics tested, we find that for N_2O_2 the values of Z_{pred} for DIG-dominated spaxels are closest to the values of Z_{meas} . This agrees with the results of Zhang et al. (2017), who state that N_2O_2 is largely insensitive to DIG contamination.

6.1 Testing DIG correction factors

Using data from the MUSE Atlas of Disks (MAD), Kumari et al. (2019) computed correction factors for the N_2O_2 and RS_{32} diagnostics that allow them to be applied to DIG-dominated regions (selected using the N2-BPT and S2-BPT diagnostics discussed in Appendix B2.2). Here, we detail the methodology used by Kumari et al. (2019) to construct these diagnostics, and test the metallicities measured using this method against predictions from our geostatistical model, as an independent assessment of their validity with a different sample of galaxies.

Spaxels from 24 nearby spiral galaxies observed as a part of the MAD program were classified as either H II or DIG/LIER regions, based on three classification schemes: the S2-BPT method, the N2-BPT method, and a surface brightness method based on a threshold flux for the $H\alpha$ brightness of H II regions of $\log \Sigma_{H\alpha} = 39$ erg kpc $^{-2}$ s $^{-1}$ (not considered in this work). A Voronoi-binning scheme was employed to ensure that all H II and DIG regions had

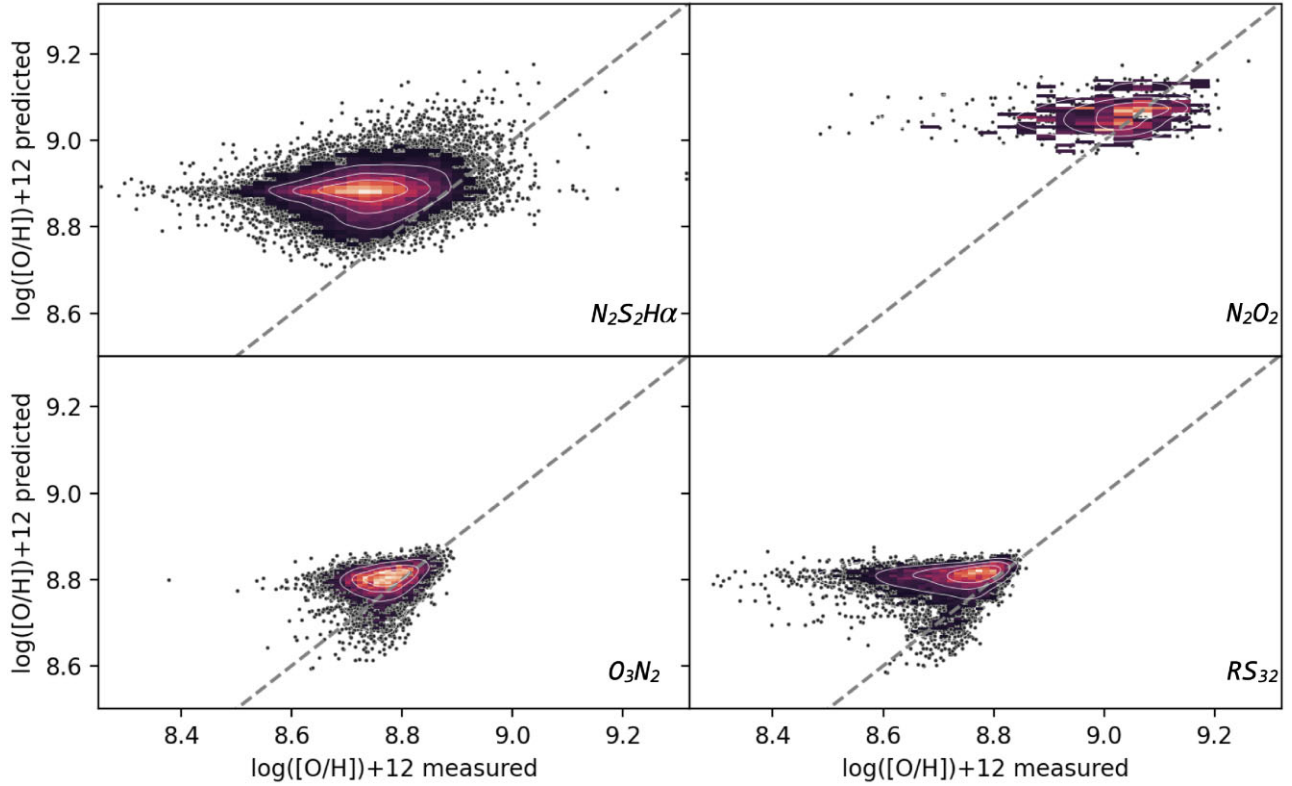


Figure 6. Scatter plots comparing the metallicities predicted for DIG-dominated regions in NGC 5236 using universal kriging, to those measured directly using H II region calibrated strong emission-line diagnostics. We find that the H II region calibrated diagnostics yield metallicities for the DIG-dominated spaxels that do not agree with our fitted models of metal mixing.

$S/N > 5$ for the emission lines [O III] λ 5007, H β , H α , [N II] λ 6584, and [S II] λ 6717, 31. For each Voronoi cell, metallicities were determined using the $N_2S_2H\alpha$, O_3N_2 , and RS_{32} diagnostics.⁷ Then, pairs of H II and DIG/LIER regions with angular separations greater than the 1 arcsec, and physical separations less than 500 pc were selected. The motivation for the first cut is that the seeing of MUSE for these observations ranges from 0.5 to 1.2 arcsec; this ensures that the correlations in metallicities between these nearby regions is not due to beam smearing. The second cut is motivated by the results of Berg et al. (2013), who found that H II regions separated by less than 500 pc had metallicity differences lower than 0.1 dex.

Using this selection of H II-DIG pairs, Kumari et al. (2019) proceed using a classical (as opposed to geo)statistical analysis, assuming that the true metallicity of each DIG region in a pair should be equal to the metallicity of the H II region of that pair. This assumption is discussed further in Section 7.1. Based on this assumption, $\Delta \log Z$, the difference between the metallicity measured for the H II region

and the DIG-dominated region in each pair was calculated, and this quantity was interpreted as the error term associated with using these metallicity diagnostics to infer metallicities of DIG/LIER. Similarly to the results of our geostatistical analysis presented in Section 6, they found that the differences between the metallicities measured directly for DIG regions and those inferred from the metallicities of nearby H II regions showed a large amount of scatter, and significant bias, especially when the RS_{32} and $N_2S_2H\alpha$ diagnostics were used. Furthermore, $\Delta \log Z$ was found to correlate significantly with the line ratio $O3$ ($= [O III]\lambda 5007 / H\beta$) for O_3N_2 and RS_{32} . Using this information, Kumari et al. (2019) fit first- and second-order polynomials to the relationship between $\Delta \log Z$ and $O3$, using both the S2-BPT and N2-BPT DIG diagnostics, to compute correction factors for O_3N_2 and RS_{32} that can be used to produce more accurate measurements of the metallicity of DIG.

In this section, we use our geostatistical model, calibrated on H II metallicities using the O_3N_2 and RS_{32} diagnostics, to predict the metallicity of DIG regions selected using the S2-BPT and N2-BPT methods. We compare these predicted metallicities to two measured metallicities: first, the metallicity of these DIG regions measured directly using the H II region calibrated O_3N_2 and RS_{32} diagnostics published in Curti et al. (2017) and Kumari et al. (2019); and secondly, adding in the $O3$ -dependent DIG correction terms computed by Kumari et al. (2019) using the method outlined above.

Fig. 8 shows the results of this analysis for the galaxy NGC 5236 as a scatter plot. Independently of whether the N2-BPT or S2-BPT diagnostic was used to separate H II regions from DIG, the metallicities of DIG regions determined using the $O3$ -corrected empirical metallicity diagnostics were found to be closer to Z_{pred}

⁷The RS_{32} diagnostic used in Kumari et al. (2019) is not the same as the RS_{32} diagnostic published in Curti et al. (2020a). Although it is computed using the same method, Kumari et al. (2019) fit only a third-order polynomial to the relationship between Z and RS_{32} , whereas Curti et al. (2020a) fit a fifth-order polynomial. Furthermore, in Kumari et al. (2019), the line ratio $\log([O III]\lambda 5007 / H\beta + [S II]\lambda 6717, 6731 / H\alpha)$ is referred to as O_3S_2 rather than RS_{32} . For consistency, we will refer to this diagnostic as RS_{32} throughout. In this and only this section, we will use the third-order RS_{32} calibration published in Kumari et al. (2019) and not the fifth-order calibration of Curti et al. (2020a), such that our results may be compared directly to those of Kumari et al. (2019).

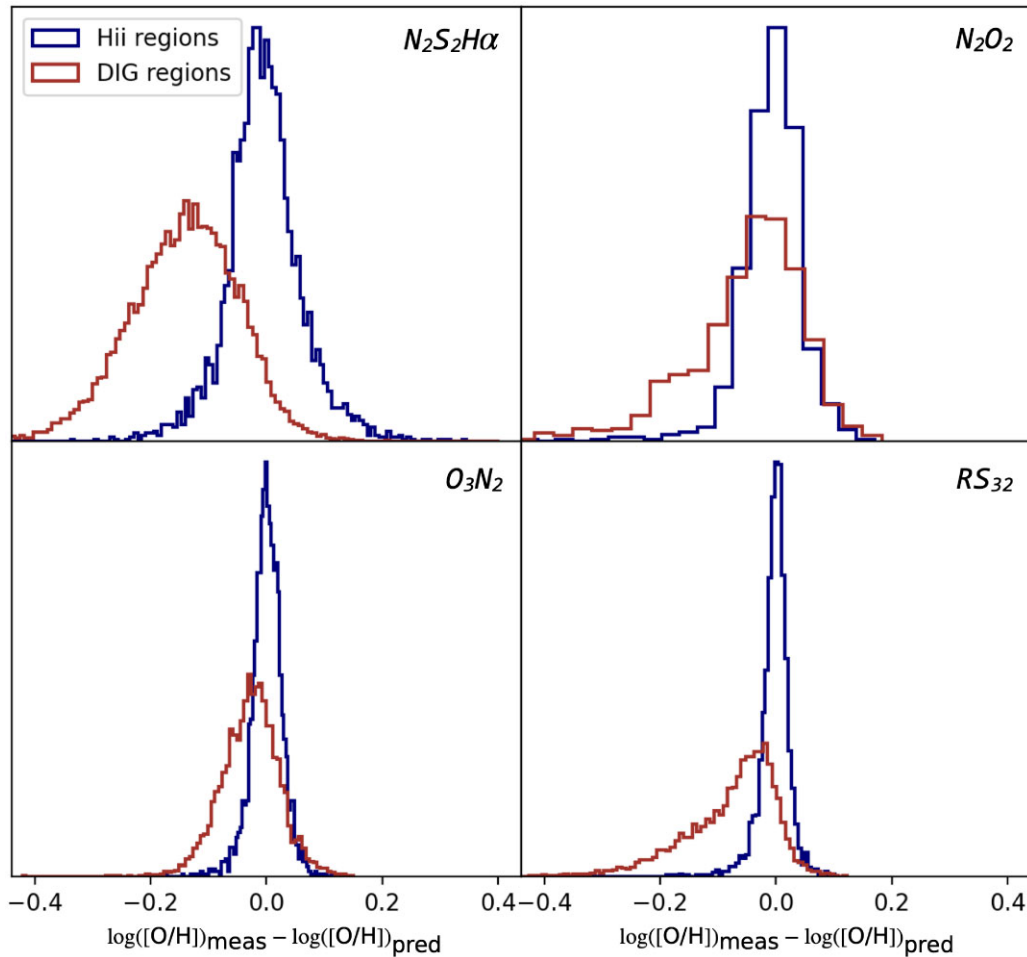


Figure 7. Histograms of the difference between the metallicities measured for each spaxel (Z_{meas}), and the metallicities predicted using universal kriging (Z_{pred}). Dark blue lines show the results of the ten fold cross-validation for H II regions described in Section 5.2. Maroon lines show the errors in predicted versus measured metallicities for DIG regions. All diagnostics show a significant negative bias – that is, the metallicities inferred from strong emission line diagnostics are lower than those expected from the geostatistical model.

than those computed using the uncorrected diagnostic.⁸ Under the assumptions that our geostatistical models are accurate (which is reasonable; see Section 5.2) and the spatial structure of metallicities throughout the DIG should be the same as the spatial structure of metallicities within H II regions (which is reasonable, as both are components of the warm ISM, and are mixed together by the same forces), this implies that the correction factors of Kumari et al. (2019) are successful at computing more accurate metallicities for regions of galaxies for which DIG contamination is appreciable. Interestingly, the O3-corrections also reduce the Z_{meas} produced, especially for RS_{32} . This may indicate that these correction terms are anticorrelated with the true metallicity of the DIG regions; when added, they bring the inferred metallicities closer to the mean. More sophisticated correction factors may need to be constructed to avoid this effect.

We show this same data as histograms in Fig. 9. Maroon lines are analogous to those shown in Fig. 7 for the N2-BPT and S2-BPT diagnostics; cyan lines show the reduction in scatter and bias gained when the O3-corrections are added. We replicate the result

of Kumari et al. (2019), in that these correction factors successfully reduce both the bias and scatter in $Z_{\text{meas}} - Z_{\text{pred}}$. However, some bias still remains. For NGC 5236, the mean value of $Z_{\text{meas}} - Z_{\text{pred}}$ for the O_3N_2 diagnostic is -0.04 before correction, decreasing to -0.03 after the O3 correction factor is added (this result does not depend on whether the N2- or S2-BPT diagnostic is used). For RS_{32} , the improvement is greater, with the mean bias in Z_{meas} decreasing from -0.12 to -0.04 when the N2-BPT diagnostic is used (likewise, from -0.14 to -0.05 with the S2-BPT diagnostic).

Unfortunately, this trend of decreasing total bias was not seen for all galaxies in our sample. In Table 4, we document the mean values and scatter in of $Z_{\text{meas}} - Z_{\text{pred}}$ for DIG regions in each galaxy, selected using both the N2- and S2-BPT diagnostics, comparing the amount of bias before and after the O3 corrections are applied. We find that generally the bias is lower when the DIG-correction factors are used. However, several cases can be seen where the O3-based correction factor introduces a larger bias than the uncorrected diagnostic. Furthermore, the scatter in $Z_{\text{meas}} - Z_{\text{pred}}$ often is just as wide when the corrected diagnostics are used as when the uncorrected diagnostics are used. We conclude that while the O3-based correction factors of Kumari et al. (2019) are successful in improving predictions of metallicities for regions with significant DIG contamination, these correction factors are not universal. Indeed, different galaxies may

⁸We note that less data points classified as DIG dominated when the S2-BPT diagnostic is used compared to when the N2-BPT diagnostic is used – this is discussed further in Appendix B.

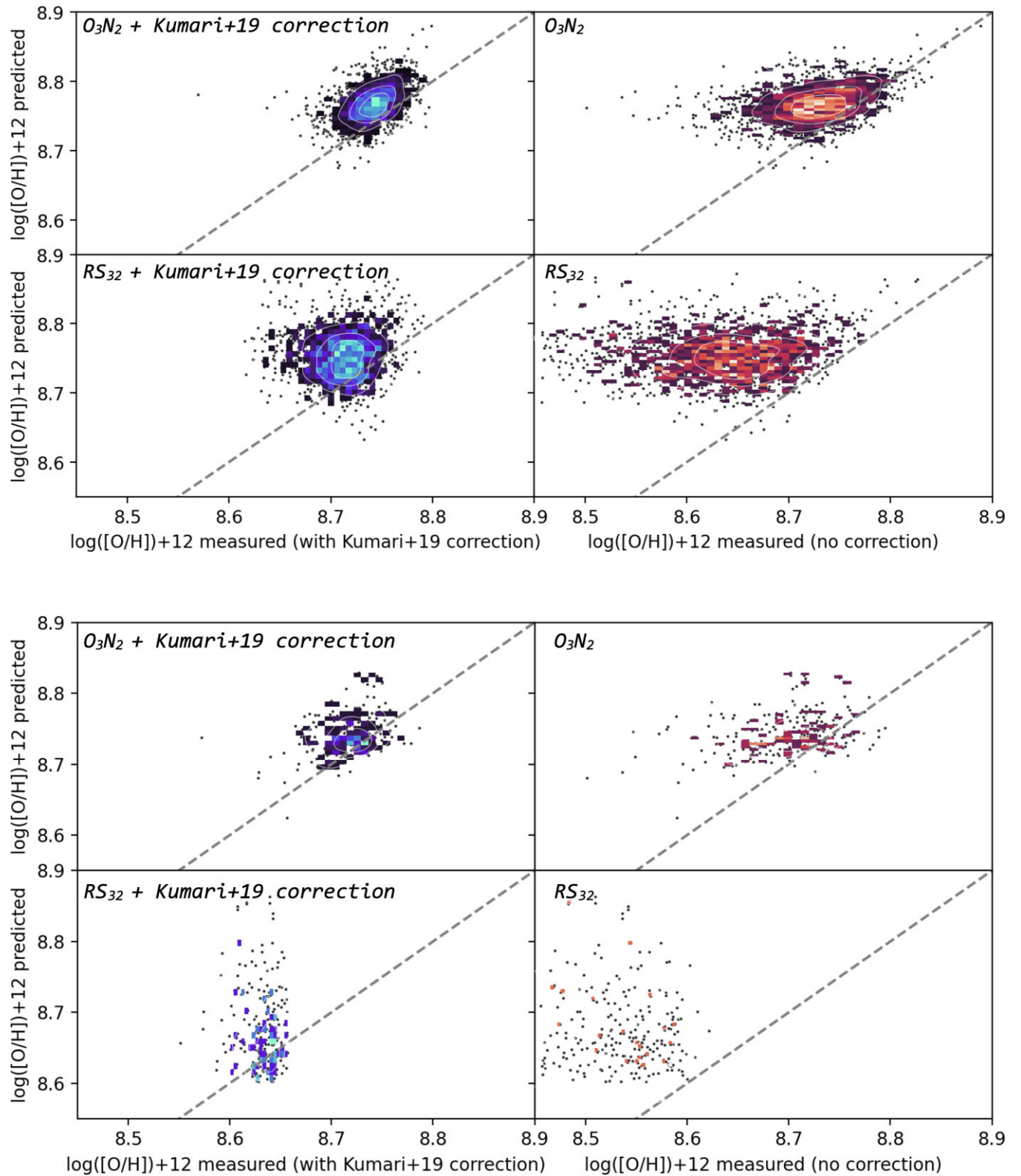


Figure 8. Left: Scatter plot showing the difference between the metallicities predicted for DIG regions, using the geostatistical model calibrated on H II regions, and the metallicities measured for the DIG regions using the DIG-corrected empirical diagnostics of Kumari et al. (2019). Right: The same, but without applying the DIG corrections to the empirical metallicity diagnostics. In the top panel, H II regions are selected using the N2-BPT method, while in the bottom panel, the S2-BPT diagnostic is used.

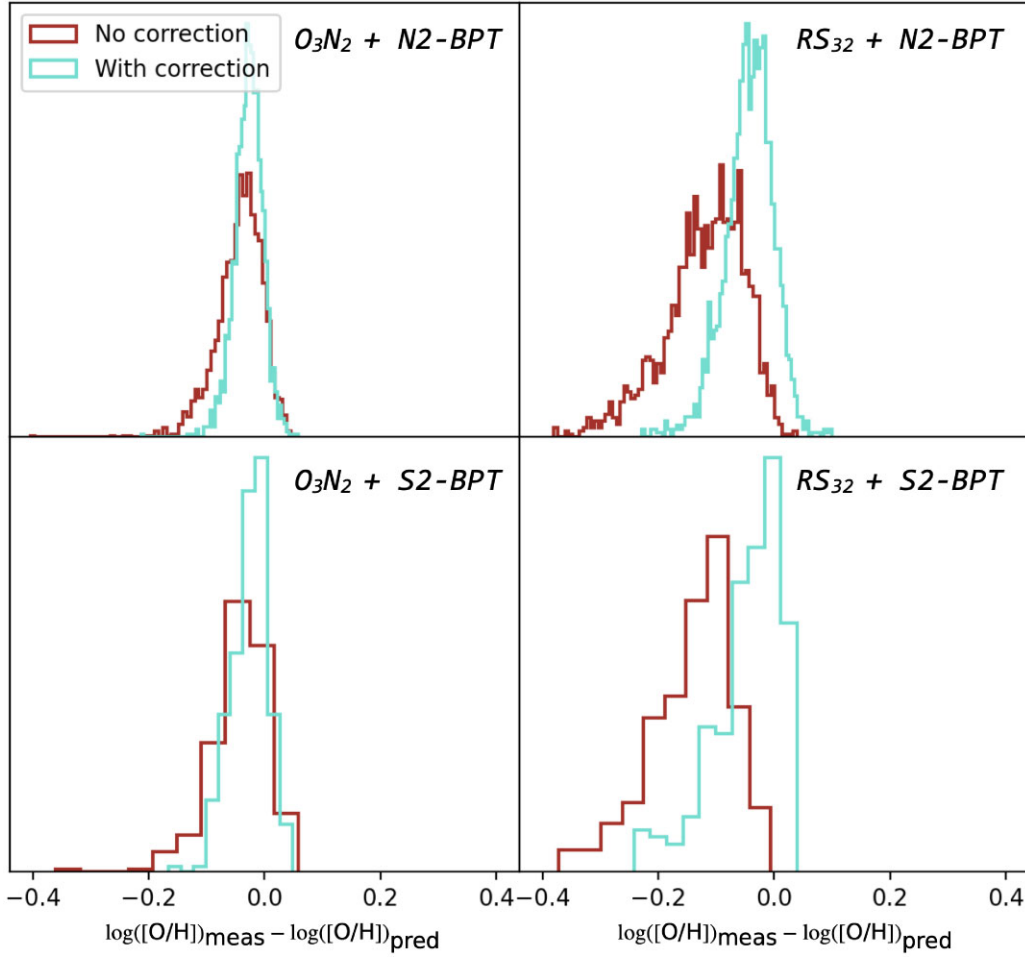


Figure 9. Histograms showing the difference between the metallicities predicted for DIG-dominated regions in NGC 5236 using a geostatistical model and those measured by the empirical metallicity diagnostics published in Kumari et al. (2019) (i) without DIG correction terms (maroon) and (ii) including the DIG correction terms (cyan). For this galaxy, the O3-dependant correction terms decrease both the bias and the scatter in $Z_{\text{meas}} - Z_{\text{pred}}$ – however, this is not always the case (see Table 4 and discussion in text).

Table 4. Mean offset (μ) and scatter (σ) in $Z_{\text{meas}} - Z_{\text{pred}}$ found for DIG regions for each galaxy, when using the corrected versus uncorrected O_3N_2 and RS_{32} metallicity diagnostics.

Galaxy	Type	N2-BPT				S2-BPT			
		$\mu-O_3N_2$	$\sigma-O_3N_2$	$\mu-RS_{32}$	$\sigma-RS_{32}$	$\mu-O_3N_2$	$\sigma-O_3N_2$	$\mu-RS_{32}$	$\sigma-RS_{32}$
NGC 1068	Before correction	-0.08	0.05	-0.17	0.08	-0.04	0.04	-0.08	0.04
	After correction	-0.02	0.02	-0.10	0.04	0.04	0.03	0.07	0.05
NGC 1365	Before correction	-0.02	0.05	-0.10	0.07	-0.03	0.05	-0.07	0.04
	After correction	0.02	0.04	-0.01	0.07	0.03	0.04	0.05	0.05
NGC 1566	Before correction	-0.01	0.07	-0.11	0.07	-0.02	0.05	-0.11	0.06
	After correction	0.02	0.06	-0.01	0.06	0.01	0.05	-0.01	0.06
NGC 2835	Before correction	-0.01	0.08	-0.07	0.05	-0.03	0.06	-0.06	0.04
	After correction	0.06	0.06	-0.06	0.04	0.01	0.05	0.06	0.05
NGC 2997	Before correction	-0.02	0.06	-0.09	0.06	-0.04	0.05	-0.12	0.06
	After correction	0.02	0.05	0.01	0.06	-0.01	0.04	-0.02	0.07
NGC 5068	Before correction	-0.01	0.06	-0.08	0.05	-0.04	0.06	-0.08	0.04
	After correction	0.05	0.04	0.06	0.04	0.00	0.04	0.03	0.05
NGC 5236	Before correction	-0.04	0.04	-0.12	0.07	-0.04	0.05	-0.14	0.07
	After correction	-0.03	0.02	-0.04	0.04	-0.03	0.03	-0.05	0.06
NGC 7793	Before correction	-0.01	0.07	-0.07	0.05	-0.04	0.05	-0.08	0.04
	After correction	0.04	0.06	0.05	0.04	-0.01	0.04	0.01	0.05

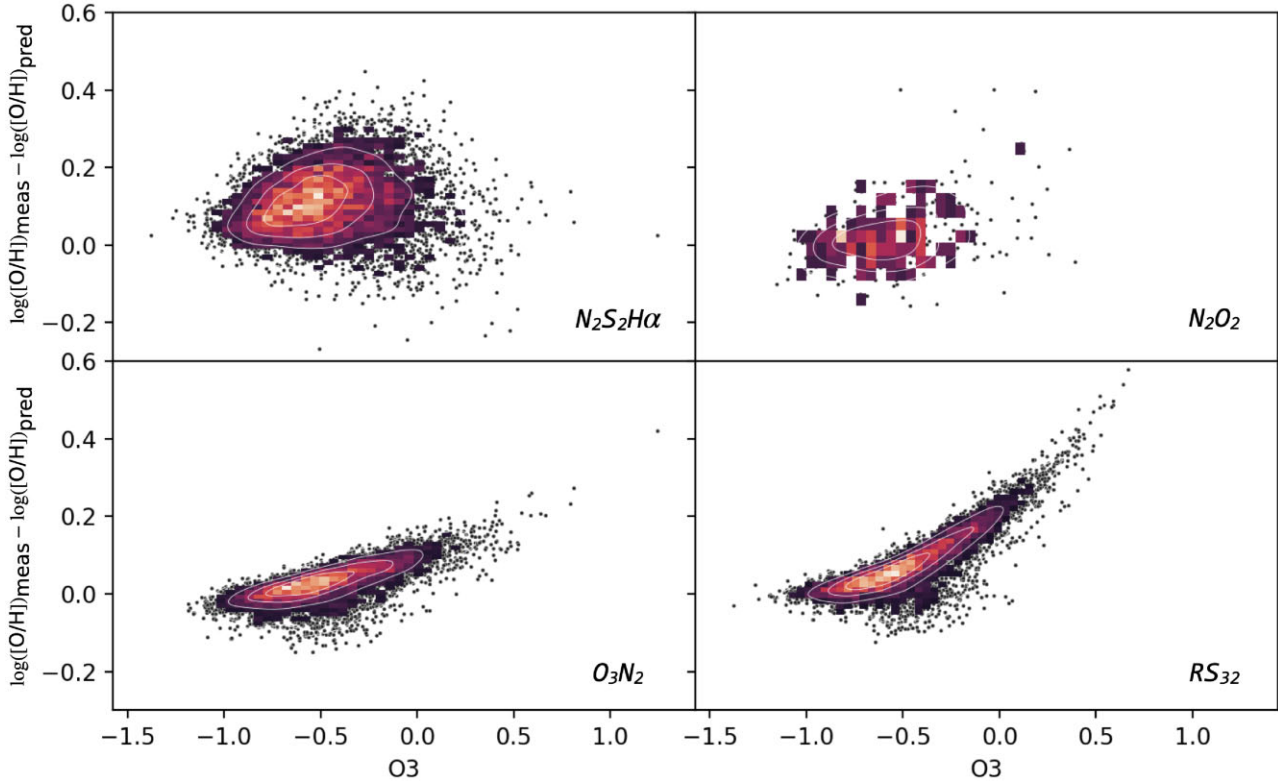


Figure 10. The deviance between metallicities measured directly using strong-emission line diagnostics for DIG-dominated regions, and those predicted via universal kriging, versus the line ratio $O3 = [O\text{ III}\lambda 5007]/H\beta$, for NGC 5236, selecting H II and DIG-dominated regions using the S2-based method described in Kaplan et al. (2016). For the two empirically calibrated diagnostics $O3N2$ and $RS32$, a strong correlation is seen between the error in the measured metallicities and the $O3$ line, in excellent agreement with the findings of Kumari et al. (2019). For the two theoretically calibrated diagnostics, no strong correlation is observed.

host different sources of DIG ionization, including evolved stars, shocks, and leaking H II regions. If this were the case, then finding a single correction factor to negate the effects of DIG contamination may not be possible, due to the different sources of DIG ionization in each galaxy. We discuss this point further in Section 7.

Finally, we demonstrate how our method of geostatistical analysis may be used to construct $O3$ -dependant correction factors, similarly to the methodology of Kumari et al. (2019). In Fig. 10, we plot $Z_{\text{meas}} - Z_{\text{pred}}$ (using the Curti et al. 2020a calibration for $RS32$) for DIG-dominated spaxels of NGC 5236, selected using the S2 method of Kaplan et al. (2016), against the value of the line ratio $O3$ for our four metallicity diagnostics. We see that for the two theoretically calibrated diagnostics, there is no significant correlation. However, for both the $O3N2$ and $RS32$ calibrations, the offset is significantly correlated with $O3$. The trend of $Z_{\text{meas}} - Z_{\text{pred}}$ with $O3$ is well-fit by a linear relation for the $O3N2$ diagnostic, and a higher order polynomial for $RS32$, in qualitative agreement with the results of Kumari et al. (2019) using other DIG diagnostics. This analysis demonstrates how the results of universal kriging may be used to construct new correction factors for empirically calibrated metallicity diagnostics to be used on DIG-dominated regions. We postpone the construction of such correction factors until the full release of data from the TYPHOON survey (Seibert et al., in preparation).

7 DISCUSSION

Thanks to the efforts of large population IFS galaxy surveys such as CALIFA (Sánchez et al. 2012a), MaNGA (Bundy et al. 2015),

and SAMI (Bryant et al. 2015), a clear unified picture of star-forming galaxies in the local universe has emerged. Global relationships between galaxy properties, such as the mass–metallicity relationship (MZR, e.g. Tremonti et al. 2004), the star-forming main sequence (Renzini & Peng 2015), and the Kennicutt–Schmidt relation (Schmidt 1959; Kennicutt 1998) have been found to be large-scale consequences of local scaling laws (e.g. Bigiel et al. 2008; Moran et al. 2012; Rosales-Ortega et al. 2012; Sánchez et al. 2013; Wuyts et al. 2013). Furthermore, observational evidence suggests that galaxies usually evolve inside-out, with inner regions showing higher star formation rates, metallicities, gas densities, and stellar ages (Sánchez 2020).

This picture holds over a large range of masses and morphologies. However, in this work, we show that the simple linear metallicity gradient is far from the whole picture. The failure of simple azimuthally symmetric models to capture the complex structures of the ISM of galaxies has been noted explicitly by many studies, especially at high redshift, and when dealing with high-resolution data. In a study of strongly lensed galaxies at cosmic noon ($1.2 < z < 2.4$), Curti et al. (2020b) find that the process of fitting a metallicity gradient can often obscure kpc-scale inhomogeneities that are common in starbursting systems. Similarly, Florian et al. (2021) suggest that star-formation in strongly lensed galaxies at $z \sim 1.4$ is better characterised by its ‘clumpiness’ rather than a large-scale gradient.

Our geostatistical approach is useful for describing both the large-scale radial trend and the small-scale clumpy structure of the ISM in a parsimonious way, using only four parameters. Furthermore,

this approach is based on over 60 yr of literature, with applications ranging from agriculture to epidemiology.

For all of the galaxies in our extended sample of eight nearby galaxies observed using the TYPHOON/PrISM survey, we find that galaxies show stochastic variations over characteristic spatial scales of ~ 1 kpc. This is seen regardless of the metallicity and DIG diagnostics used to construct H II metallicity maps. In Section 5, we show that even a simple exponential model of the stochastic small-scale metallicity structure is able to account for a large amount of the variance found around the large-scale metallicity trend. Furthermore, in Section 5.2, we show that by accounting for these small-scale fluctuations, the accuracy of the prediction of the metallicity at all regions of a galaxy can be vastly improved, especially when the metallicity gradient of the galaxy is close to zero. This makes this technique particularly appealing for the analysis of high-redshift galaxies, where most galaxies do not exhibit large-scale metallicity gradients (Simons et al. 2021). Results of our universal kriging procedure will be presented for the seven currently unpublished galaxies listed in Table 1 in a future paper (Metha et al., in preparation).

This technique has potential applications in many areas of extragalactic astronomy where accurate metallicity predictions are useful. In particular, this technique may be valuable in determining the value of the Hubble constant, H_0 , using observations of the local universe. Type Ia supernovae (SNe Ia) are the prototypical standard candles, with a narrow scatter in their intrinsic luminosity (Folatelli et al. 2010). These bright transient events can be used as a cosmological distance ladder when calibrated using other standard candles, such as RR Lyrae variables and Cepheids, which have well-defined period luminosity (PL) relations. However, these PL relations are known to depend on metallicity (Nemec, Nemec & Lutz 1994; Groenewegen et al. 2004). To calibrate these relations, accurate predictions of the metallicity of the interstellar gas at the locations of these stars are needed. Metallicity gradients alone have a large prediction variance at all radii, which can contribute significantly to the error budget of the cosmological distance ladder (Beaton et al. 2016). In Fig. 3, we show that this prediction variance can be decreased when universal kriging is used. These more accurate predictions would lead to a more accurate calibration of the zero-point of the SNe Ia luminosity function, leading to tighter constraints on the value of H_0 in the local universe.

In order to compute high-resolution molecular hydrogen column density maps for nearby galaxies, the PHANGS-ALMA survey must estimate the metallicity dependant CO-to-H₂ conversion factor, α_{CO} . In Sun et al. (2020), this is done by assuming a linear gradient model, with Z_c estimated from a galaxy’s stellar mass using the MZR, and a constant assumed metallicity gradient of $0.1 \text{ dex } R_e^{-1}$ for all galaxies. Therefore, estimates of the molecular gas density distribution could be improved upon by using these methods, together with PHANGS-MUSE data. We emphasize that our universal kriging method could provide metallicities at all locations of the galaxies, even at regions of DIG contamination.

As telescope instrumentation continues to improve, more and more IFS surveys are being conducted which allow the ISM of nearby galaxies to be resolved to scales of 100 pc or less, below the size of an individual H II region (Mannucci et al. 2021). Recent and upcoming surveys with resolutions similar to TYPHOON/PrISM include PHANGS (Emsellem et al. 2022), MAD (Erroz-Ferrer et al. 2019), the SDSS-V LVM surveys of the LMC and SMC (Kollmeier et al. 2017), TIMER (Gadotti et al. 2019), SIGNALS (Rousseau-Nepton et al. 2019), GASP (Poggianti et al. 2017), and AMUSING (Galbany et al. 2016). The geostatistical techniques described in

this work are naturally applicable to this new generation of IFS surveys. Comparisons between the small-scale metallicity structure of the ISM of these galaxies and the output of dedicated zoom-in comparison simulations (Jeffreson et al. 2020; Utreras et al. 2020) will allow us to constrain models of metal mixing through stellar feedback. Furthermore, these techniques may be used in conjunction with data from the TIMER survey to reveal the effects of spiral bars and nuclear structures on a galaxy’s internal metallicity distribution, or the GASP survey to better understand the mixing effects in interacting galaxies.

In Fig. 10, we show that the line ratio O3 correlates with the offset between Z_{meas} and Z_{pred} for two empirically calibrated metallicity diagnostics, confirming the result of Kumari et al. (2019). Current state-of-the-art Bayesian metallicity fitting programs such as IZI (Blanc et al. 2015) are not able to correct for DIG contamination in regions (Zhang et al. 2017). Our Fig. 10 implies that the O3 line ratio may be useful in correcting for DIG contamination in order to produce a robust metallicity fit.

7.1 Comparisons to other works

This study is not the first that attempts to predict the metallicity throughout the ISM using data from H II regions. In this section, we briefly review some other key studies, and highlight how our analysis differs from the current state of the art.

Using data from 365 galaxies from the MANGA survey, Zhang et al. (2017) derive the bias occurring from DIG contamination associated with different strong emission line based metallicity diagnostics. Assuming that the proportion of DIG contamination within a spaxel is proportional to the surface brightness of H α , a linear relation was fit between the metallicity derived using several strong-line diagnostics, and the amount of DIG contamination. The effect of metallicity gradients was avoided by only selecting spaxels within $0.4\text{--}0.6R_e$ within each galaxy. No local metallicity variations or metallicity measurement errors were accounted for. In the language of our geostatistical model, this would be equivalent to setting $\eta(\vec{x}) = 0$ and $\epsilon(\vec{x}) = 0$. Out of seven diagnostics tested, Zhang et al. (2017) concluded that the two least sensitive to contamination from DIG were the N₂S₂H α diagnostic with the calibration of Dopita et al. (2016), and the O₃N₂ diagnostic with the calibration of Pettini & Pagel (2004).

This method, while statistically robust, does not account for any local variations in metallicity. For this reason, a substantial fraction of the data need to be discarded in order to make this technique work. In contrast, by using geostatistical modelling, all the data in the galaxy could be used, leading to a larger sample of spaxels, which could strengthen the conclusions of such a study.

Similarly, the approach of Kumari et al. (2019) can be casted as a geostatistical analysis, but with a very specific form for the correlation function. Specifically, the correlation in metallicity measurements at two different data points is treated as a top hat function: $\text{Corr}(Z(\vec{x}), Z(\vec{y})) = 1$ if $|\vec{x} - \vec{y}| < 0.5$ kpc, and 0 otherwise. This correlation structure would induce a shape for the theoretical semivariogram of the following form for all galaxies:

$$\gamma(h) = \begin{cases} 0, & \text{if } h < 0.5 \\ \sigma^2, & \text{otherwise} \end{cases} \quad (12)$$

Our Fig. 2 shows that such a model is not a good fit for local star-forming galaxies. Furthermore, the method of Kumari et al. (2019) does not account for any large-scale metallicity trends, such as metallicity gradients. Finally, such a step-based correlation function

does not lead to a positive definite covariance function.⁹ Therefore, the model of Kumari et al. (2019) for describing the small-scale homogeneity of the ISM is not statistically valid, and its application may thus lead to hard to quantify biases in the inference drawn from it.

Using a hierarchical model optimized with the Integrated Nested Laplace Approximations method (INLA; Rue et al. 2017), González-Gaitán et al. (2019) fit a model of the spatial variation of the age and metallicity of stellar populations in 721 galaxies from the CALIFA and PISCO (Galbany et al. 2018) surveys. The model that they fit is similar to our own in many respects: in both cases, the mean trend μ is modelled as being linearly dependant on the radial distance from the galaxy centre; and in both cases, the observed metallicity at each location is assumed to be drawn from a latent Gaussian process with some known observation error. Our analysis differs from theirs in two key respects: first, in that work the covariance kernel of the small-scale metallicity variation η is a Matérn function, rather than an exponential function, allowing one extra free parameter to be fit.¹⁰ Secondly, the fitting method that they use is Bayesian, whereas ours uses a maximum likelihood approach. Using a Bayesian method may make the results dependant on the model priors, and be computationally costly to compute, even though the speed of computations is significantly increased by using the INLA approximate Bayesian computing method.

Similarly, Clark et al. (2019) also use Bayesian methods to fit a Gaussian process model of the metallicity variation within galaxies, using a Matérn function of order 1.5 to capture deviations from a mean radial trend. Using a cross-validation analysis (their Appendix C), they found that this method produced accurate predictions of the metallicities of H II regions, with 80.5 per cent (85.4 per cent) of predicted H II region metallicities agreeing with their measured metallicity values for M74 (M83) – larger than the 68.3 per cent proportion expected for a Gaussian process. This result may be caused by an overestimation of the uncertainty of the difference between the predicted metallicity and the actual metallicity: in their equation (C1), the variance of this difference is derived under the assumption that the predicted metallicity and the actual metallicity are independent of each other at each testing point. However, since the predicted metallicity is constructed using the metallicities observed, and these observed metallicity are spatially correlated with the true metallicity at each testing point, the predicted metallicity and the actual metallicity are likely to be correlated. Accounting for this positive correlation between predictions and the true metallicity would likely reduce or remove this tension.

The methodology of Williams et al. (2022) follows Clark et al. (2019) very closely, fitting the same model using the same software (GaussianProcessRegressor from the PYTHON machine learning package `scikit-learn`), and focusing on analysing the results of their best-fitting models in order to understand the spatial scales over which galaxies may be considered well-mixed. The length-scales of best fit for their Matérn kernels range from 1.6 to 30 kpc for the 12 galaxies with significant metallicity variations beyond a radial trend (see left-hand panel of fig. 9 in Williams

et al. 2022), with a median length-scale of ~ 16 kpc, comparable to the size of the galaxy. This length-scale is much larger than the ~ 0.1 – 0.4 kpc mixing scales we found with our exponential kernel. Williams et al. (2022) also computed the 50 per cent correlation scale for these PHANGS galaxies, finding mixing lengths of 0.2–1.1 kpc (see right-hand panel of fig. 9 in their work), that are in agreement with our findings. Therefore, we conclude that it is not the data, but the modelling methodology of Williams et al. (2022) that leads to the observation of non-radial metallicity variations on a different physical scale to the model presented here. In our analysis, we separate the signal $Z(x)$ into a large-scale global trend ($\mu(x)$, e.g. a metallicity gradient) and small-scale fluctuations ($\eta(x)$; characterised through the semivariogram analysis), whereas the models of Williams et al. (2022) capture galaxy-scale trends that are none the less significantly different from a simple metallicity gradient-based approach. We emphasize that the small-scale fluctuations we uncover in our work cannot be consistent with measurement error, as they are spatially correlated on scales of ~ 1 kpc, whereas measurement error is expected to have no spatial correlation.

In our work, we model the small-scale stochastic structure of the ISM as a stationary, isotropic random field. Other models (e.g. Sale & Magorrian 2014; Lee & Gammie 2021) relax this assumption, employing non-stationary models that allow parameters such as the correlation length-scale to vary throughout a galaxy. This procedure introduces additional complexity that makes inference more challenging. We find that a stationary process with a constant mixing scalelength, while perhaps less realistic, is sufficient to reproduce accurately the metallicity substructures observed in galaxies, for the purpose of making reliable predictions (see Section 5.2).

All of these examples show how our novel geostatistical techniques provide a complementary approach to existing literature approaches to spatially resolved studies of metal distribution in galaxies, and offer distinct advantages.

7.2 Caveats

No model exists without assumptions. In this section, we detail the assumptions that we make in constructing our geostatistical model, and our justifications for making them.

Crucially, this analysis assumes that the metallicity of DIG-dominated regions is governed by the same physical processes that determine the metallicity of H II regions. This assumption allows us to use a geostatistical metallicity model trained on H II region data to estimate the metallicities of DIG-dominated regions. If there was a reason to believe that the metallicity of DIG regions was expected to be intrinsically lower or higher than the metallicity of H II regions, this bias would need to be accounted for before any metallicities could be predicted. Based on current evidence, we see no physical reason as to why this would be the case. Both the DIG and H II regions exist in the warm phase of the ISM. Mixing effects due to interstellar turbulence, bar-driven motion, and other phenomena that act to redistribute metals through the ISM should not be significantly affected by the ionization phase of the medium. Furthermore, the model we fit for η , the small-scale metallicity structure of the ISM, is so simple that it will likely not be affected by accounting for these effects, if they exist.

This analysis also ignores the multiphase structure of the ISM and assumes each phase has a similar mixing scale. This assumption may not be accurate (see e.g. De Cia et al. 2021). Gas particles in the hot phase of the ISM have higher velocities than warm and cold ISM constituents, which may lead to shorter time-scales and spatial scales for mixing. For a first approximation, we assume that an exponential

⁹This can be seen from Bochner’s theorem: a Fourier transformation of a top-hat function yields a sinc function, which is not positive for all real numbers.

¹⁰The Matérn function is a stationary positive definite correlation function containing two parameters: the order ν , which sets the shape of the function, and the range, ϕ . The exponential function is a special case of the Matérn function where $\nu = 0.5$.

kernel with a single scalelength is sufficient to capture the overall behaviour of the ISM. We will explore this assumption by analysing the spatial distribution of other ISM components in future papers.

Thirdly, this analysis does not account for aperture effects. Mannucci et al. (2021) posit that this might be a problem. As the size of our spaxels is often below the size of a H II region, some of our spaxels may only include the inner regions of H II regions, where higher ionization species dominate, and lower ionization species are under-represented. A tradeoff exists between integrating fluxes between H II spaxels to get more accurate metallicities, and losing resolution because of binning. H II region detection software such as `HiPhot` (Thilker, Braun & Walterbos 2000, used by Williams et al. 2022) or `HiExplorer` (Sánchez et al. 2012b) may be used to construct integrated spectra of H II regions; however, these software packages rely on multiple parameters that are tuned by the user, and may convolve multiple nearby poorly resolved H II regions (e.g. Grasha et al. 2022). Since we are trying to understand how the metallicity changes between H II regions that are separated by a small distance, taking this step may limit the ability to measure the shape of the semivariogram at low separation, impacting our abilities to construct geostatistical models. While aperture effects may lead to inaccuracies in metallicities computed from emission-line ratios, we do not believe these effects will be the same for all four metallicity diagnostics tested in this work. Therefore, any conclusions that hold across all diagnostics may still be trusted. Exploring how our results change when integrated H II region properties are used may be the focus of future work.

Finally, this model assumes that the stochastic metallicity fluctuations $\eta(\vec{x})$ are Gaussian, with an exponential correlation function. Of all the assumptions that we make in this work, we regard this assumed correlation structure as the main area for future improvement. From the semivariograms presented in Fig. 2, we can see that this model does a good job of describing the small-scale metallicity fluctuations; but it is not perfect. Below 1 kpc, semivariograms were seen to exhibit power-law like behaviour that is not captured by a simple exponential model, but is predicted from general models of turbulence. High-resolution datacubes such as the ones produced by the TYPHOON/PrISM survey allow testing and verification of detailed models of chemical mixing in the ISM. The question of how the small-scale structure of metals in the ISM of local star-forming galaxies should best be modelled remains open and is an active area of investigation.

8 SUMMARY AND CONCLUSIONS

In M21, the potential of geostatistical analysis was introduced as a tool to (1) understand the small-scale structures of galaxies, (2) compare theoretical models to high-resolution IFU data, (3) constrain models of stellar feedback and metal mixing, and (4) move beyond the metallicity gradient to understand the 2D metallicity distribution of galaxies. Here, we extend upon the initial work, introducing analytical techniques for fitting a geostatistical metallicity model to galaxy data, and using these models to predict the metallicity at unmeasured points using universal kriging. We summarize our key findings below:

(i) For NGC 5236, significant correlations are seen in the metallicity of spaxels separated by less than 0.4–1.2 kpc, with the scale of these correlations dependant on the metallicity diagnostic used. Small-scale metallicity structures are shown to cause deviations in the metallicities of spaxels from a radial metallicity trend with

amplitude ~ 1 order-of-magnitude larger than what is expected from measurement errors alone. This small-scale metallicity structure appears to be fairly well-described by an exponential covariance structure with a scale parameter of $\phi = 133\text{--}400$ pc and a variance of $\sigma^2 = 0.0013\text{--}0.01$; however, we believe a more physically motivated model could be constructed in the future.

(ii) Universal kriging produces better predictions for metallicities at unknown data points than the commonly used linear metallicity gradient model. Small-scale metallicity trends are important for understanding the structure of the ISM of local galaxies and thus quantifying them gives distinct advantages for testing models of turbulent metal mixing in galaxies, and determining the metallicities of specific locations within galaxies, such as the sites of transient events.

(iii) The metallicity of a DIG-dominated region may be estimated using the technique of universal kriging, together with the geostatistical modelling methods presented in this work. We argue that this method is to be preferred over directly applying H II region calibrated metallicity diagnostics.

(iv) The $\text{N}_2\text{S}_2\text{H}\alpha$, N_2O_2 , O_3N_2 , and RS_{32} diagnostics have limited accuracy in determining the metallicities of DIG-dominated regions. Out of the 4 metallicity diagnostics tested, N_2O_2 is the least sensitive to DIG contamination, in concordance with the findings of Zhang et al. (2017).

(v) For O_3N_2 and RS_{32} , the offset between the metallicities inferred from a geostatistical model via kriging and from those estimated using these H II-region calibrated diagnostics is tightly correlated with O_3 .

(vi) Comparisons to predictions from geostatistical models calibrated on H II region data reveal that the O_3N_2 and RS_{32} diagnostics together with the O_3 -dependant DIG correction factors of Kumari et al. (2019) are found to generally exhibit less bias than their uncorrected counterparts. However, we caution that for some galaxies, these corrected metallicity diagnostics still introduce a greater bias than when no correction is used. This suggests that the dominant sources of DIG may be different for different galaxies, again highlighting the benefits of taking a more rigorous geostatistical approach to data analysis.

ACKNOWLEDGEMENTS

The authors thank the anonymous referee for their comments, which helped improve the structure and readability of this paper. We would like to give thanks to Prof. Barry F. Madore (Carnegie Observatories and the University of Chicago) PI of the TYPHOON Programme, to members of that team, especially Dr. Henry Poetrodjojo, Prof. Lisa Kewley, Dr. Mark Seibert, and Dr. Jeff Rich for providing the data and numerous insights; to Dr Alex Cameron for his expertise on BPT diagnostics; and to Dr. Nicha Leethochawalit for their helpful comments in preparing this manuscript. BM acknowledges support from an Australian Government Research Training Program (RTP) Scholarship. This research is supported in part by the Australian Research Council Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), through project number CE170100013.

DATA AVAILABILITY

Any data products presented in this work, as well as the PYTHON code used to construct them, are available from the corresponding author upon reasonable request.

REFERENCES

- Aller L. H., 1984, *Physics of Thermal Gaseous Nebulae*. Springer, Dordrecht
- Andrews B. H., Martini P., 2013, *ApJ*, 765, 140
- Baldwin J. A., Phillips M. M., Terlevich R., 1981, *PASP*, 93, 5
- Beaton R. L. et al., 2016, *ApJ*, 832, 210
- Berg D. A., Skillman E. D., Garnett D. R., Croxall K. V., Marble A. R., Smith J. D., Gordon K., Kennicutt Robert C. J., 2013, *ApJ*, 775, 128
- Bigiel F., Leroy A., Walter F., Brinks E., de Blok W. J. G., Madore B., Thornley M. D., 2008, *AJ*, 136, 2846
- Binette L., Flores-Fajardo N., Raga A. C., Drissen L., Morisset C., 2009, *ApJ*, 695, 552
- Blanc G. A., Heiderman A., Gebhardt K., Evans Neal J. I., Adams J., 2009, *ApJ*, 704, 842
- Blanc G. A., Kewley L., Vogt F. P. A., Dopita M. A., 2015, *ApJ*, 798, 99
- Boettcher E., Gallagher J. S. I., Zweibel E. G., 2017, *ApJ*, 845, 155
- Bresolin F., Kennicutt Robert C. J., 2002, *ApJ*, 572, 838
- Bresolin F., Ryan-Weber E., Kennicutt R. C., Goddard Q., 2009, *ApJ*, 695, 580
- Bresolin F., Kudritzki R.-P., Urbaneja M. A., Gieren W., Ho I. T., Pietrzyński G., 2016, *ApJ*, 830, 64
- Bryant J. J. et al., 2015, *MNRAS*, 447, 2857
- Bundy K. et al., 2015, *ApJ*, 798, 7
- Cappellari M., 2017, *MNRAS*, 466, 798
- Cappellari M., Copin Y., 2003, *MNRAS*, 342, 345
- Cappellari M., Emsellem E., 2004, *PASP*, 116, 138
- Cardelli J. A., Clayton G. C., Mathis J. S., 1989, *ApJ*, 345, 245
- Christensen C. R., Davé R., Brooks A., Quinn T., Shen S., 2018, *ApJ*, 867, 142
- Clark C. J. R. et al., 2019, *MNRAS*, 489, 5256
- Cressie N., 1990, *Math. Geol.*, 22, 239
- Cressie N., 1993, *Statistics for Spatial Data*. John Wiley & Sons, New York
- Curti M., Cresci G., Mannucci F., Marconi A., Maiolino R., Esposito S., 2017, *MNRAS*, 465, 1384
- Curti M., Mannucci F., Cresci G., Maiolino R., 2020a, *MNRAS*, 491, 944
- Curti M. et al., 2020b, *MNRAS*, 492, 821
- Dale D. A. et al., 2009, *ApJ*, 703, 517
- Davies R. L., Rich J. A., Kewley L. J., Dopita M. A., 2014a, *MNRAS*, 439, 3835
- Davies R. L., Kewley L. J., Ho I. T., Dopita M. A., 2014b, *MNRAS*, 444, 3961
- de Avillez M. A., Mac Low M.-M., 2002, *ApJ*, 581, 1047
- De Cia A., Jenkins E. B., Fox A. J., Ledoux C., Ramburuth-Hurt T., Konstantopoulou C., Petitjean P., Krogager J.-K., 2021, *Nature*, 597, 206
- Di Matteo P., Haywood M., Combes F., Semelin B., Snaith O. N., 2013, *A&A*, 553, A102
- Diggle P. J., Riberio P. J., 2007, *Model-based Geostatistics*. Springer, New York
- Dopita M. A., Sutherland R. S., 1995, *ApJ*, 455, 468
- Dopita M. A., Kewley L. J., Heisler C. A., Sutherland R. S., 2000, *ApJ*, 542, 224
- Dopita M. A., Sutherland R. S., Nicholls D. C., Kewley L. J., Vogt F. P. A., 2013, *ApJS*, 208, 10
- Dopita M. A., Kewley L. J., Sutherland R. S., Nicholls D. C., 2016, *Ap&SS*, 361, 61
- Edmunds M. G., Greenhow R. M., 1995, *MNRAS*, 272, 241
- Emsellem E. et al., 2022, *A&A*, 659, A191
- Erroz-Ferrer S. et al., 2019, *MNRAS*, 484, 5009
- Florian M. K. et al., 2021, *ApJ*, 916, 50
- Folatelli G. et al., 2010, *AJ*, 139, 120
- Freeman K., Bland-Hawthorn J., 2002, *ARA&A*, 40, 487
- Gadotti D. A. et al., 2019, *MNRAS*, 482, 506
- Galbany L. et al., 2016, *MNRAS*, 455, 4087
- Galbany L. et al., 2018, *ApJ*, 855, 107
- Gneiting T., 1997, PhD thesis, Univ. Bayreuth
- González-Gaitán S., de Souza R. S., Krone-Martins A., Cameron E., Coelho P., Galbany L., Ishida E. E. O., COIN Collaboration, 2019, *MNRAS*, 482, 3880
- Grand R. J. J. et al., 2016, *MNRAS*, 460, L94
- Grasha K. et al., 2022, *ApJ*, 929, 118
- Groenewegen M. A. T., Romaniello M., Primas F., Mottini M., 2004, *A&A*, 420, 655
- Haffner L. M. et al., 2009, *Rev. Mod. Phys.*, 81, 969
- Heckman T. M., Armus L., Miley G. K., 1990, *ApJS*, 74, 833
- Ho L. C., Li Z.-Y., Barth A. J., Seigar M. S., Peng C. Y., 2011, *ApJS*, 197, 21
- Ho I.-T. et al., 2016, *Ap&SS*, 361, 280
- Ho I.-T. et al., 2017, *ApJ*, 846, 39
- Hogg D. W., Bovy J., Lang D., 2010, preprint ([arXiv:1008.4686](https://arxiv.org/abs/1008.4686))
- Hoopes C. G., Walterbos R. A. M., 2003, *ApJ*, 586, 902
- Hopkins P. F., Kereš D., Oñorbe J., Faucher-Giguère C.-A., Quataert E., Murray N., Bullock J. S., 2014, *MNRAS*, 445, 581
- Huijbregts G., Matheron G., 1971, *Canada Inst. Min. Metall.*, 12, 185
- James G., Witten D., Hastie T., Tibshirani R., 2013, *An Introduction to Statistical Learning with Applications in R*. Springer, New York
- Jeffreson S. M. R., Kruijssen J. M. D., Keller B. W., Chevance M., Glover S. C. O., 2020, *MNRAS*, 498, 385
- Kaplan K. F. et al., 2016, *MNRAS*, 462, 1642
- Kauffmann G. et al., 2003, *MNRAS*, 346, 1055
- Kennicutt Robert C. J., 1998, *ApJ*, 498, 541
- Kewley L. J., Dopita M. A., 2002, *ApJS*, 142, 35
- Kewley L. J., Heisler C. A., Dopita M. A., Lumsden S., 2001, *ApJS*, 132, 37
- Kewley L. J., Nicholls D. C., Sutherland R. S., 2019, *ARA&A*, 57, 511
- Kim C.-G. et al., 2020, *ApJ*, 900, 61
- Kollmeier J. A. et al., 2017, preprint ([arXiv:1711.03234](https://arxiv.org/abs/1711.03234))
- Kreckel K. et al., 2020, *MNRAS*, 499, 193
- Krumholz M. R., Ting Y.-S., 2018, *MNRAS*, 475, 2236
- Kumari N., Maiolino R., Belfiore F., Curti M., 2019, *MNRAS*, 485, 367
- Lacerda E. A. D. et al., 2018, *MNRAS*, 474, 3727
- Lee D., Gammie C. F., 2021, *ApJ*, 906, 39
- Leroy A. K. et al., 2019, *ApJS*, 244, 24
- Madsen G. J., Reynolds R. J., Haffner L. M., 2006, *ApJ*, 652, 401
- Maiolino R., Mannucci F., 2019, *A&AR*, 27, 3
- Mannucci F. et al., 2021, *MNRAS*, 508, 1582
- Mardia K. V., Marshall R. J., 1984, *Biometrika*, 71, 135
- Matérn B., 1960, *Meddelanden från Statens Skogsforskningsinstitut*, 49, 1
- Matheron G., 1969, *Cahiers du Centre de Morphologie Mathématique de Fontainebleau*, Vol. 1, École nationale supérieure des Mines de Paris, Paris
- Metha B., Trenti M., Chu T., 2021, *MNRAS*, 508, 489 (M21)
- Moran S. M. et al., 2012, *ApJ*, 745, 66
- Naab T., Ostriker J. P., 2017, *ARA&A*, 55, 59
- Nemec J. M., Nemec A. F. L., Lutz T. E., 1994, *AJ*, 108, 222
- Oey M. S. et al., 2007, *ApJ*, 661, 801
- Pan L., Scannapieco E., 2010, *ApJ*, 721, 1765
- Persson S. E., Carr D. M., Jacobs J. H., 1990, *Exp. Astron.*, 1, 195
- Pettini M., Pagel B. E. J., 2004, *MNRAS*, 348, L59
- Pilyugin L. S., Grebel E. K., 2016, *MNRAS*, 457, 3678
- Pilyugin L. S., Thuan T. X., 2005, *ApJ*, 631, 231
- Poetrodjojo H., D'Agostino J. J., Groves B., Kewley L., Ho I.-T., Rich J., Madore B. F., Seibert M., 2019, *MNRAS*, 487, 79
- Poggianti B. M. et al., 2017, *ApJ*, 844, 48
- Rasmussen C. E., Williams C. K. I., 2006, *Gaussian Processes for Machine Learning*. MIT Press, Cambridge
- Renzini A., Peng Y.-J., 2015, *ApJ*, 801, L29
- Reynolds R. J., 1985, *ApJ*, 294, 256
- Reynolds R. J., Hausen N. R., Tufté S. L., Haffner L. M., 1998, *ApJ*, 494, L99
- Rosales-Ortega F. F., Sánchez S. F., Iglesias-Páramo J., Díaz A. I., Vílchez J. M., Bland-Hawthorn J., Husemann B., Mast D., 2012, *ApJ*, 756, L31
- Rousseau-Nepton L. et al., 2019, *MNRAS*, 489, 5530
- Rue H., Riebler A., Sørbye S. H., Illian J. B., Simpson D. P., Lindgren F. K., 2017, *Ann. Rev. Stat. Appl.*, 4, 395
- Sale S. E., Magorrian J., 2014, *MNRAS*, 445, 256
- Sánchez Almeida J., Elmegreen B. G., Muñoz-Tuñón C., Elmegreen D. M., 2014, *A&AR*, 22, 71

- Sánchez S. F., 2020, *ARA&A*, 58, 99
 Sánchez S. F. et al., 2012a, *A&A*, 538, A8
 Sánchez S. F. et al., 2012b, *A&A*, 546, A2
 Sánchez S. F. et al., 2013, *A&A*, 554, A58
 Scalo J., Elmegreen B. G., 2004, *ARA&A*, 42, 275
 Schmidt M., 1959, *ApJ*, 129, 243
 Simons R. C. et al., 2021, *ApJ*, 923, 203
 Stasińska G. et al., 2008, *MNRAS*, 391, L29
 Sun J. et al., 2020, *ApJ*, 901, L8
 Sutherland R., Dopita M., Binette L., Groves B., 2018, Astrophysics Source Code Library, record ascl:1807.005
 Thilker D. A., Braun R., Walterbos R. A. M., 2000, *AJ*, 120, 3070
 Tremonti C. A. et al., 2004, *ApJ*, 613, 898
 Tully R. B., Fisher J. R., 1988, *Catalog of Nearby Galaxies*. Cambridge Univ. Press, Cambridge
 Tully R. B., Shaya E. J., Karachentsev I. D., Courtois H. M., Kocevski D. D., Rizzi L., Peel A., 2008, *ApJ*, 676, 184
 Tully R. B. et al., 2013, *AJ*, 146, 86
 Tully R. B., Courtois H. M., Sorce J. G., 2016, *AJ*, 152, 50
 Utreras J. et al., 2020, *ApJ*, 892, 94
 Vale Asari N., Couto G. S., Cid Fernandes R., Stasińska G., de Amorim A. L., Ruschel-Dutra D., Werle A., Florido T. Z., 2019, *MNRAS*, 489, 4721
 Vazdekis A., Ricciardelli E., Cenarro A. J., Rivero-González J. G., Díaz-García L. A., Falcón-Barroso J., 2012, *MNRAS*, 424, 157
 Veilleux S., Osterbrock D. E., 1987, *ApJS*, 63, 295
 Vogt F. P. A., Pérez E., Dopita M. A., Verdes-Montenegro L., Borthakur S., 2017, *A&A*, 601, A61
 Wikle C. K., Zammit-Mangion A., Cressie N., 2019, *Spatio-Temporal Statistics with R*. CRC Press, Taylor & Francis Group, Boca Raton
 Williams T. G. et al., 2022, *MNRAS*, 509, 1303
 Wolfire M. G., McKee C. F., Hollenbach D., Tielens A. G. G. M., 2003, *ApJ*, 587, 278
 Wuyts S. et al., 2013, *ApJ*, 779, 135
 Zhang K. et al., 2017, *MNRAS*, 466, 3217
 Zurita A., Rozas M., Beckman J. E., 2000, *A&A*, 363, 9
 Zurita A., Beckman J. E., Rozas M., Ryder S., 2002, *A&A*, 386, 801

SUPPORTING INFORMATION

Supplementary data are available at [MNRAS](https://academic.oup.com/mnras/article/514/3/4465/6603854) online.

Supplementary Figures.zip

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

APPENDIX A: COMMENTS ON OTHER TYPHOON GALAXIES

In this paper, we largely restrict our discussion to NGC 5236, using the DIG-diagnostic procedure of Kaplan et al. (2016). To ensure our results are valid for a wider variety of galaxies, we repeat our analysis for a selection of seven other galaxies for which high-resolution metallicity data has been obtained by the TYPHOON/PrISM survey. Properties of this extended sample of galaxies are listed in Table 1.

Broadly, conclusions found for NGC 5236 hold for the other seven galaxies in the sample, using any of the DIG isolation methods discussed in Appendix B2 [except for the effect of applying O3-based DIG correction factors to the empirically calibrated metallicity diagnostics of Curti et al. (2020a); see Section 6.1]. Semivariograms constructed for all galaxies reveal significant structures in the metallicity distributions, with correlations on spatial scales of ~ 1 kpc. For all galaxies, the size of these correlations was much larger (~ 1 order of magnitude) than the uncorrelated measurement error. In all cases, the method of universal kriging as described in the text was verified to produce more accurate predictions of the metallicities of unknown regions than a simple metallicity-gradient model using 10-fold cross-validation (Metha et al., in preparation). Fig. A1 shows the semivariograms for all galaxies in our sample for the RS₃₂ diagnostic to illustrate that in all cases a consistent correlation length is inferred. Full plots analogous to those presented in Figs 2–3 for all galaxies with all combinations of DIG/metallicity diagnostics and kriged maps showing the metallicity distribution of each galaxy will be released in a forthcoming paper (Metha et al., in preparation).

One notable exception was found when the geostatistical model-fitting method described in Section 4.1 was applied to N₂O₂ based metallicity maps for HII regions selected using either the N2-BPT or S2-BPT methods for the 7 galaxies in this sample other than NGC 5236. When this was done, the amount of uncorrelated variance revealed by a semivariogram analysis was much greater than the amount predicted by our geostatistical model. We recall that our model explains all metallicity variance within a galaxy using a combination of three factors: a large-scale metallicity trend (μ , modelled here as a radial metallicity gradient), small-scale metallicity fluctuations (η), and measurement errors stemming from uncertainties in the emission-line fluxes (ϵ , computed using equation B1). Only the measurement error term, ϵ , is capable of changing the semivariance at zero separation, implying that the failure in our model to match the data must lie in an underestimation of this term. We further observe that the semivariograms computed using the N₂S₂H α and O₃N₂ diagnostics show good agreement between the modelled values of ϵ and the revealed semivariogram, implying that the problem must not lie in the error reporting associated with the [N II] emission line. Therefore, we conclude that the failure of our geostatistical model to correctly model the uncorrelated variance of N₂O₂ is possibly due to under-reporting of the error associated with the [O II] emission line. This issue will be corrected before kriged metallicity maps for the full TYPHOON sample are released to the public (Metha et al., in preparation).

This analysis highlights the power of the semivariogram analysis in being able to separate small-scale metallicity variations from measurement errors. While we focus in this study, and in M21, on using this analysis to understand the small-scale structure of metallicity fields after isolating and subtracting the effects of measurement error, these methods can also be used to understand errors associated with metallicity measurement without the interfering effects of true spatial variation of the data. This may be useful for validating data pipelines, and testing assumptions about the size and correlation structure of errors between spaxels in IFS data.

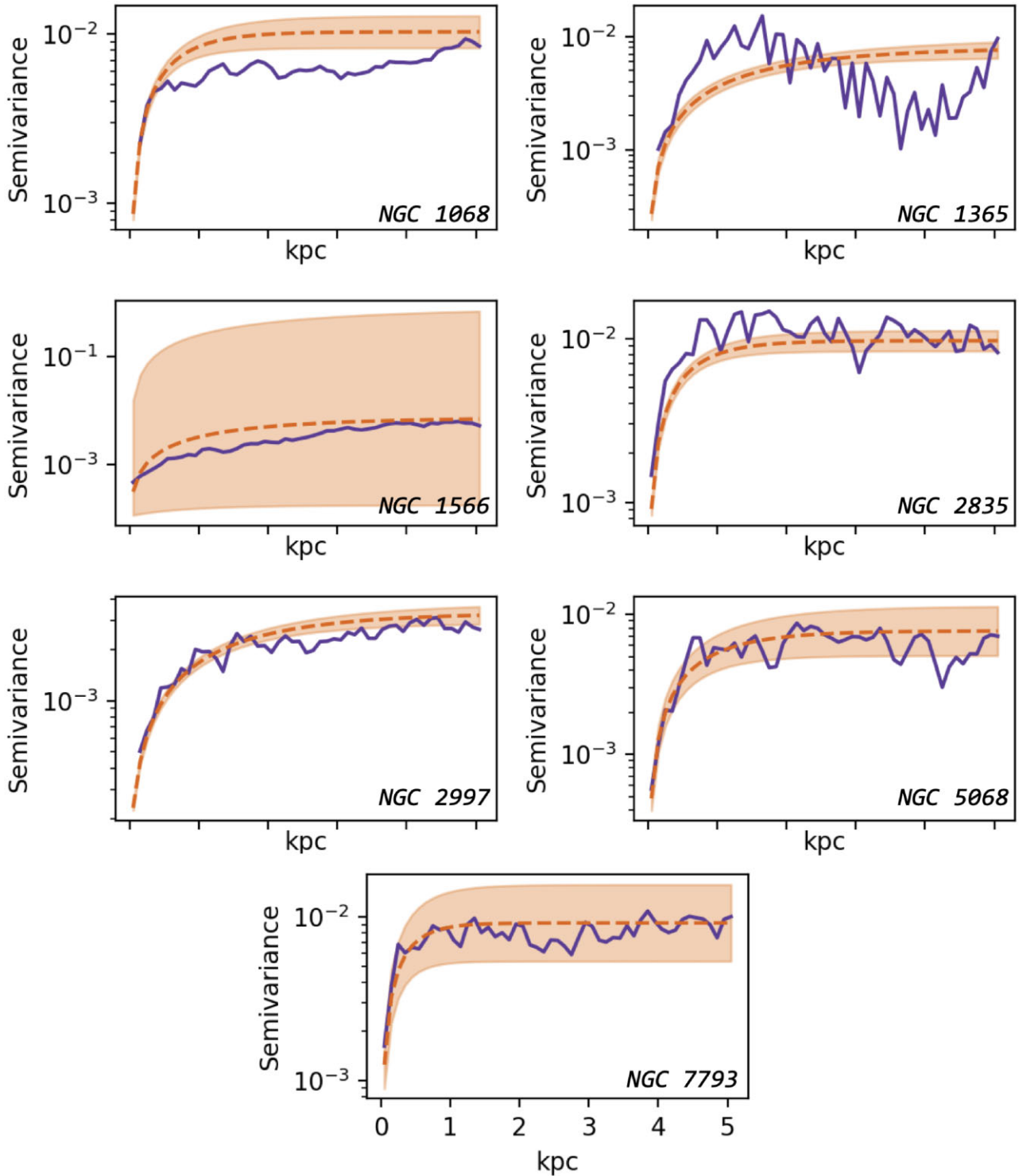


Figure A1. Semivariograms showing the small-scale metallicity structure of all other galaxies investigated in this work, using the RS_{32} diagnostic. Purple solid lines show the empirical semivariogram. Orange dashed lines show the semivariogram predicted from the model of best fit. Orange shaded regions show the $1 - \sigma$ error in the fitted curve (errors on the semivariance estimation from the data at each separation shown are negligible). For NGC 1365, an exponential model does not fit the data (likely a simple metallicity gradient does not subtract off all of the large-scale trends), and for NGC 1566, the uncertainties in the best-fitting model are very large. In all other cases, we see that a simple exponential model does a fair job of fitting the data, and correlations are present between data points separated by ~ 1 kpc.

APPENDIX B: STRONG EMISSION LINE DIAGNOSTICS

Many different diagnostics exist that can be used to estimate metallicities from ratios of strong-emission lines in H II regions (e.g. Kewley et al. 2019; Maiolino & Mannucci 2019). Similarly, many different diagnostics are currently being developed in order to distinguish regions of galaxies dominated by DIG emission from H II regions, such as using the surface brightness of H α emission (Reynolds et al. 1998; Zhang et al. 2017), the equivalent width of the H α emission line (Lacerda et al. 2018; Vale Asari et al. 2019), BPT diagrams (Baldwin, Phillips & Terlevich 1981; Kumari et al. 2019), and the [S II]/H α line ratio (Reynolds 1985; Blanc et al. 2009; Kaplan et al. 2016; Poetrodjojo et al. 2019). In this section, we will describe the four metallicity diagnostics and three DIG diagnostics used in this work to distinguish H II regions from DIG-dominated regions and identify the metallicities of H II regions, and how these diagnostics are calibrated.

B1 Metallicity diagnostics

In this study, we consider four different metallicity diagnostics, two of which have been calibrated theoretically using photoionization models, and two of which are empirically calibrated against metallicities determined using the direct (T_e -based) method.

The primary metallicity diagnostic we consider is the N₂S₂H α diagnostic, devised by Dopita et al. (2016), theoretically calibrated using the MAPPINGS 5.0 code (Sutherland et al. 2018). The three emission lines used in this diagnostic ([N II] λ 6583, [S II] λ λ 6717,6731, and H α) are all very close to each other in wavelength, reducing the sensitivity of this diagnostic to dust extinction. Furthermore, this diagnostic shows little dependence on the ionization parameter.

As a secondary theoretically calibrated metallicity diagnostic, we use the N₂O₂ diagnostic, based on the line ratio N₂O₂ = log([N II] λ 6583/[O II] λ λ 3726, 29). Zhang et al. (2017) found that metallicities calculated using this diagnostic showed very little dependence on DIG contamination. To match this study, we use the theoretical calibration of Dopita et al. (2013), which was computed using the astrophysical plasma modelling code MAPPINGS IV. An advantage of using this line ratio is that it is insensitive to variations in the ionization parameter (Dopita et al. 2000). One disadvantage of this diagnostic for the specific data set we used is that the du Pont Telescope at the Las Campanas Observatory is not very sensitive at the wavelength of [O II]. For this reason, errors associated with this metallicity diagnostic are much larger than those found with any of the other metallicity diagnostics tested.

Additionally, we consider two metallicity diagnostics with empirical calibrations; one based on the O₃N₂ line ratio (=log([O III] λ 5007/H α) – log([N II] λ 6583/H β)) (Curti et al. 2017), and a similar calibration based on RS₃₂ = log([O III] λ 5007/H β + [S II] λ λ 6717, 6731/H α) (Curti et al. 2020a). These calibrations were constructed by fitting polynomials up to 5th order between the line ratios and metallicities computed using the T_e method for stacked spectra of galaxies observed by the *Sloan Digital Sky Survey*, and each relationship show scatters of the order of $\sigma = 0.1$ dex.

One limitation of the O₃N₂ and RS₃₂ diagnostics is that they have only been calibrated for a range of metallicities, from log(O/H) + 12 = 7.6 to 8.9. A negligible number of spaxels with a metallicity larger than this value are discarded (13 out of 2472 H II spaxels were discarded for the O₃N₂ diagnostic, and 36 out of 2518 for RS₃₂). This cutoff can be seen in Fig. 1, especially for the O₃N₂ diagnostic. It is important to note this as a caveat; however, the

small number of high-metallicity spaxels is unlikely to affect our results, and developing a correction term to extend the range of these diagnostics to larger metallicities in order to handle these few spaxels is beyond the scope of this study. As an additional caveat, the RS₃₂ diagnostic is double-valued for metallicities below log(O/H) + 12 = 8. However, the spiral galaxies considered in this study are all large, local spirals, with metallicities far greater than this limiting value for all spaxels. Hence, this problem does not require further consideration.

For each of these diagnostics, we compute the error in the deduced metallicity of each spaxel from the uncertainty in each line flux I_i using linear error propagation, and assuming the error in all line fluxes I_i are independent:

$$\sigma_Z^2 = \sum \left(\frac{\partial Z}{\partial I_i} \sigma_{I_i} \right)^2. \quad (\text{B1})$$

B2 DIG diagnostics

B2.1 [S II]/H α

Our primary diagnostic for separating DIG-dominated regions from star-forming spaxels uses the [S II]/H α line ratio, and the methodology of Kaplan et al. (2016). Madsen, Reynolds & Haffner (2006) found that this line ratio is higher in DIG regions. Using this information, Blanc et al. (2009) created an equation that relates $C_{\text{H II}}$, the fraction of H α emission originating from H II regions, from the [S II]/H α line ratio of each spaxel:

$$\frac{[\text{S II}]}{\text{H}\alpha} = C_{\text{H II}} \left(\frac{[\text{S II}]}{\text{H}\alpha} \right)_{\text{H II}} + (1 - C_{\text{H II}}) \left(\frac{[\text{S II}]}{\text{H}\alpha} \right)_{\text{DIG}}. \quad (\text{B2})$$

To properly estimate $C_{\text{H II}}$ for each spaxel using rigorous statistical methods, three things would need to be known: (i) the distribution of the [S II]/H α line ratio for DIG regions within each galaxy; (ii) the distribution of the [S II]/H α line ratio for H II regions with no DIG contamination; and (iii) the prior expected distribution of $C_{\text{H II}}$ for each spaxel. In reality, none of these three distributions are known. Therefore, some assumptions must be made.

The method of Kaplan et al. (2016) makes two main assumptions: first, that the intrinsic distributions of [S II]/H α are narrow, with little intrinsic scatter within each galaxy; and secondly, that the spaxels with the brightest H α flux in each galaxy have zero DIG contamination ($C_{\text{H II}} = 1$), while the faintest spaxels are completely contaminated by DIG ($C_{\text{H II}} = 0$). Under these two assumptions, $\left(\frac{[\text{S II}]}{\text{H}\alpha} \right)_{\text{H II}}$ and $\left(\frac{[\text{S II}]}{\text{H}\alpha} \right)_{\text{DIG}}$ are taken to be the median value of [S II]/H α for the 100 spaxels with the brightest and faintest flux of H α , respectively, for each galaxy. Following Poetrodjojo et al. (2019), we take any spaxels with $C_{\text{H II}} > 0.9$ to be H II spaxels, and classify all other spaxels as DIG-dominated. For NGC 5236, this leads to 3664 spaxels being classified as H II spaxels, and 19607 being identified as having non-negligible DIG contributions.

Unlike $\Sigma_{\text{H}\alpha}$, [S II]/H α is an intensive quantity, and does not suffer the conceptual limitations associated with line-of-sight projections detailed in Lacerda et al. (2018). But it also agrees with the $\Sigma_{\text{H}\alpha}$ diagnostic, making it a reliable intrinsic proxy that is consistent with the $\Sigma_{\text{H}\alpha}$ method (Blanc et al. 2009; Kaplan et al. 2016; Poetrodjojo et al. 2019). A drawback of this method is that the decision to use 100 spaxels to classify the reference ratio values for DIG/H II regions is arbitrary. In the galaxies in our sample, this corresponds to 0.4 – 3.8 per cent of spaxels. We leave analysis on how this hyperparameter affects the realised DIG/H II distributions to dedicated studies of DIG classification.

B2.2 BPT diagnostics

As an additional measure, we also classify spaxels as H II/DIG-dominated using the classical emission line ratio diagnostic diagrams, commonly known as BPT diagrams (Baldwin et al. 1981; Veilleux & Osterbrock 1987). This methodology is also used by Kumari et al. (2019) when constructing DIG-corrected versions of the O_3N_2/RS_{32} metallicity diagnostics. For consistency with this study, we adopt the same diagnostics to distinguish H II regions from DIG used by Kumari et al. (2019), using the demarcation line published by Kauffmann et al. (2003) on the $[O III]/H\beta$ – $[N II]/H\alpha$ diagram, and the Kewley et al. (2001) demarcation line on the $[O III]/H\beta$ – $[S II]/H\alpha$ diagram. These diagnostics are some of the oldest, most well tested and trusted ways in which ionization from H II regions is distinguished from other sources of ionization and show quantitative agreement with other diagnostics (Lacerda et al. 2018). While they are traditionally used to classify galaxies

and not regions within galaxies, these diagnostics are applicable to all spaxels in IFU datacubes with sufficient S/N (see e.g. Davies et al. 2014a,b; Kewley et al. 2019; Sánchez 2020).

We note that these different diagnostics predict a different number of H II spaxels and DIG regions for each galaxy. For NGC 5236, using the N2-based BPT diagnostic, 7232 spaxels are identified as H II spaxels, and 2008 are classified as being DIG-dominated; however, when the S2-based BPT diagnostic is used, 8819 spaxels are identified as H II spaxels, and only 307 are classified as DIG-dominated. For this reason, we caution the reader that the classification of each spaxel as being H II/DIG dominated depends on the DIG diagnostic used. To overcome this dependence, analogous Figures to those presented in this work are shown for all DIG diagnostics in the supplementary material (available online).

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.