Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **499,** 193–209 (2020)
Advance Access publication 2020 September 15

doi:10.1093/mnras/staa2743

Measuring the mixing scale of the ISM within nearby spiral galaxies

Kathryn Kreckel [1]★ I.-Ting Ho [2] Guillermo A. Blanc,[3,4] Simon C. O. Glover,[5] Brent Groves,[6] Erik Rosolowsky [7] Frank Bigiel,[8] Médéric Boquíen,[9] Mélanie Chevance,[1] Daniel A. Dale,[10] Sinan Deger,[11] Eric Emsellem,[12,13] Kathryn Grasha [14] Jenny J. Kim [1] Ralf S. Klessen,[5,15] J. M. Diederik Kruijssen [1] Janice C. Lee,[11] Adam K. Leroy,[16] Daizhong Liu,[2] Rebecca McElroy,[17] Sharon E. Meidt,[18] Ismael Pessa,[2] Patricia Sanchez-Blazquez,[19] Karin Sandstrom,[20] Francesco Santoro,[2] Fabian Scheuermann,[1] Eva Schinnerer,[2] Andreas Schruba,[21] Dyas Utomo,[16] Elizabeth J. Watkins[1] and Thomas G. Williams [2]

Affiliations are listed at the end of the paper

Accepted 2020 September 4. Received 2020 September 4; in original form 2020 July 8

ABSTRACT

The spatial distribution of metals reflects, and can be used to constrain, the processes of chemical enrichment and mixing. Using PHANGS-MUSE optical integral field spectroscopy, we measure the gas-phase oxygen abundances (metallicities) across 7138 H II regions in a sample of eight nearby disc galaxies. In Paper I, we measure and report linear radial gradients in the metallicities of each galaxy, and qualitatively searched for azimuthal abundance variations. Here, we examine the 2D variation in abundances once the radial gradient is subtracted, Δ(O/H), in order to quantify the homogeneity of the metal distribution and to measure the mixing scale over which H II region metallicities are correlated. We observe low (0.03–0.05 dex) scatter in Δ(O/H) globally in all galaxies, with significantly lower (0.02–0.03 dex) scatter on small (<600 pc) spatial scales. This is consistent with the measurement uncertainties, and implies the 2D metallicity distribution is highly correlated on scales of $\lesssim$600 pc. We compute the two-point correlation function for metals in the disc in order to quantify the scale lengths associated with the observed homogeneity. This mixing scale is observed to correlate better with the local gas velocity dispersion (of both cold and ionized gas) than with the star formation rate. Selecting only H II regions with enhanced abundances relative to a linear radial gradient, we do not observe increased homogeneity on small scales. This suggests that the observed homogeneity is driven by the mixing introducing material from large scales rather than by pollution from recent and on-going star formation.

Key words: ISM: abundances – ISM: evolution – galaxies: abundances – galaxies: ISM.

1 INTRODUCTION

Dynamical conditions within the interstellar medium (ISM) play an important role in regulating galaxy evolution. The cloud-scale turbulent state (Larson 1981) combined with the surrounding ISM pressure (Hughes et al. 2013; Schruba, Kruijssen & Leroy 2019; Sun et al. 2020) and large-scale dynamical processes (Colombo et al. 2014; Jeffreson & Kruijssen 2018; Meidt et al. 2018, 2020) set the conditions for cloud stability and cloud collapse, regulating future star formation. Cloud collapse is additionally regulated by the physical and chemical conditions in the ISM, as the presence of heavy elements facilitates gas cooling (Dopita & Sutherland 2003; Smith et al. 2017; Klessen & Glover 2016). As a result, understanding the homogeneity (or inhomogeneity) in the gas-phase abundances (metallicity) is important for informing our understanding of the star formation process. The distribution of elements in the gas phase has further implications for star formation history studies based on observed stellar metallicities, such as Galactic archaeology and (globular) cluster formation, which pre-suppose a certain level of homogeneity (see review in Krumholz, McKee & Bland -Hawthorn 2019).

Many processes are expected to drive mixing of metals within the ISM, increasing the homogeneity. This includes bar-driven radial mixing (Di Matteo et al. 2013), spiral-arm-driven large-scale systematic streaming motions (Grand et al. 2016; Sánchez-Menguiano et al. 2016), kiloparsec-scale mixing-induced dilution due to the spiral density waves passage (Ho et al. 2017), thermal and gravitational instabilities (Yang & Krumholz 2012; Petit et al. 2015), and interstellar turbulence (de Avillez & Mac Low 2002; Klessen & Lin 2003; Krumholz & Ting 2018). Many of these processes derive from non-axisymmetric features like bars and spiral arms, and are expected to preserve and/or introduce azimuthal variations in the metallicity distribution. To observationally distinguish these models, we need to quantify the level of homogeneity, the spatial scales over which mixing is effective, and how it relates to large-scale dynamics and environments.

H II regions, photoionized by the most massive OB stars during their short (<10 Myr) lifetimes, exhibit line emission that probes the instantaneous local ISM abundances. With a sufficiently large

★ E-mail: kathryn.kreckel@uni-heidelberg.de

Table 1. Key parameters of our galaxy sample, adopted from the z0mgs (z = 0 multiwavelength galaxy survey, Leroy et al. 2019) data base (https://github.com/akleroy/z0mgs), where the original references can be found.

Name	D (Mpc)	Type	$\log_{10}\left(\frac{M_*}{M_\odot}\right)$	$\log_{10}\left(\frac{SFR}{M_\odot yr^{-1}}\right)$	PA (deg)	i (deg)	Physical resolution (pc)	Number of H II regions
NGC 0628	9.77 ± 0.82	Sc	10.2	0.26	20.9	8.7	47	1277
NGC 1087	14.4 ± 4.8	SBc	9.8	0.05	177.3	41.3	70	679
NGC 1672	11.9 ± 4.0	SBb (S)	10.2	0.48	135.7	37.5	46	880
NGC 2835	10.1 ± 3.4	SBc	9.6	− 0.08	1.6	47.8	44	699
NGC 3627	10.6 ± 0.9	SBb (S3)	10.6	0.55	174.1	55.0	51	692
NGC 4254	16.8 ± 5.6	Sc	10.5	0.74	67.7	37.8	57	1824
NGC 4535	15.8 ± 2.3	SBc	10.4	0.35	179.8	40.7	38	864
IC 5332	9.95 ± 3.3	SBc	9.6	− 0.31	74.7	24.0	39	223

sample of H II regions, it is possible to use these as tracers of the underlying 2D metallicity field, mapping the effectiveness of mixing in the ISM across a galaxy disc. Until the last decade, H II region metallicity studies were limited to samples of tens of regions per galaxy (Pilyugin, Grebel & Kniazev 2014), sufficient only for measuring the radial metallicity gradient. Long-slit multi-object spectroscopy has achieved samples of ∼100 H II regions per galaxy (Berg et al. 2015; Croxall et al. 2015, 2016; Berg et al. 2020), however in recent years, optical integral field unit (IFU) spectroscopic techniques have increased these numbers to cover 1000s of H II regions within individual galaxies (Rosales-Ortega et al. 2011; Sánchez-Menguiano et al. 2016; Ho et al. 2017), achieving the necessary statistics to investigate azimuthal metallicity variations.

Mapping both the radial and azimuthal variations in metallicity across a sample of galaxies is a key science goal of the PHANGS[1] collaboration. With our PHANGS-MUSE large observing program (PI: Schinnerer; 1100.B-0651), we use the Very Large Telescope/Multi Unit Spectroscopic Explorer (VLT/MUSE) to mosaic the central 4–8 kpc disc of 19 galaxies with optical IFU observations. Preliminary results from the first eight completed galaxies in the PHANGS-MUSE survey (Kreckel et al. 2019, hereafter Paper I) reveal that half exhibit significant azimuthal variations, and suggest that spiral arms play a role in organizing and mixing the ISM. These statistics have been recently confirmed with a larger sample of 45 disc galaxies (Sánchez-Menguiano et al. 2020).

Building on Paper I, we use our sample of 7138 H II regions across the first eight PHANGS-MUSE galaxies (Section 2) to quantify the chemical homogeneity of the ISM (Section 3) and measure the characteristic physical scale below which the gas-phase ISM is chemically homogeneous (Section 4). In Section 5, we compare these results with theoretical models, and relate them to the spiral structure of the host galaxy, before concluding in Section 6.

2 DATA

For this work, we analyse results from the first eight galaxies observed with the MUSE spectrograph (Bacon et al. 2010) at the VLT as part of the PHANGS-MUSE survey (Table 1; Paper I). In general, PHANGS galaxies have been selected to be nearby ($D < 17$ Mpc), low inclination ($i < 60°$) main-sequence disc galaxies. These first eight galaxies span a range in stellar mass of only one order of magnitude, such that they all have approximately solar or slightly subsolar metallicities.

[1]Physics at High Angular resolution in Nearby GalaxieS; http://www.phangs.org

All galaxies are at sufficiently small distances such that the physical resolution is better than 70 pc given the ∼1 arcsec point spread function sizes. This sets the minimum H II region separation length to which we are sensitive. Median H II region separations within our catalogues range from 100–150 pc, and median H II region diameters are ∼80 pc (marginally resolved). All physical size scales throughout this paper are calculated based on deprojected separations, assuming the galaxy distance (D), position angle (PA), and inclination (i) listed in Table 1.

Images of all targets and associated H II region catalogues, along with details on the observing setup, data reduction, techniques for selecting H II regions, and prescriptions applied to infer the gas-phase oxygen abundances are described in Paper I, and only briefly summarized here.

2.1 H II region metallicities

Using the Balmer decrement, assuming case B recombination and $T_e = 10^4$ K, we apply the Fitzpatrick (1999) extinction law assuming a value of $R_V = 3.1$ in order to determine de-reddened line flux measurements for all strong emission lines (Hβ, [O III] λ4959,5007, [N II]λ6548,6583, Hα, [S II]λ6716,6731). All metallicity calculations in this paper adopt the empirical Pilyugin & Grebel (2016) S-calibration (Scal). It relies on the following three standard diagnostic line ratios:

$$N_2 = I_{[\mathrm{N\,II}]\lambda6548+\lambda6584}/I_{\mathrm{H}\beta},$$
$$S_2 = I_{[\mathrm{S\,II}]\lambda6717+\lambda6731}/I_{\mathrm{H}\beta},$$
$$R_3 = I_{[\mathrm{O\,III}]\lambda4959+\lambda5007}/I_{\mathrm{H}\beta}.$$

Here, we have measured only the stronger line in both the [O III] and [N II] doublets and assume a fixed ratio of 3:1 (Storey & Zeippen 2000). The Scal prescription is defined separately over the upper and lower branches in logN_2, with almost all H II regions in our sample falling on the upper branch (log$N_2 \geq -0.6$), where

$$12 + \log(\mathrm{O/H}) = 8.424 + 0.030\ \log(R_3/S_2) + 0.751\ \log N_2 + (-0.349 + 0.182\ \log(R_3/S_2) + 0.508 \log N_2) \times \log S_2. \quad (1)$$

For the lower branch (log$N_2 \leq -0.6$), the oxygen abundance is instead calculated as

$$12 + \log(\mathrm{O/H}) = 8.072 + 0.789\ \log(R_3/S_2) + 0.726\ \log N_2 + (1.069 - 0.170\ \log(R_3/S_2) + 0.022 \log N_2) \times \log S_2. \quad (2)$$

In Appendix A, we repeat our calculations using other common metallicity calibrations. These yield qualitatively similar results to our preferred Scal method. Ho et al. (2019) used auroral line detections to directly measure abundance variations for one galaxy in our sample, NGC 1672, across H II regions in one spiral arm. Those authors find very good qualitative agreement with the trends identified using the Scal strong-line method, increasing our confidence in the use of this calibration, particularly in relation to relative metallicity trends.

Typical uncertainties in the measured metallicities, based on propagation of systematic errors in the data reduction and random errors associated with the line flux measurement, are 0.02 dex. Because our analysis compares variations in metallicity, constant offsets due to calibration uncertainties are effectively removed from the analysis.

In our results, we refer to the metallicity variation once the radial gradient is subtracted as Δ(O/H) (see Paper I). Subtracting a median radial gradient instead of a linear fit does not change the results presented here. We refer to the 1σ scatter observed in Δ(O/H) as σ(O/H). Both Δ(O/H) and σ(O/H) are reported in logarithmic abundance units (e.g. log(O/H)). Fig. 1 shows the Hα map and selected H II regions within NGC 4254, colour coded by Δ(O/H) as an example.

2.2 Ancillary data

In attempting to understand the physical mechanisms affecting our local metallicity variations, we compare our results to a variety of physical conditions inferred from ancillary data. As many parameters vary radially, and disc rotation impacts the azimuthal mixing of the ISM, we choose to do this in radial bins. For our analysis, we exclude the central 1 kpc and break down the remaining area of our galaxies into 2 kpc wide annuli. This size scale is chosen to ensure a sufficiently large sample of H II regions ($\sim$100) in most annuli, and is larger than the poorest angular resolution available in our ancillary data (H I maps at 25 arcsec). Results presented here are not sensitive to the width of the annuli.

A natural potential source of turbulence and mixing is the mechanical and chemical feedback from the star formation processes itself within any given annulus. To calculate the star formation rate (SFR) and SFR surface density ($\Sigma_{\rm SFR}$), we sum the extinction-corrected H II region Hα luminosities, L(Hα), across the annulus. We then convert to SFR following Murphy et al. (2011) as

$$\mathrm{SFR}(\mathrm{M}_\odot\ \mathrm{yr}^{-1}) = 5.37 \times 10^{-42} L(\mathrm{H}\alpha)\ (\mathrm{erg\ s}^{-1}) \quad (3)$$

which assumes a constant SFR over 100 Myr and a fully sampled Kroupa IMF (initial mass function, Kroupa 2001). This should be a reasonable approximation for our 2-kpc wide rings, which each cover a large part of the galaxy. Our results do not change qualitatively if we include the diffuse Hα emission within the annulus in this analysis.

The orbital time within any given annulus is determined from high-resolution rotation curve fits to the molecular gas kinematics, as traced by PHANGS-ALMA observations of CO (2$-$1) emission (Lang et al. 2020). IC 5332 is excluded from this analysis, as it contains no detected pixel above the necessary signal-to-noise (S/N) cut. As the molecular gas is largely confined to a thin mid-plane disc, we expect the derived rotation curve to serve as a reasonable tracer of the orbital time experienced by the H II regions.

Accounting for the molecular and atomic gas relative to the stellar mass, we measure the gas fraction in each annulus. All three tracers are currently available for only six of our eight galaxies, as IC 5332 and NGC 1672 currently have no H I observations available and are thus excluded. The total gas fraction f_g is then calculated as

$$f_g = \frac{\Sigma(M_{\rm H_2}) + \Sigma(M_{\rm HI})}{\Sigma(M_{\rm H_2}) + \Sigma(M_{\rm HI}) + \Sigma(M_*)}. \quad (4)$$

The stellar mass is mapped directly from the MUSE data cubes (Pessa et al. in preparation). Voronoi bins are constructed with a target continuum S/N of 35, and the stellar population is fit using the Penalized Pixel-Fitting method (PPXF, Cappellari & Emsellem 2004; Cappellari 2017) without regularization, and masking of all emission lines and strong sky lines. As input templates for PPXF, we take the eMILES stellar population (SSP) models (Vazdekis et al. 2010) using a Kroupa IMF, Basti isochrones, and base (matching the Galactic abundances) α/Fe. Stellar masses are then calculated by taking into account the mass-to-light ratio of the stellar SSP templates. With this approach, we find typical agreements within 40 per cent to estimates derived from mid-infrared (mid-IR) stellar emission (e.g. Querejeta et al. 2015), where a constant M/L ratio was assumed after removing the non-stellar emission. This difference has proved to be strongly correlated with the age of the underlying stellar population in a given spaxel, where the overestimation of the mass value from mid-IR emission with respect to MUSE is more severe in spaxels dominated by younger stellar populations. Querejeta et al. (in preparation) will address these differences more in depth.

The molecular gas mass is measured from the PHANGS-ALMA survey CO(2$-$1) moment-zero maps at 1–2 arcsec resolution (Leroy et al., in preparation), assuming a fixed CO(2$-$1)/CO(1$-$0) ratio of 0.7 (Leroy et al. 2013; den Brok et al., in preparation; Saito et al., in preparation) and a Milky Way CO-to-H_2 conversion factor of 4.4 $\mathrm{M}_\odot\ \mathrm{pc}^{-2}\ (\mathrm{K\ km\ s}^{-1})^{-1}$ (Bolatto, Wolfire & Leroy 2013), which includes a factor of 1.36 to account for helium. Most of our targets are approximately Milky Way mass galaxies at solar metallicity, making this a reasonable assumption. Choosing to adopt a metallicity-dependent conversion factor (Bolatto et al. 2013) would result in a $\sim$20 per cent increase to the molecular gas estimates. Molecular gas maps for six of our eight targets have been reported by our earlier works (Sun et al. 2018; Kreckel et al. 2018; Utomo et al. 2018), and we refer the reader to those papers for additional visualizations and details.

The atomic gas mass is measured from 21 cm H I emission observed at the Karl G. Jansky Very Large Array (VLA) as part of archival (Walter et al. 2008; Chung et al. 2009) and new data (Legacy ID: AU157, PI: D. Utomo), and includes a correction for helium. The angular resolution ranges from 15 to 25 arcsec, well matched to the $\sim$2 kpc scales considered in our annuli for the most distant galaxies.

The Hα velocity dispersion ($\sigma_{\rm H\alpha}$, deconvolved from the instrumental response, but not corrected for thermal broadening) is measured within each H II region. We note, however, that the MUSE spectral resolution ($\sim$49 km s^{-1}at Hα; Bacon et al. 2017) is much larger than our measured values (10–40 km s^{-1}), and without an extremely careful determination of the line spread function we caution against placing high confidence in these measurements. However, given the exceptionally high S/N in Hα within the H II regions, this provides the most robust determination, and we take $\sigma_{\rm H\alpha}$ to be the median value of all H II regions within the annulus. As a test, we calculated the median Hα velocity dispersion using all pixels that fall within the annulus, which results in $\sim$5 km s^{-1} higher Hα velocity dispersions. This is likely due to either the increased turbulence or the higher scale height in the diffuse ionized gas (Della Bruna et al. 2020).

The CO velocity dispersion ($\sigma_{\rm CO}$) is measured from PHANGS-ALMA data (line widths for the PHANGS-ALMA data have been

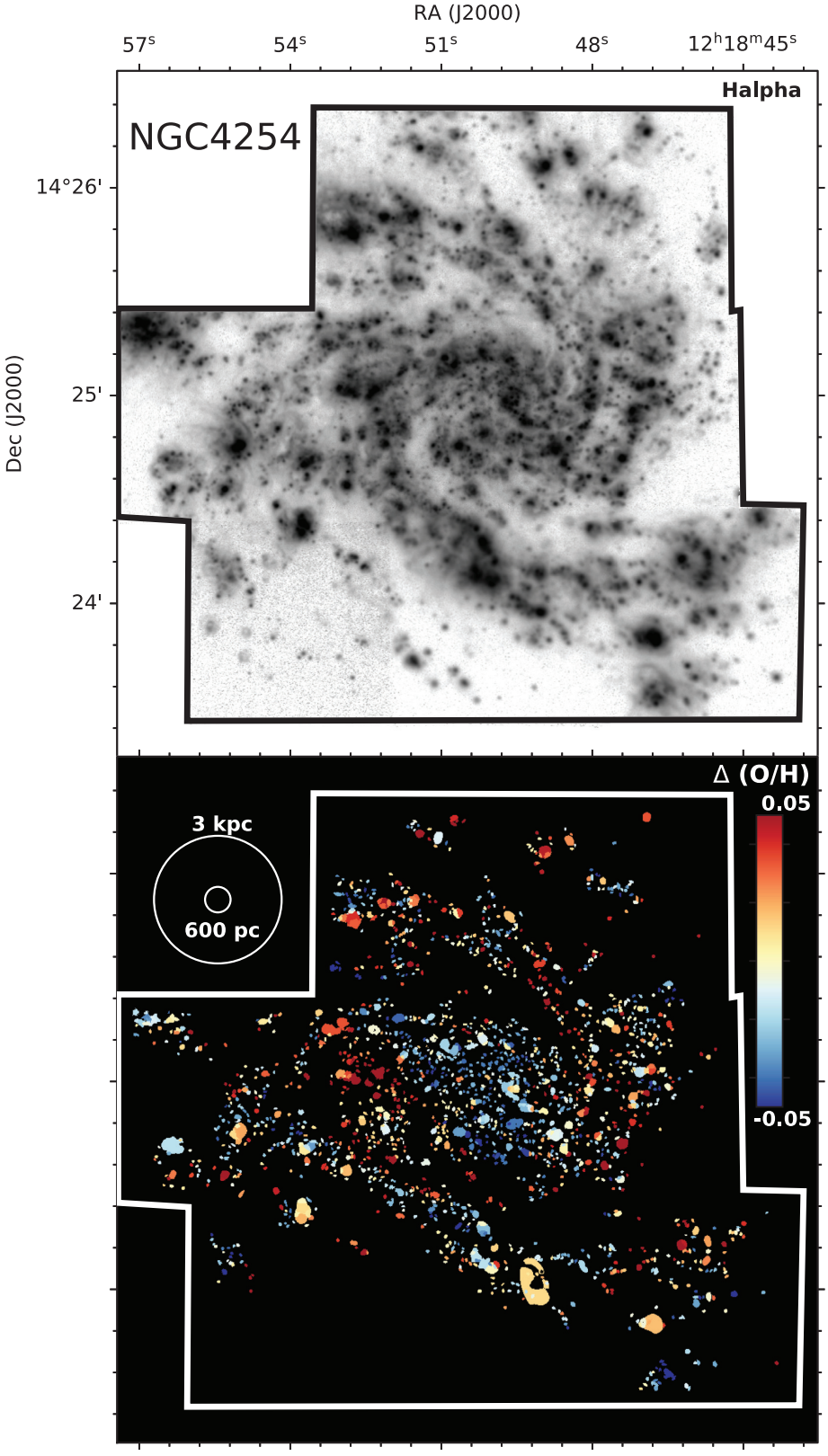

Figure 1. Mapping the H II regions in NGC 4254. Top: Hα emission is detected across the disc of the galaxy. In Paper I, we describe how the Hα morphology is used to identify H II region boundaries, resulting in a sample of H II regions that have been isolated from the surrounding diffuse ionized gas. Bottom: colour coding each H II region by its metallicity offset from a radial gradient, Δ(O/H), it is clear that we sample the gas-phase oxygen abundances throughout the disc. In this paper, we quantify the structures revealed in these Δ(O/H) maps.

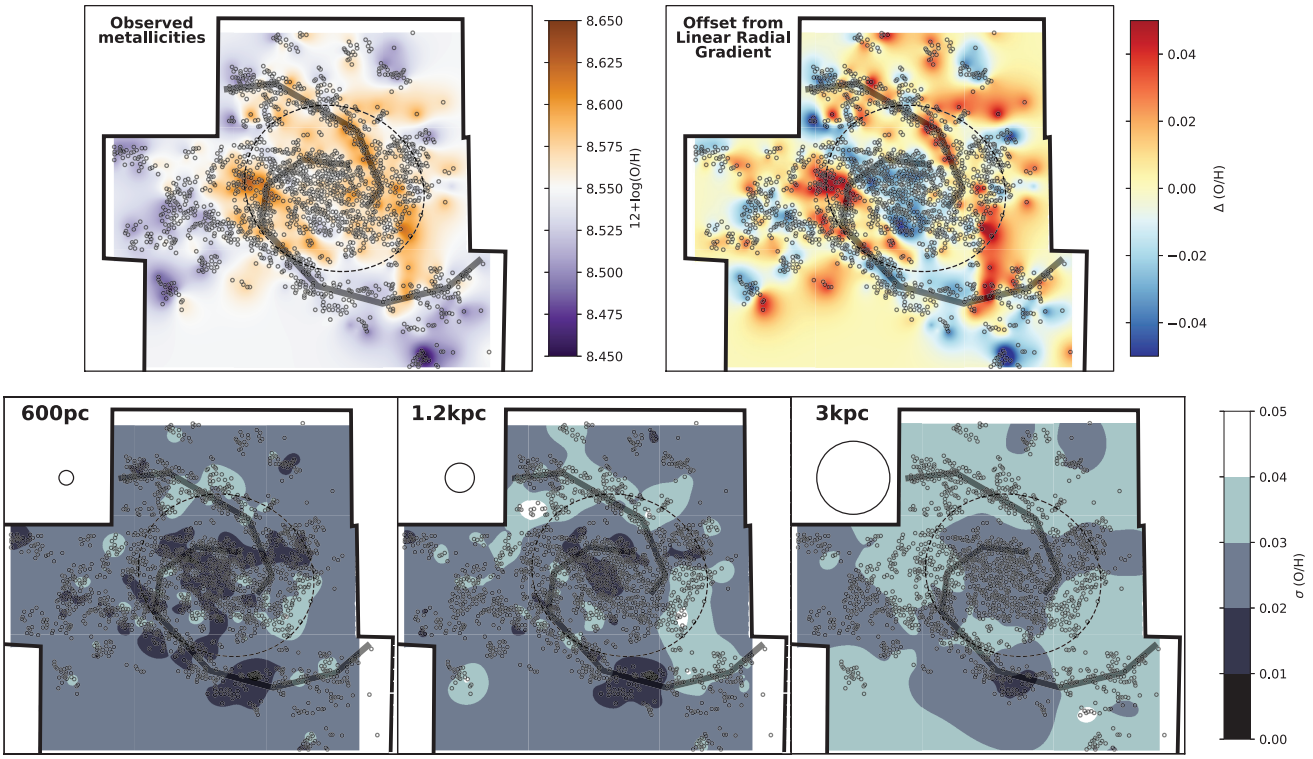

Figure 2. Visualizing the large-scale homogeneity in metallicities within NGC 4254. Circles represent individual H II regions. The dashed circle is at $0.3R_{25}$, and grey lines roughly outline the spiral structure. In colour scale, we show an interpolation between the values using `kriging`, a local estimation technique that derives the best linear unbiased prediction of the intermediate values. This is done purely for visualization purposes, and assumes an exponential model with 100 pc scale length, which is approximately twice our spatial resolution (see Table 1). We map the observed metallicity (12 + log(O/H), top left) and the offset from a linear radial gradient (Δ(O/H), top right; see Paper I). For the bottom panels, we have calculated the standard deviation of metallicity values between each H II region and its neighbours over increasing distances (600 pc, and 1.2 and 3 kpc diameters). We see systematically reduced scatter on the smallest 600 pc spatial scales. This is also visible in the maps of Δ(O/H), where this $\sim$600 pc scale is well matched to the scale length over which uniform values are observed.

presented and investigated in Sun et al. 2018, 2020). Here, we use the velocity dispersion estimated from the rms scatter of emission about the intensity-weighted mean velocity (the 'moment 2' line width). Use of an alternate parametrization of the molecular gas line width does not significantly change our results. As this emission predominantly arises from within cold molecular clouds, it shows systematically lower velocity dispersions than the ionized gas.

3 QUANTIFYING THE CHEMICAL HOMOGENEITY OF THE ISM

From the image atlas presented in Paper I, we see that the metallicity distribution in all galaxies is dominated by a radial gradient. In this paper, we do not include images for all galaxies, but present NGC 4254 as an example in Fig. 2. The top left panel maps the observed metallicities. In the top right panel, we have subtracted the linear radial metallicity gradient from H II region measurements, and map the variations in metallicity (Δ(O/H)). These are generally small (0.03–0.05 dex) when measured globally across the full galaxy disc. In Paper I, we estimate that the maximum level of enrichment we observe could be caused locally by a 5–10 Myr long period of continuous star formation.

A striking detail apparent in the maps is that after subtracting a linear radial gradient, neighbouring H II regions show very little scatter in Δ(O/H). To visualize the size scale over which we observe this increased homogeneity, in Fig. 2 (bottom panels) we calculate the standard deviation in Δ(O/H) (σ(O/H)) at each H II region position over fixed 600 pc, and 1.2 and 3 kpc deprojected diameter apertures. We require each aperture to have at least four H II regions in order to calculate the standard deviation, which holds for 90–95 per cent of 600 pc deprojected diameter apertures. When calculating the standard deviation on 600 pc scales, we see connected, $\sim$kpc sized regions with very low scatter (σ(O/H) < 0.02), consistent with the metallicity measurement uncertainties. Similar large regions with low scatter (σ(O/H) $<$ 0.02) are still seen at 1.2 kpc scales, but at 3 kpc scales most of the map is consistent with the global value (σ(O/H) = 0.03–0.04).

3.1 Global statistics

In Fig. 3, we compare the global distribution in Δ(O/H) at different physical scales, across all galaxies in our sample. To highlight the width of the distribution, for this figure we show Δ(O/H) measured relative to the local median. To minimize repeated sampling of clustered H II regions, we use independently sampled 600 pc and 1.2 kpc, and global apertures. Thus, on local scales, a narrower width in the distribution corresponds to a smaller intrinsic scatter in metallicities. We see directly the reduced scatter in the histograms for the smallest 600 pc scale size, with a minimum observed scatter of $\sim$0.02 dex that is consistent with our quoted errors. Statistics relating to these distributions for each galaxy individually are given in Table 2, including the global scatter (standard deviation), skewness

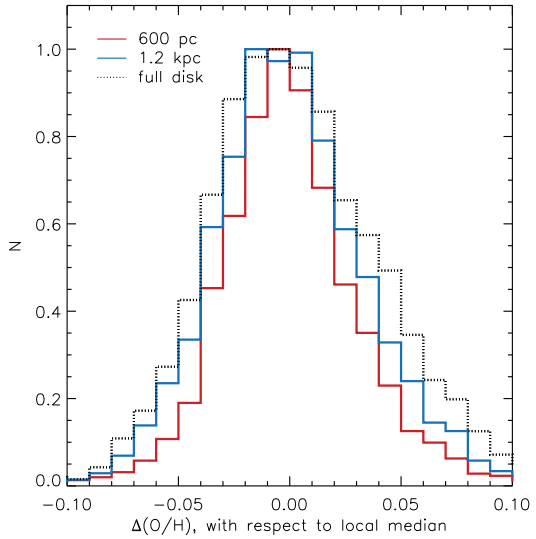

Figure 3. Distribution of metallicity offsets from a radial gradient, Δ(O/H), at different physical scales. To compare the widths of these distributions, we measure Δ(O/H) with respect to the local median within independent apertures with diameters of 600 pc (red), 1.2 kpc (blue), and globally (dotted) for all galaxies in our sample taken together. The typical (statistical) uncertainty in the measured metallicity is ∼0.02 dex. Metallicities are more homogeneous on smaller scales, and well represented by a Gaussian distribution (Table 2).

and kurtosis in Δ(O/H), and the average scatter over 600 pc diameter apertures for each galaxy separately.

We note that in Fig. 3 the global distribution in Δ(O/H) is not symmetric. This is because, as shown in Paper I, the linear radial gradients are fit accounting for the uncertainties in each measurement. The higher metallicity regions are often brighter and hence have correspondingly smaller uncertainties, driving the asymmetry. In this figure, our distribution is shown relative to the median value, which sits at negative Δ(O/H), and results in the apparent skewness towards positive values.

3.2 Trends with physical scale

To investigate any trends with physical scale, we take circular apertures with increasing diameter around each H II region in order to calculate σ(O/H) and plot the median value as a function of scale (Figs 4 and 5). Here, we have required a minimum of four H II regions be contained in the aperture. We have chosen this limit as small sample sizes result in a systematic underestimation of the standard deviation. To ensure that we achieve an unbiased estimation of the standard deviation, we apply a correction based on Cochran's theorem (assuming a normal distributions of values) when calculating σ(O/H). This corresponds to a correction factor of ∼8 per cent for a sample of four and <3 per cent for a sample of 10. At 300 pc (600 pc) scales, we find that ∼90 per cent (∼20 per cent) of apertures contain 10 or fewer H II regions. Fig. 4 shows the resulting standard deviation in absolute values (left) and relative to the global scatter (right).

All galaxies show very similar behaviour, with decreased scatter (increased homogeneity) on the smallest sub-kpc scales. We tested if this could be produced by instrumental effects, and found that the signal persists both within and across different MUSE pointings. Global values of σ(O/H) also vary, and the apparent quantization is striking, but we find no obvious correlation with galactic properties such as stellar mass, SFR, metallicity, or metallicity gradient (though our sample size is limited). Repeating this analysis with independent, regularly sampled apertures gives quantitatively similar results but with increased scatter (due to more limited statistics).

As the H II regions are not uniformly distributed across the disc, but often clustered in and around spiral arms, we also test the null hypothesis (that metallicities are uncorrelated) by randomly shuffling all metallicities across the disc and repeating the analysis (in grey). We repeat this shuffling procedure 100 times for each galaxy, in order to determine the confidence interval of the signal we detect (in light grey). In all galaxies, the increased homogeneity on sub-kpc scales is significantly different from the null hypothesis (a randomized distribution). We also perform a similar test, modelling all H II to have a simple linear radial gradient, introducing a scatter equivalent to the global σ(O/H). This test produces qualitatively similar trends, and is unable to reproduce the observed homogeneity on small spatial scales.

Fig. 5 demonstrates that for each individual galaxy (except IC 5332) the small-scale homogeneity is significantly different from the null hypothesis (that metallicities are uncorrelated), even when taking into account the observational uncertainties. For each galaxy, the coloured range indicates the uncertainty after bootstrapping the error in metallicity over 100 realizations. The light and dark grey bands also confirm the confidence of our detection, including both different realizations of shuffling the metallicities (light grey) and resampling the errors (dark grey). IC 5332, the least massive galaxy in the sample, shows the least significant trend, but it also has the most limited statistics, with only 223 H II regions detected. The remaining galaxies show significantly increased homogeneity out to scales of 600 pc and larger.

4 MEASURING THE CORRELATION SCALE

We now quantify the spatial scale below which the metallicity exhibits reduced scatter (and thus enhanced homogeneity). Consider the vector position of any given H II region in a galaxy, $\boldsymbol{r_1}$, which provides a measurement at that location of the 2D gas-phase oxygen abundance field, S_X. In practice, we measure $S_X(\boldsymbol{r_1})$ as the offset from the radial gradient (Δ(O/H)), removing any radial trends from this global analysis. We can then calculate the two-point correlation (ξ) of the observed gas-phase oxygen abundance as a function of spatial scale (r) for each galaxy as

$$\xi(r) = \left\langle \frac{\overline{S_X(\boldsymbol{r_1})S_X(\boldsymbol{r_2})} - \overline{S_X}^2}{\overline{\left(S_X - \overline{S_X}\right)^2}} \right\rangle, \quad (5)$$

where $|\boldsymbol{r_1} - \boldsymbol{r_2}| \leq r$. The horizontal lines indicate averaging over all H II regions in the galaxy and angle brackets indicate averaging over all choices of $\boldsymbol{r_1}$. A 100 per cent correlation is expected at scales of $r = 0$, as each H II region correlates perfectly with itself. The shape of the correlation function traces the scale over which the metal field retains some level of homogeneity. In the following analysis, we parametrize these curves by the scale at which the metallicities remain correlated at the 30 per cent or 50 per cent level.

Fig. 6 compares the correlation functions of all galaxies, while Fig. 7 shows each of our eight galaxies individually. In all galaxies except NGC 628, the correlation does not drop below 50 per cent until scales become larger than ∼300–400 pc and remains above 30 per cent out to ∼kpc scales. This is a significant result in all galaxies (including NGC 628) when compared to the null hypothesis, where we have randomized S_X. This significance persists when including random sampling of the metallicity uncertainties. NGC 628 shows shorter correlation scales, but is still significantly different

Table 2. Measured values quantifying the homogeneity and correlation length scale of the ISM.

Name	Global scatter (standard deviation in Δ(O/H))	Global skewness in Δ(O/H)	Global kurtosis in Δ(O/H)	Average 600 pc scatter in Δ(O/H)	50 per cent correlation scale (pc)	30 per cent correlation scale (pc)
NGC 628	0.049	0.355	−0.365	0.040	230 ± 8	360 ± 14
NGC 1087	0.033	−0.184	0.009	0.020	380 ± 7	710 ± 54
NGC 1672	0.033	0.193	0.058	0.021	370 ± 6	760 ± 62
NGC 2835	0.042	0.012	0.259	0.033	290 ± 2	510 ± 30
NGC 3627	0.035	0.393	0.025	0.024	370 ± 5	820 ± 79
NGC 4254	0.032	0.454	0.028	0.022	340 ± 8	740 ± 39
NGC 4535	0.043	0.372	−0.295	0.032	290 ± 2	550 ± 35
IC5332	0.051	0.025	0.036	0.039	270 ± 8	510 ± 77

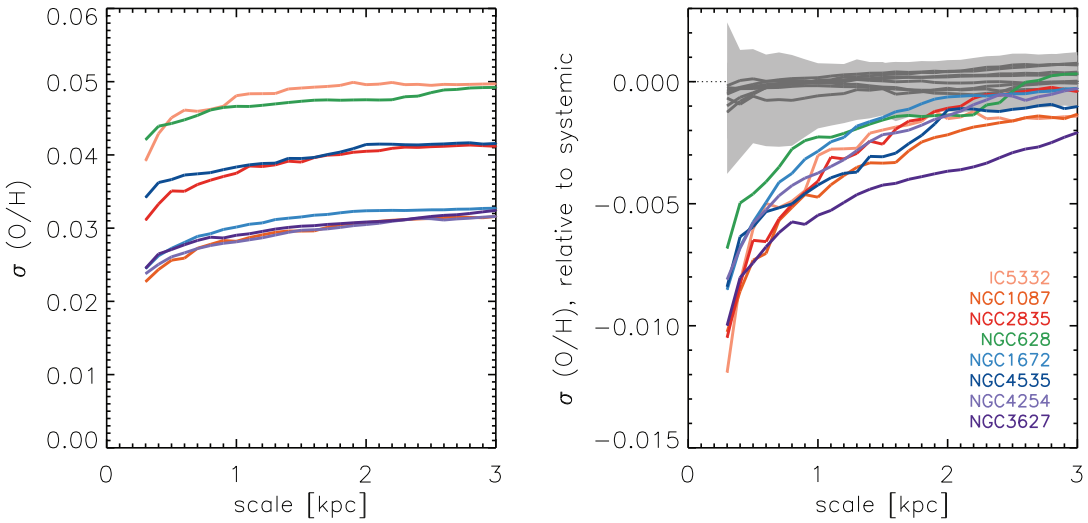

Figure 4. Median scatter in Δ(O/H) over a given spatial scale, for apertures centred on H II regions and containing at least 4 H II regions. Here, we apply a correction when measuring σ based on Cochran's theorem (assuming a normal distribution of values) to ensure that we achieve an unbiased estimation of the standard deviation. The observed trends (left) and trends in each galaxy measured with respect to the global scatter (right) show consistent homogenization on sub-kpc scales. Galaxy names are sorted by total stellar mass, from low (orange) to high (purple). In grey, we test the null hypothesis by randomly shuffling Δ(O/H) 100 times for each galaxy. In dark grey, we show the median of each galaxy separately for all realizations, and the light grey band traces our 1σ confidence interval across the sample. In our test of the null hypothesis, we are unable to reproduce the pronounced increase in homogeneity at smallest spatial scales, and find that our signal is detected at the 3σ level.

from the null hypothesis. The global 30 per cent and 50 per cent correlation scales for each galaxy are given in Table 2, and all results presented here are not sensitive to our choice of threshold.

We compare these observed correlations with the model proposed by Krumholz & Ting (2018) for a stochastically forced diffusion model. This model uses an analytic and semi-analytic approach to mimick the diffusion of metals by solving an evolution equation including random (star-formation-driven) injections, combined with linear diffusion as a proxy for turbulent transport. We find good agreement with their fiducial spiral and dwarf galaxy models (their fig. 5), although their model does not incorporate any structural disc features (e.g. spiral arms, bars, or differential rotation). A more detailed comparison of our results with their model is given in Section 5.3.

Breaking each galaxy down into 2 kpc wide (de-projected) radial bins, we explore how the correlation scale changes as a function of local properties that might be expected to drive turbulence and mixing in the ISM. This annular width is chosen to ensure that most annuli contain at least 100 H II regions and is well matched to the spatial scales available from ancillary data (see Section 2.2). Annuli with less than 80 per cent of their area contained in the MUSE maps are excluded, as is the inner $r < 500$ pc. We parametrize each annulus by the scale at which the correlation falls below 30 per cent. An analysis performed using the 50 per cent correlation scale gives similar results.

Fig. 8 (top) compares the correlation scale with radius and $\Sigma_{\rm SFR}$. By the Spearman's rank correlation coefficients calculated over the full sample, radius shows no correlation and $\Sigma_{\rm SFR}$ shows a positive correlation ($\rho \sim 0.3$) but with only weak statistical significance ($p \sim 0.05$). This figure (bottom) also shows a comparison of correlation scale with orbital time and total gas fraction. A weak correlation is seen for longer correlation scales at shorter orbital times. However, the Spearman's rank correlation coefficient is not significant ($p \sim 0.1$), and no significant correlation is seen with gas fraction.

Fig. 9 explores the impact of gas velocity dispersion, measured in three ways. $\sigma_{\rm H\alpha}$ is the median value from pixels across the full annulus (left) and from only pixels within H II regions (centre). Measurements are corrected for instrumental broadening, but not for thermal broadening. We measure $\sigma_{\rm CO}$ (right) as the median from the second moment CO map, which has already been masked to

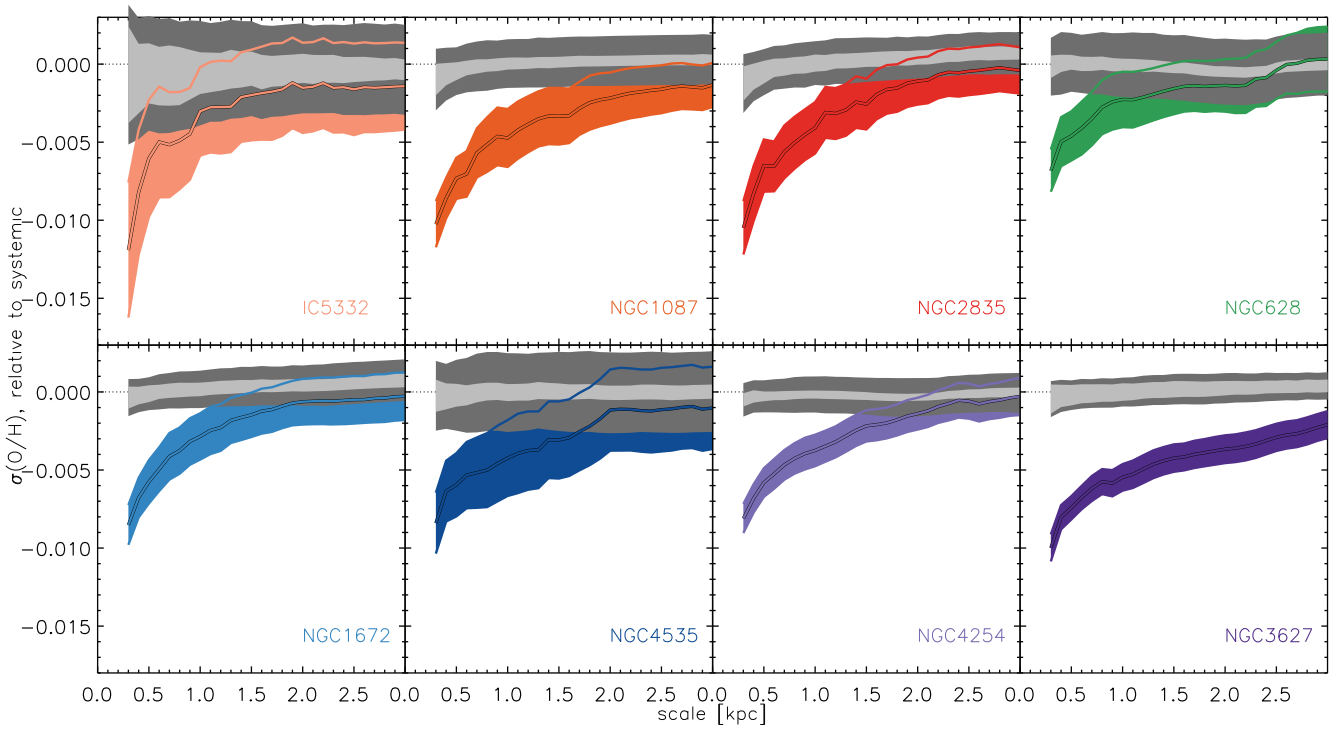

Figure 5. Median scatter in Δ(O/H) relative to the global scatter over a range of spatial scales, as in Fig. 4, for each galaxy individually. The coloured range indicates the uncertainty after bootstrapping the error in metallicity over 100 realizations. In grey, we test the null hypothesis by randomly shuffling Δ(O/H). The bands show the 1σ distribution given 100 different realizations of the shuffling (light grey) and 100 different instances of Gaussian sampling a single randomized realization (dark grey). IC 5332 has the smallest sample size, and we are unable to disprove the null hypothesis at 3σ confidence. The remaining seven galaxies show increased homogeneity on small <1 kpc scales at the 1σ–3σ level.

include only lines of sight with significant CO emission. This CO emission is concentrated in a thin (∼150 pc) layer within the galaxy mid-plane, a difference reflected by the systematically lower velocity dispersion compared to the ∼kpc scales traced by the ionized ISM in Hα. In grey, we show how we can scale the CO velocity dispersions to disc (∼1 kpc, Levy et al. 2019) scale heights assuming a power-law relation between velocity dispersion and spatial scale. Adopting a power-law index of 1/2 (Falgarone et al. 2009; Klessen & Glover 2016), these scaled CO velocity dispersions compare reasonably well with what is measured in Hα, and retain the strong correlation with the 30 per cent correlation scale. Across all gas phases, significantly larger correlation scales are seen for annuli with higher gas velocity dispersion with high significance ($p < 0.01$).

Seven of our galaxies have been analysed with the 'uncertainty principle for star formation' formalism, introduced by Kruijssen & Longmore (2014) and Kruijssen et al. (2018). This technique compares the distribution of spatially resolved (∼100 pc) CO and Hα emission peaks, constraining and quantifying the underlying evolutionary timeline associated with the molecular cloud lifetime, the feedback time-scale (during which clouds are dispersed), and the (unobscured) H II region lifetime. One parameter measured by this technique is the feedback outflow velocity, defined as the ratio of the average CO emission peak radius over the feedback time-scale. This velocity is derived independently of the CO or Hα kinematics, and is solely determined using the CO and Hα morphology. These measurements are available globally (Chevance et al. 2020b; Kim et al., in preparation) as well as in radial bins (Chevance et al., in preparation), and we compare both data sets to our 30 per cent correlation scales in Fig. 10. Although the sample size is small, we do observe that galaxies with homogeneity over a larger size scale have larger feedback velocities, with a Spearman's rank correlation coefficient of $\rho = 0.78$ and high ($p = 0.005$) significance.

5 DISCUSSION

As we report in Paper I, we observe very small systematic scatter (0.03–0.05 dex) globally across all galaxies in our sample after subtracting the radial gradients. This agrees well with the ∼0.05 dex systematic scatter observed by Sánchez-Menguiano et al. (2018) using strong-line methods to measure metallicities within 102 spiral galaxies. This level of scatter is much smaller than the large-scale scatter typically measured using direct temperature methods (∼0.1 dex; Rosolowsky & Simon 2008; Croxall et al. 2015; Berg et al. 2015; Croxall et al. 2016), and it has been previously shown that strong-line methods systematically produce smaller scatter in their radial gradients than direct temperature methods (Arellano-Córdova et al. 2016). In this work, we report even lower scatter (∼0.02 dex) on the smallest <600 pc scales (Table 2). This discrepancy between strong-line and direct temperature methods is perplexing, and in fact our observed minimum σ(O/H) is consistent with our quoted errors.

While the metallicities presented here are measured assuming the Pilyugin & Grebel (2016) calibration, we note that all results are qualitatively reproduced with the theoretical Dopita et al. (2016) N2S2 calibration as well as the empirical (Marino et al. 2013) N2 and O3N2 calibrations (see Appendix A). In addition, the detailed azimuthal trends we reported in Paper I based on strong-line abundances have been confirmed with auroral line electron temperature determinations in one of our galaxies (NGC 1672, Ho et al. 2019), increasing our confidence in the statistical trends we report in this paper.

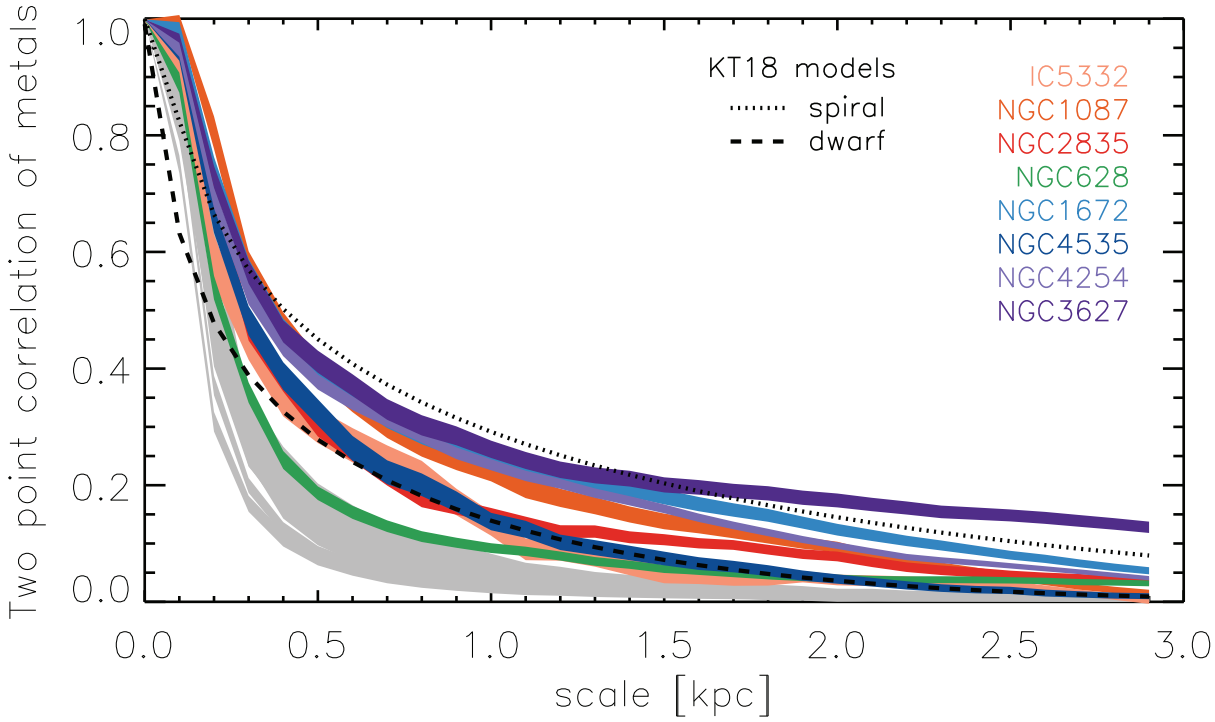

Figure 6. Two-point correlation of metallicity, Δ(O/H) (after removing the radial gradient) as a function of separation between H II regions, comparing all galaxies. The bands show the 1σ distribution given 100 different instances of Gaussian sampling the metallicity uncertainties. The null hypothesis (in grey), that metallicities are uncorrelated, is tested by performing this analysis when all values have been randomized, where the bands show the 1σ distribution given 100 different realizations of the shuffling. Most galaxies show correlations at the 30 per cent level out to nearly kpc scales (see also Fig. 7). Black lines are from the KT18 (Krumholz & Ting 2018) model. The dotted and dashed lines represent the fiducial spiral and dwarf galaxies, respectively, to give a sense of how increased star formation and stellar mass impact the model shape. Galaxy names are ordered by stellar mass, from low (orange) to high (purple).

We measure 30 per cent correlation scales of 300–800 pc (Table 2). This indicates the typical physical scales over which mixing in the ISM is most effective. This has not been well quantified in previous literature. Sánchez et al. (2015) define a 'mixing length' based on the dispersion of H II region galactocentric distances with respect to a linear metallicity gradient in NGC 6754. They measure a 4.6 kpc 'mixing length', which parametrizes how far a certain H II region has moved from its expected location based on pure inside-out chemical enrichment. This is much larger than the 30 per cent correlation scales we report here. In fact, by our analysis most galaxies appear to have reached their global scatter by ∼3 kpc scales.

We note that a measurement of length scales in this context is inextricably linked to the time-scales involved in the ongoing pollution of the ISM by stars and supernovae. In Fig. 10, we observe a strong correlation between our 30 per cent correlation scales and the feedback outflow velocities reported in Chevance et al. (2020b), which derives from the time-scales associated with the dissolution of molecular clouds. This, combined with the correlations with Σ_{SFR} and gas velocity dispersion (Figs 8 and 9) suggests that stellar feedback drives much of the mixing while (presumably) simultaneously contributing to the ongoing enrichment of the ISM.

In this section, we further discuss how the homogeneity measured in Section 3 and the correlation length measured in Section 4 can be interpreted within the context of the injection and dilution of metals, as well as in relation to the disc morphology and spiral structure, and finally how this compares with existing diffusion and mixing models.

5.1 Injection versus dilution of metals

The mixing process is responsible for taking material that is enriched within stars and injected through winds and supernovae into the ISM, and redistributing it throughout the disc. The overall efficiency of this process is reflected in the well-established mass–metallicity relation (Tremonti et al. 2004; Sánchez et al. 2019), that ties the buildup of stellar mass to the buildup of metals in the disc and may lead to the negative radial metallicity gradients measured in almost all nearby galaxies (e.g. Magrini et al. 2016). We further explore the trends reported in Section 3, focusing on only those H II regions with enhanced abundances relative to the radial gradient (1σ < Δ(O/H)), and only those H II regions with reduced abundances (Δ(O/H)<−1σ). As these are defined to be the 1σ outliers (Fig. 3), these two samples are roughly equal in size. The use of a fixed threshold for the definition of both samples (e.g. 0.03 dex) produces similar results.

As Paper I showed, within each galaxy the H II regions with enhanced abundances are distributed with wider separation than the H II regions with reduced abundances. In Fig. 11, we reproduce the analysis in Fig. 4 but using only these two populations. Measuring the median σ(O/H) as a function of scale for only H II regions with reduced abundances, we recover the same systematic trend for increased homogeneity on small scales. Here, all trends are shown as the offset from the global scatter (Table 2), and compared with the null hypothesis (that all metallicities are uncorrelated), which we test by randomizing all H II region metallicities. In contrast, the median σ(O/H) of H II regions with enhanced abundances show no such increased homogeneity on small scales, particularly in relation to the null hypothesis.

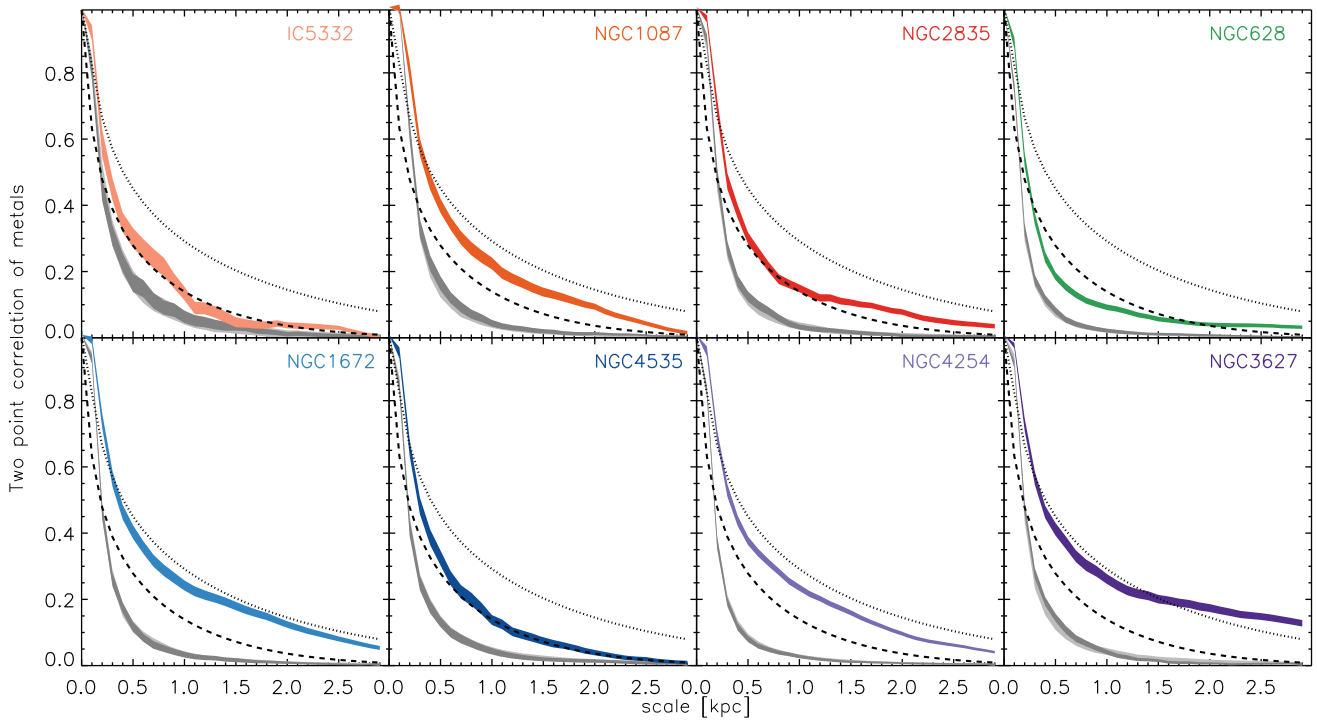

Figure 7. Two-point correlation of metallicity, Δ(O/H) (after removing the radial gradient) as a function of separation between H II regions, as in Fig. 6, for each galaxy individually. The coloured region indicates the uncertainty estimated as a function of scale using bootstrap sampling of the reported uncertainties in the metallicity measurements. The null hypothesis (in grey), that metallicities are uncorrelated, is tested by performing a similar analysis where all values have been randomized. The bands show the 1 sigma distribution given 100 different realizations of the shuffling (light grey) and 100 different instances of Gaussian sampling a single randomized realization (dark grey), which introduce similar levels of uncertainty. Black lines are from the Krumholz & Ting (2018) model. The dotted and dashed represent the fiducial spiral and dwarf galaxies (respectively). In all galaxies, significant ($>1\sigma$) differences are seen across a range of scales, with all showing significantly higher correlations out to >2 kpc scales. Even for NGC 628, which exhibits shorter correlation lengths, this trend is significantly different from the null hypothesis. Galaxies are ordered by stellar mass (low to high; top left to bottom right), with a suggestion of larger correlation scales in the more massive galaxies, roughly consistent with model predictions.

The stark difference observed in Fig. 11 between these populations indicates that the mechanisms driving homogeneity in the ISM act similarly on H II regions with average abundances and H II regions with decreased abundances relative to the radial gradient. In contrast, the enriched pockets appear to represent a local phenomenon, and show that the injection of metals is not responsible for the observed chemical homogeneity. This is in good agreement with our hypothesis in Paper I that these enhancements reflect local enrichment produced by 5–10 Myr length episodes of star formation. In this case, the new metals have not yet diffused more broadly into the ISM.

The lengths scales over which we observe homogeneity are similar to the ionized and atomic gas scale heights (Ferrière 2001), and hint towards large-scale processes being at work. For example, the persistence of star formation over cosmic time (Prochaska & Wolfe 2009) as well as the presence of deuterium in many regions of the Milky Way (Linsky 2003; Lubowich et al. 2000) suggests that large-scale accretion of fresh material on to a galaxy must be a common and continuous process (see also Sancisi et al. 2008). If the new gas is accreted from the cosmic web (Dekel et al. 2009) it has considerably lower metallicity than the galactic ISM, and because it falls predominantly on to the outer regions of the disc, the accretion process will steepen the radial metallicity gradient. On the other hand, if the infalling gas is part of a galactic fountain flow (Marasco, Fraternali & Binney 2012; Fraternali 2017), then it has been enriched and lifted out of the disc by stellar feedback in the inner regions of the galaxy and typically comes back at larger radii. In this case, it will flatten any pre-existing metallicity gradient. Regardless of its origin, the kinetic energy associated with the infalling motion can contribute significantly to the observed turbulence in the galactic ISM (Klessen & Hennebelle 2010) and enhance to the global mixing efficiency. In combination with the spiral-driven 'carousel' model of enrichment and mixing (see Section 5.2 below), this could provide an additional mechanism that facilitates the dilution of gas in the disc while simultaneously driving mixing.

5.2 Relating disc dynamics and environments with mixing

A similar analysis separating H II region populations by environment (bar, arm, and interarm) shows no clear systematic trends. However, the difficulties faced in robustly identifying spiral features make this result difficult to interpret. Individual case studies (Sánchez-Menguiano et al. 2016; Ho et al. 2017; Vogt et al. 2017; Ho et al. 2018) have revealed that spiral structure plays an important role in establishing azimuthal enrichment patterns in the ISM, but in our sample of eight galaxies we find galaxies can vary dramatically, with only half-containing subsections of spiral arms with clearly associated azimuthal metallicity gradients (Paper I). As the expected trends depend on a determination of the pattern speed and corotation radius (Spitoni et al. 2019), a full dynamical model of individual galaxies would be needed to reveal trends (see e.g. the case study of NGC 1672, where clear chemical enrichment is observed along the spiral arm ridge, Ho et al. 2019), but this is beyond the scope of this paper.

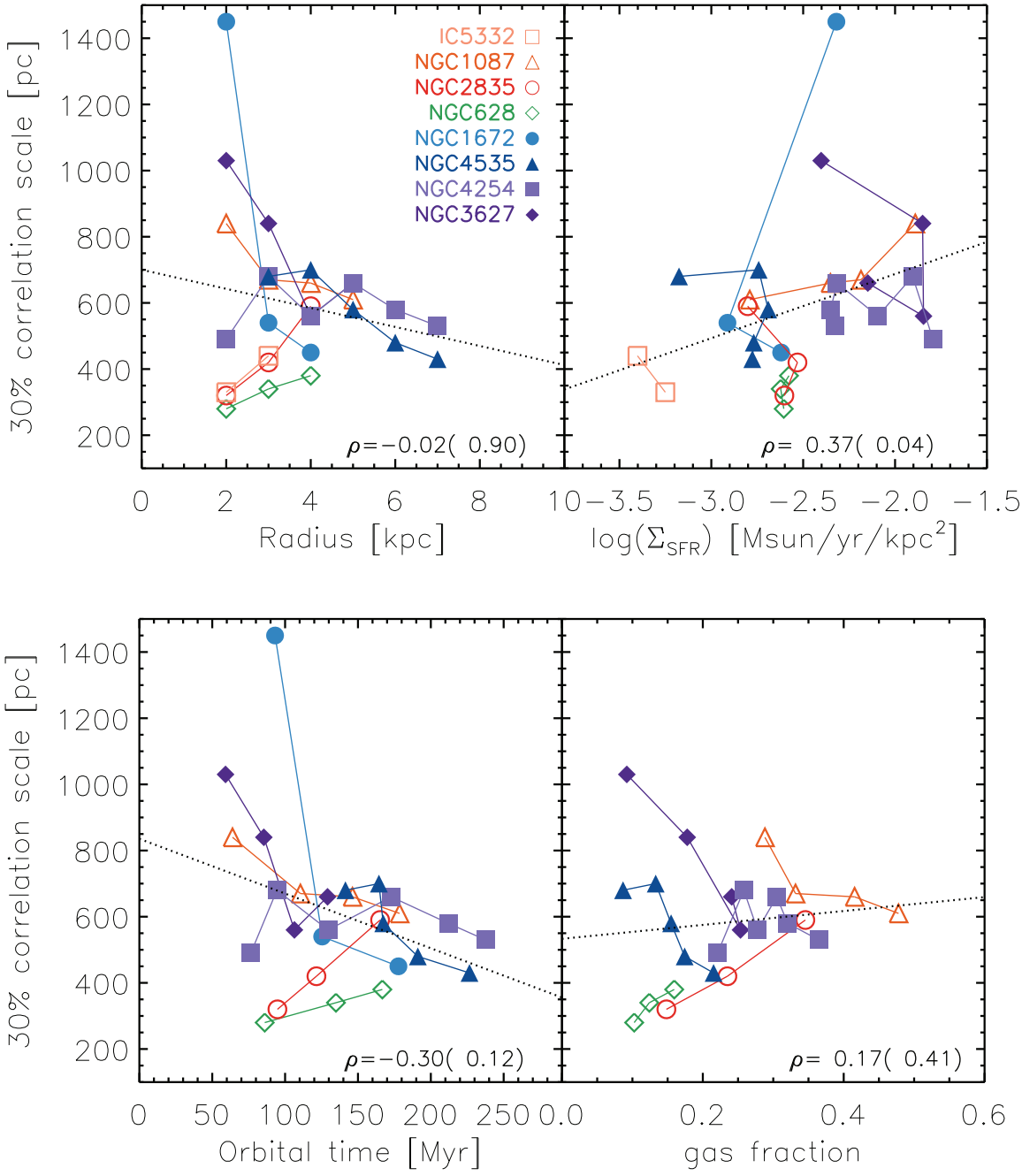

Figure 8. The size scale associated with a 30 per cent correlation (see Fig. 6) within 2 kpc wide annuli, as a function of radius (top left), Σ_{SFR} (top right), orbital time (bottom left), and total gas fraction (bottom right). Larger correlation scales are seen for annuli with higher Σ_{SFR}, and shorter orbital time. However, the Spearman's rank correlation coefficient is not significant ($p \sim 0.1$). No clear trend is seen with gas fraction. Coefficients and significance are shown for each panel in the lower right corners. A linear fit to all points is shown to guide the eye.

Since it is clear that the spiral arms play some role in organizing and mixing the ISM, we can interpret our results within the context of two existing simple models for how the spiral structure could be influential. Given that most galaxies show negative radial abundances gradients, radial gas flows driven along spiral structures could introduce azimuthal patterns in the gas-phase abundances (Grand et al. 2016; Sánchez-Menguiano et al. 2016). The mixing induced by spiral-driven streaming motions would be characterized (inside of corotation) by a metal-rich trailing edge and a metal-poor leading edge, as relatively metal-rich material is driven outward (along the trailing edge) and relatively metal-poor material is driven inward (along the leading edge). In this scenario, we would expect H II regions with both enhanced and reduced abundances relative to the radial gradient to symmetrically show similar spatial correlations. As we observe distinct differences in these two populations, we believe this model is not uniformly supported across our sample.

Ho et al. (2017) present an alternative 'carousel' model, where enrichment can proceed slowly in small pockets as gas passes through the interarm region, with enrichment peaking at the spiral arm. Passage through the spiral arm then triggers large-scale mixing.

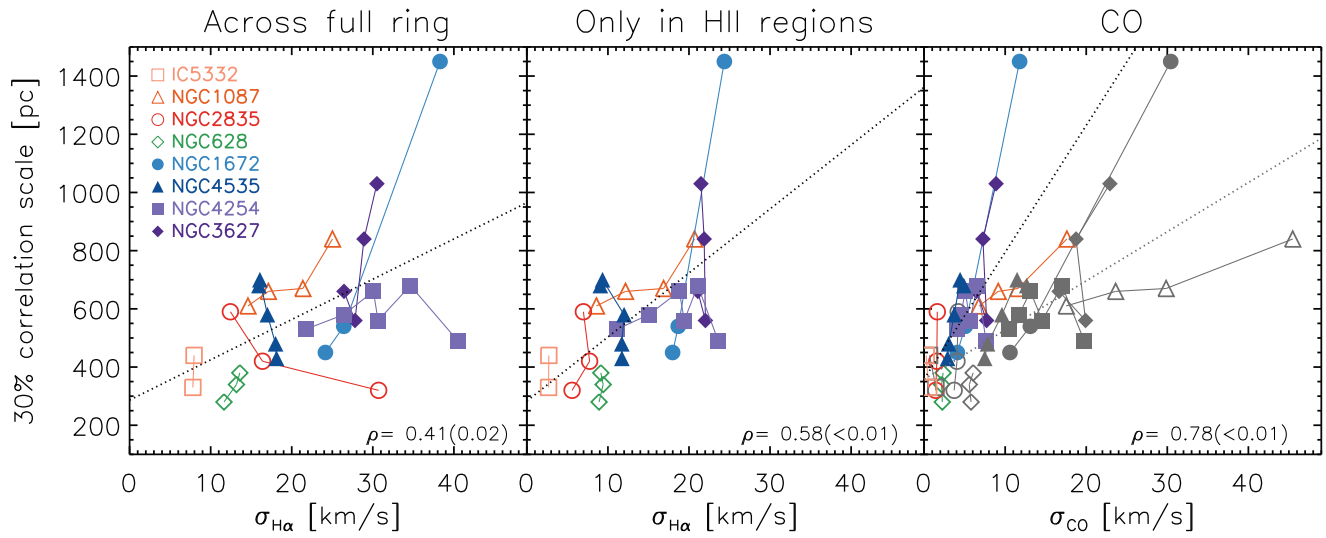

Figure 9. The size scale associated with a 30 per cent correlation (see Fig. 6) within 2 kpc wide annuli, as a function of the azimuthally averaged ionized gas velocity dispersion. All measurements are based on pixel-scale (50–100 pc) measurements, and are corrected for instrumental broadening. $\sigma_{\mathrm{H}\alpha}$ is the median value from pixels across the full annulus (left) and from only pixels within H II regions (centre). σ_{CO} (right) is the median from the second moment map. In grey, we also show a scaling of the CO velocity dispersions to disc (∼1 kpc) scale heights assuming a power-law index of 1/2 (Falgarone, Pety & Hily-Blant 2009). These scaled values are used as input for the Krumholz & Ting (2018) model, and are roughly consistent with the Hα values. Larger correlation scales are seen for annuli with higher gas velocity dispersion across both ionized gas and molecular gas tracers. The Spearman's rank correlation coefficient is significant ($p < 0.01$) for all gas phases (lower right corners of each panel). A linear fit to all points is shown to guide the eye.

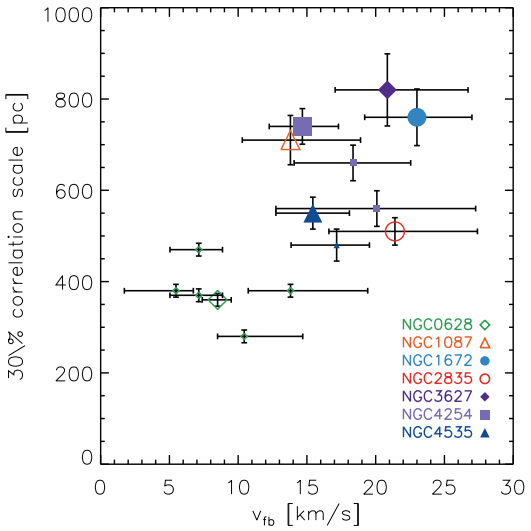

Figure 10. The size scale associated with a 30 per cent correlation (see Fig. 6) as a function of the feedback outflow velocity (v_{fb}) reported in Chevance et al. (2020b), Kim et al. (in preparation). The quantity v_{fb} is defined as the ratio of the average CO emission peak radius over the feedback time-scale (during which CO and Hα are cospatial), as measured using the 'uncertainty principle for star formation' formalism introduced by Kruijssen & Longmore (2014) and Kruijssen et al. (2018). Global results for the seven galaxies that overlap between samples are shown with large symbols, individual radial bins are shown with smaller symbols. We see a clear correspondence between these two parameters (Spearman's rank correlation coefficient of $\rho = 0.78$ with high significance; $p = 0.005$), supporting the indications from Figs 8 and 9 that these mixing scales are set by star formation feedback.

This homogenization effectively redistributes any enriched pockets of material, resulting in overall lower metallicities on the trailing side of the arm compared to along the spiral arm ridge (where metals have maximally accumulated before redistribution). Given the nearly kpc scale mixing we observe, this could potentially also introduce relatively more pristine gas from the surroundings or from freshly accreted material at large scale heights, systematically decreasing the abundances. This is more consistent with our observations, as both the average H II regions and the H II regions with reduced abundances show increased homogeneity (Fig. 11), and the correlation scale is larger for annuli with higher gas velocity dispersion (Fig. 9, left).

5.3 Comparison with mixing models

Chemical mixing within the ISM is expected to be driven by a combination of processes operating on different physical scales. On large scales, turbulent diffusion can be driven by supernovae or by thermal instability, aided by the effects of differential rotation (Roy & Kunth 1995; de Avillez & Mac Low 2002; Klessen & Lin 2003; Yang & Krumholz 2012; Petit et al. 2015). Hydrodynamical studies of these effects have shown them to be capable of reproducing the observed level of homogeneity required by the shallow negative metallicity gradients and flattening observed in outer galaxy discs. However, these focused studies generally do not account for the ongoing injection of enriched material. While they are useful for developing our understanding of the physical processes involved in mixing the ISM, they do not provide predictions that can be directly compared with observations.

More recently, Krumholz & Ting (2018) developed a theoretical model of stochastically forced diffusion that is capable of predicting the multiscale statistics in a disc galaxy metallicity distribution. Their formalism attempts to balance the stochasticity of injection events (e.g. from Type Ia and Type II supernovae, asymptotic giant branch stars, and neutron star mergers) with the mixing driven by interstellar turbulence that acts to homogenize the ISM within a simple uniform disc model. Their resulting fiducial dwarf galaxy and spiral galaxy models (dashed and dotted lines, Fig. 6) exhibit very good qualitative agreement with the trends we observe, including the high degree of correlation (30–50 per cent) observed on sub-kpc scales. Note

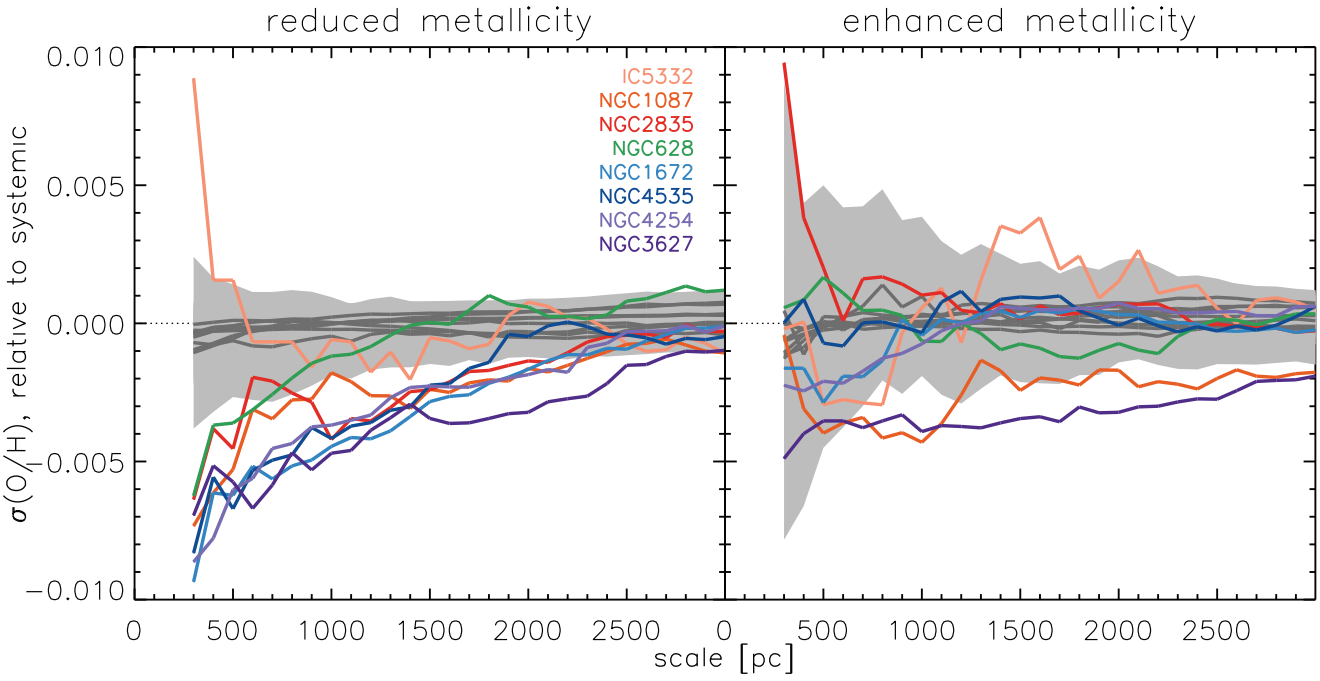

Figure 11. Centring on H II regions with either reduced (left) or enhanced (right) metallicity compared to the local average, we plot the median scatter in Δ(O/H) as a function of aperture scale. Grey bands show 1σ confidence intervals when randomly shuffling the metallicity offsets (over 100 realizations) to test how this analysis appears assuming the null hypothesis (that metallicities are uncorrelated). Dark grey lines are the median of this sample. Regions with reduced abundances show a similar trend for increased homogeneity on small scales as in Fig. 4. Only IC 5332 shows no such trend, but this galaxy also contains the smallest H II region sample size. Regions with enhanced abundances do not show a significant difference from the null hypothesis.

that the difference between the modelled and observed two-point correlation functions at the smallest separation scales reflects our observational limitations, given our H II region minimum separation length of ∼100 pc (roughly twice the physical resolution).

As the Krumholz & Ting (2018) formalism takes as an input a combination of gas fraction, gas-phase velocity dispersion, orbital time, and age of the system (their equations 109 and 111–113), we can apply their model directly to the annuli shown in Figs 8 and 9 to predicate both the correlation scale and the scatter in metallicity. Here, we assume a fixed system age of 10 Gyr, consistent with their fiducial spiral and dwarf galaxy models, and a fraction of metals retained in the disc of $f_d = 0.5$ (Tumlinson et al. 2011). Breaking the galaxies down into annuli enables us to better account for some of the radial structure in observed galaxies that is missed in this simple model.

As shown in Fig. 9, velocity dispersions within both the ionized and molecular phases positively correlate with the correlation scale. As our limited instrumental resolution in the ionized phase significantly increases our uncertainty in that measurement, we prefer to base this on the measured CO velocity dispersions ($\sigma_{\rm CO}$). However, the molecular gas is confined to a thin (∼150 pc) layer, that is not representative of the turbulence at play on ionized disc (∼1 kpc) scale heights. To convert $\sigma_{\rm CO}$ to this larger disc scale, we assume a power-law relation between scale length (L) and velocity dispersion as $\sigma \propto L^{1/2}$ (Falgarone et al. 2009; Klessen & Glover 2016). These scaled values agree reasonably well with the ionized gas velocity dispersions. Decreasing the target scale height by a factor of 2 results in slightly better agreement between observed and modelled correlation scales, but worse agreement between observed and modelled scatter in metallicities. Finally, we convert our 1D velocity dispersion ($\sigma_{\rm 1D}$, measured along the line of sight) into a 3D velocity dispersion ($\sigma_{\rm 3D}$) required as input for the model as $3 \times \sigma_{\rm 1D}^2 = \sigma_{\rm 3D}^2$.

Fig. 12 (top) directly compares the modelled 30 per cent correlation scales with our observations. A comparison at 50 per cent correlation scales shows similar results. In general, we see good agreement given the model's lack of large-scale inhomogeneities (such as spiral arms, bars, or differential rotation). NGC 1087 and NGC 4254 in particular appear to be systematically poorly modelled, with the correlation scale overpredicted by a factor of 3. NGC 4254 resides in the outskirts of the Virgo cluster, and shows strong evidence in the H I distribution of a recent interaction (Vollmer, Huchtmeier & van Driel 2005). Perhaps driven by these external dynamical effects, the ionized gas velocity dispersion across the galaxy is particularly high (Fig. 9, left), and drives the model towards longer correlation scales. On the other hand, NGC 3627 is also interacting (Haynes, Giovanelli & Roberts 1979), showing high velocity dispersion and hosting a strong stellar bar. It is unclear how much of the observed dispersion is actually turbulent (and hence would be expected to contribute to mixing) and how much is due to imperfect modelling of the bulk velocity on our resolution scale – given the high velocities associated with the bar and the substantial inclination. NGC 3627 is, however, better matched by the Krumholz & Ting (2018) model.

The predicted metallicity dispersions (Fig. 12, bottom) are typically about twice as large and show more variation between galaxies than what is observed, though it also clearly correlated with our observed σ(O/H). One free parameter in the Krumholz & Ting (2018) model is f_d, the fraction of metals retained in the disc. Based on Milky Way observations, they assume a fiducial value of 0.5. However, they note that the true value could be closer to unity. Increasing f_d to 0.90 (grey points, Fig. 12) decreases σ(O/H) by up to a factor of 2, bringing the model into better agreement with our observations. Such a large value of f_d might be plausible locally, but could be difficult to reconcile with the complicated picture of

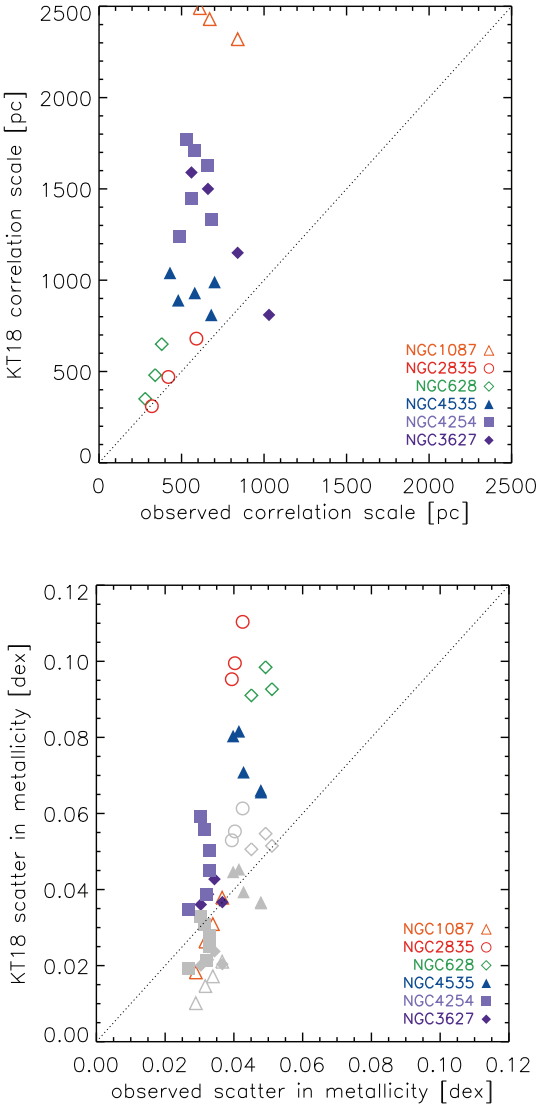

Figure 12. Comparison of the predictions from the Krumholz & Ting (2018) stochastically forced diffusion model with our observed values within 2 kpc wide annuli. Measured input parameters for their model in each annulus are shown in Fig. 9. A 1-to-1 line is drawn for comparison. Galaxies without H I (NGC 1672 and IC 5332) are omitted. Top: the correlation scale associated with the two-point correlation function at the 30 per cent level. Only NGC 4254, an interacting galaxy on the Virgo cluster outskirts, falls systematically in disagreement with their model. Bottom: the scatter in metallicity (σ(O/H)). Typical disagreement is by a factor of about 2 when assuming 50 per cent of metals are retained in the disc for elements whose primary origin site is Type II SNe. Increasing this fraction to 90 per cent (in grey) improves the agreement of the model with observations.

inflows, outflows, and gas recycling that regulate galaxy evolution (Tumlinson, Peeples & Werk 2017).

6 CONCLUSIONS

We examine the statistical distribution of gas-phase abundances within the ISM of eight nearby ($D < 17$ Mpc) galaxies. Each galaxy has hundreds to thousands of H II regions with measured 12 + log(O/H) based on strong-line metallicity prescriptions, as observed by the PHANGS-MUSE survey (Kreckel et al. 2019; Emsellem et al., in preparation). After fitting and removing a linear gradient in Paper I, we find rms scatter in residual metallicity of 0.04–0.05 dex. This dispersion becomes even lower (0.02–0.03 dex) when we consider the scatter on local 600 pc scales, indicating a remarkable level of chemical homogeneity.

We apply two-point correlation function statistics to directly measure the 30 per cent and 50 per cent correlation scales of the oxygen abundance in each galaxy, both globally and in 2 kpc wide annuli. These size scales correlate most strongly with the gas velocity dispersion, and weakly with SFR surface density. Our observed statistical correlation lengths are found to be in good agreement with predictions from the stochastically forced diffusion model of Krumholz & Ting (2018), despite the relative simplicity of their disc model.

H II regions with reduced abundances relative to the radial gradient show consistent and larger scale levels of homogeneity than the H II regions with enhanced abundances, suggesting the uniformity is driven predominantly by turbulence mixing material from large scales and potentially introducing relatively more pristine material from the halo. This is consistent with a spiral-arm-driven 'carousel' model of enrichment (Ho et al. 2017), where gas passing through the interarm region is slowly enriched and metallicities peak at the spiral arm (as shown in Paper I). However, enriched pockets of gas requiring at least ∼10 Myr time-scales to allow pollution (assuming a continuous star formation history) are observed to survive within this turbulent mixing environment. Overall, the high level of chemical homogeneity over large spatial scales demonstrates that efficient mixing within the ISM is a common feature in spiral galaxies.

ACKNOWLEDGEMENTS

We thank the anonymous referee for their comments which helped improve the clarity of the work. KK and FS gratefully acknowledges funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) in the form of an Emmy Noether Research Group (grant number KR4598/2-1, PI: Kreckel). SCOG and RSK acknowledge support from the DFG via SFB 881 'The Milky Way System' (project-ID 138713538; subprojects B1, B2, and B8) and from the Heidelberg cluster of excellence EXC 2181-390900948 'STRUCTURES: A unifying approach to emergent phenomena in the physical world, mathematics, and complex data', funded by the German Excellence Strategy. RSK furthermore thanks for funding from the European Research Council via the ERC Synergy Grant ECOGAL (grant 855130). ER acknowledges the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), funding reference number RGPIN-2017-03987. FB acknowledges funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 726384/Empire). JMDK and MC gratefully acknowledge funding from the DFG through an Emmy Noether Research Group (grant number KR4801/1-1). JMDK, MC, and JJK gratefully acknowledge funding from the DFG through the DFG Sachbeihilfe (grant number KR4801/2-1). JMDK gratefully acknowledges funding from the ERC under the European Union's Horizon 2020 research and innovation programme via the ERC Starting Grant MUSTANG (grant agreement number 714907). EW acknowledges support from the DFG via SFB 881 'The Milky Way System' (project-ID 138713538; subproject P2). TGW acknowledges funding from the ERC under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 694343).

This work was carried out as part of the PHANGS collaboration. Based on observations collected at the European Organisation for Astronomical Research in the Southern Hemisphere under ESO programme IDs 094.C-0623(A), 098.C-0484(A),

1100.B-0651(A) and 1100.B-0651(B). This paper makes use of the following ALMA data: ADS/JAO.ALMA#2012.1.00650.S, ADS/JAO.ALMA#2015.1.00925.S, ADS/JAO.ALMA#2015.1.00956.S, ADS/JAO.ALMA#2017.1.00392.S, ALMA is a partnership of ESO (representing its member states), NSF (USA), and NINS (Japan), together with NRC (Canada), MOST and ASIAA (Taiwan), and KASI (Republic of Korea), in cooperation with the Republic of Chile. The Joint ALMA Observatory is operated by ESO, AUI/NRAO, and NAOJ. The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc.

DATA AVAILABILITY

The data underlying this article are available in Kreckel et al. (2019).

REFERENCES

Arellano-Córdova K. Z., Rodríguez M., Mayya Y. D., Rosa-González D., 2016, MNRAS, 455, 2627
Bacon R., et al., 2010, in Ground-based and Airborne Instrumentation for Astronomy III, SPIE, Bellingham, p. 773508
Bacon R. et al., 2017, A&A, 608, A1
Berg D. A., Skillman E. D., Croxall K. V., Pogge R. W., Moustakas J., Johnson-Groh M., 2015, ApJ, 806, 16
Berg D. A., Pogge R. W., Skillman E. D., Croxall K. V., Moustakas J., Rogers N. S. J., Sun J., 2020, ApJ, 893, 96
Bolatto A. D., Wolfire M., Leroy A. K., 2013, ARA&A, 51, 207
Cappellari M., 2017, MNRAS, 466, 798
Cappellari M., Emsellem E., 2004, PASP, 116, 138
Chevance M. et al., 2020b, MNRAS, 493, 2872
Chung A., van Gorkom J. H., Kenney J. D. P., Crowl H., Vollmer B., 2009, AJ, 138, 1741
Colombo D. et al., 2014, ApJ, 784, 4
Croxall K. V., Pogge R. W., Berg D. A., Skillman E. D., Moustakas J., 2015, ApJ, 808, 42
Croxall K. V., Pogge R. W., Berg D. A., Skillman E. D., Moustakas J., 2016, ApJ, 830, 4
de Avillez M. A., Mac Low M.-M., 2002, ApJ, 581, 1047
Dekel A. et al., 2009, Nature, 457, 451
Della Bruna L. et al., 2020, A&A, 635, A134
Di Matteo P., Haywood M., Combes F., Semelin B., Snaith O. N., 2013, A&A, 553, A102
Dopita M. A., Sutherland R. S., 2003, Astrophysics of the Diffuse Universe, Springer, Berlin
Dopita M. A., Kewley L. J., Sutherland R. S., Nicholls D. C., 2016, Ap&SS, 361, 61
Falgarone E., Pety J., Hily-Blant P., 2009, A&A, 507, 355
Ferrière K. M., 2001, Rev. Mod. Phys., 73, 1031
Fitzpatrick E. L., 1999, PASP, 111, 63
Fraternali F., 2017, Gas Accretion via Condensation and Fountains. Springer, Cham, p. 323
Grand R. J. J. et al., 2016, MNRAS, 460, L94
Haynes M. P., Giovanelli R., Roberts M. S., 1979, ApJ, 229, 83
Ho I. T. et al., 2017, ApJ, 846, 39
Ho I. T. et al., 2018, A&A, 618, A64
Ho I. T. et al., 2019, ApJ, 885, L31
Hughes A. et al., 2013, ApJ, 779, 46
Jeffreson S. M. R., Kruijssen J. M. D., 2018, MNRAS, 476, 3688
Klessen R. S., Glover S. C. O., 2016, Saas-Fee Adv. Course, 43, 85
Klessen R. S., Hennebelle P., 2010, A&A, 520, A17
Klessen R. S., Lin D. N., 2003, Phys. Rev. E, 67, 046311
Kreckel K. et al., 2018, ApJ, 863, L21
Kreckel K. et al., 2019, ApJ, 887, 80
Kroupa P., 2001, MNRAS, 322, 231
Kruijssen J. M. D., Longmore S. N., 2014, MNRAS, 439, 3239
Kruijssen J. M. D., Schruba A., Hygate A. P. S., Hu C.-Y., Haydon D. T., Longmore S. N., 2018, MNRAS, 479, 1866
Krumholz M. R., Ting Y.-S., 2018, MNRAS, 475, 2236
Krumholz M. R., McKee C. F., Bland -Hawthorn J., 2019, ARA&A, 57, 227
Lang P. et al., 2020, ApJ, 897, 122
Larson R. B., 1981, MNRAS, 194, 809
Leroy A. K. et al., 2013, AJ, 146, 19
Leroy A. K. et al., 2019, ApJS, 244, 24
Levy R. C. et al., 2019, ApJ, 882, 84
Linsky J. L., 2003, Space Sci. Rev., 106, 49
Lubowich D. A., Pasachoff J. M., Balonek T. J., Millar T. J., Tremonti C., Roberts H., Galloway R. P., 2000, Nature, 405, 1025
Magrini L., Coccato L., Stanghellini L., Casasola V., Galli D., 2016, A&A, 588, A91
Marasco A., Fraternali F., Binney J. J., 2012, MNRAS, 419, 1107
Marino R. A. et al., 2013, A&A, 559, A114
Meidt S. E. et al., 2018, ApJ, 854, 100
Meidt S. E. et al., 2020, ApJ, 892, 73
Murphy E. J. et al., 2011, ApJ, 737, 67
Petit A. C., Krumholz M. R., Goldbaum N. J., Forbes J. C., 2015, MNRAS, 449, 2588
Pilyugin L. S., Grebel E. K., 2016, MNRAS, 457, 3678
Pilyugin L. S., Grebel E. K., Kniazev A. Y., 2014, AJ, 147, 131
Prochaska J. X., Wolfe A. M., 2009, ApJ, 696, 1543
Querejeta M. et al., 2015, ApJS, 219, 5
Rosales-Ortega F. F., Díaz A. I., Kennicutt R. C., Sánchez S. F., 2011, MNRAS, 415, 2439
Rosolowsky E., Simon J. D., 2008, ApJ, 675, 1213
Roy J. R., Kunth D., 1995, A&A, 294, 432
Sánchez S. F. et al., 2015, A&A, 573, A105
Sánchez S. F. et al., 2019, MNRAS, 484, 3042
Sánchez-Menguiano L. et al., 2016, ApJ, 830, L40
Sánchez-Menguiano L. et al., 2018, A&A, 609, A119
Sánchez-Menguiano L., Sánchez S. F., Pérez I., Ruiz-Lara T., Galbany L., Anderson J. P., Kuncarayakti H., 2020, MNRAS, 492, 4149
Sancisi R., Fraternali F., Oosterloo T., van der Hulst T., 2008, A&AR, 15, 189
Schruba A., Kruijssen J. M. D., Leroy A. K., 2019, ApJ, 883, 2
Smith B. D. et al., 2017, MNRAS, 466, 2217
Spitoni E., Cescutti G., Minchev I., Matteucci F., Silva Aguirre V., Martig M., Bono G., Chiappini C., 2019, A&A, 628, A38
Storey P. J., Zeippen C. J., 2000, MNRAS, 312, 813
Sun J. et al., 2018, ApJ, 860, 172
Sun J. et al., 2020, ApJ, 892, 148
Tremonti C. A. et al., 2004, ApJ, 613, 898
Tumlinson J. et al., 2011, Science, 334, 948
Tumlinson J., Peeples M. S., Werk J. K., 2017, ARA&A, 55, 389
Utomo D. et al., 2018, ApJ, 861, L18
Vazdekis A., Sánchez-Blázquez P., Falcón-Barroso J., Cenarro A. J., Beasley M. A., Cardiel N., Gorgas J., Peletier R. F., 2010, MNRAS, 404, 1639
Vogt F. P. A., Pérez E., Dopita M. A., Verdes-Montenegro L., Borthakur S., 2017, A&A, 601, A61
Vollmer B., Huchtmeier W., van Driel W., 2005, A&A, 439, 921
Walter F., Brinks E., de Blok W. J. G., Bigiel F., Kennicutt R. C., Jr, Thornley M. D., Leroy A., 2008, AJ, 136, 2563
Yang C.-C., Krumholz M., 2012, ApJ, 758, 48

APPENDIX A: ALTERNATE METALLICITY PRESCRIPTIONS

We reproduce two key figures from this paper (Figs 4 and 6) using three additional strong-line metallicity prescriptions, also presented in Paper I (Appendix C). Due to the wavelength coverage of MUSE, a limited number of calibrations are available.

Our preferred calibration (Scal; Pilyugin & Grebel 2016) is based on an empirical comparison with H II regions where direct temperature measurements are available, and relies on a combination

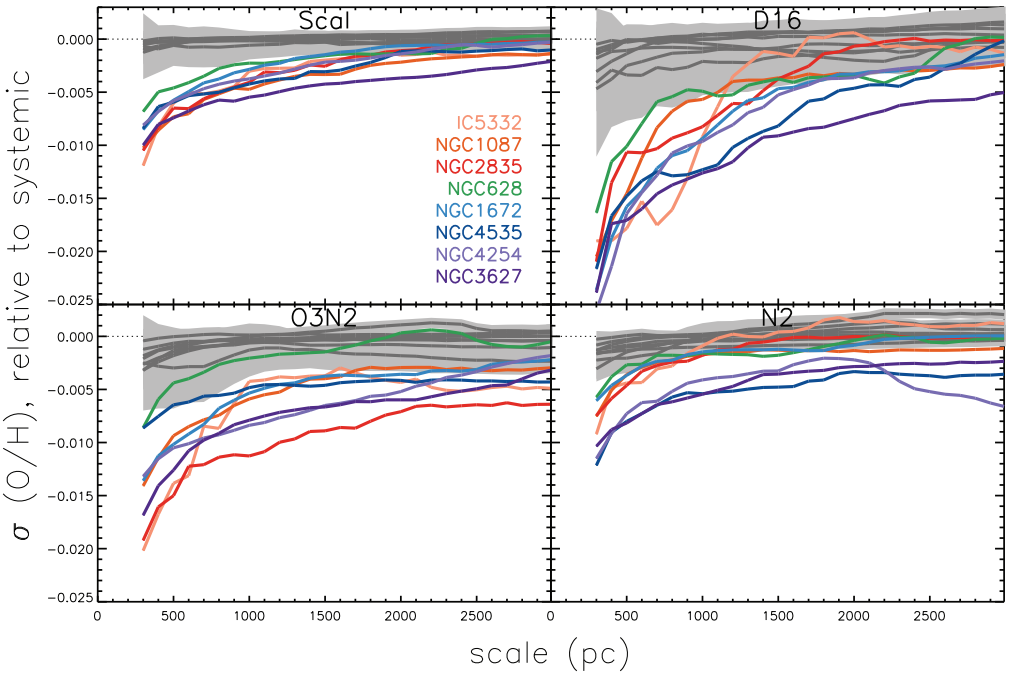

Figure A1. Revisiting Fig. 4, using different metallicity calibrations, to show the median scatter in Δ(O/H) over a given spatial scale, for apertures centred on H II regions. Trends in each galaxy are measured with respect to the systemic scatter. Our preferred empirical calibration (Scal; Pilyugin & Grebel 2016) shows similar qualitative trends as the three other calibrations explored in Paper I. This includes photoionization model-based calibrations (D16; Dopita et al. 2016) and alternate empirical prescriptions (O3N2, N2; Marino et al. 2013). In all calibrations, we see increased homogeneity on small, scales, in excess of what is predicted by the null hypothesis (in grey).

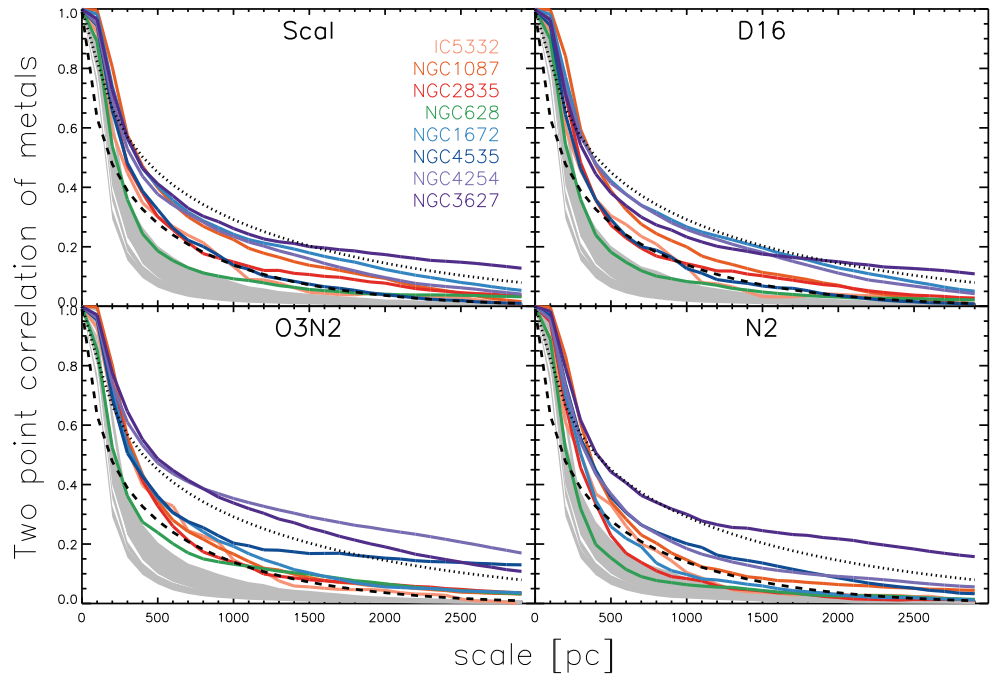

Figure A2. Revisiting Fig. 6, using different metallicity calibrations, to show the two-point correlation function of metals as a function of scale. Our preferred empirical calibration (Scal; Pilyugin & Grebel 2016) shows similar qualitative trends as the three other calibrations explored in Paper I. While the photoionization model-based calibration D16 (Dopita et al. 2016) shows larger global scatter (σ(O/H); Paper I, table 3), quantitatively the correlation scales agree quite well with the results reported here. The empirical O3N2 and N2 calibrations (Marino et al. 2013) are unable to account for variations in ionization parameter, and both show on average shorter correlation scales (but still in excess of the null hypothesis, in grey).

of three emission-line ratios. This relatively high dimensionality in the prescription helps break the degeneracy between ionization parameter and metallicity variations. For comparison, we show here results using two other empirical calibrations (O3N2 and N2; Marino et al. 2013), which include fewer diagnostic ratios and thus are less able to account for the influence of ionization parameter variations. We further compare with a photoionization model based calibration (D16; Dopita et al. 2016), which by design is also intended to break the degeneracy between ionization parameter and metallicity. In Paper I, we found the Scal and D16 calibrations to show very similar qualitative trends in azimuthal variations, though D16 produced significantly higher global scatter in metallicity variations, σ(O/H).

As is seen in Figs A1 and A2, the Scal-based results presented in this paper show good qualitative agreement with all other calibrations.

[1]*Astronomisches Rechen-Institut, Zentrum für Astronomie der Universität Heidelberg, Mönchhofstraße 12-14, D-69120 Heidelberg, Germany*
[2]*Max-Planck Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany*
[3]*The Observatories of the Carnegie Institution for Science, 813 Santa Barbara Street, Pasadena, CA 91101, USA*
[4]*Departamento de Astronomía, Universidad de Chile, Casilla 36-D, Santiago, Chile*
[5]*Institut für theoretische Astrophysik, Zentrum für Astronomie, Universität Heidelberg, Albert-Ueberle-Str. 2, D-69120 Heidelberg, Germany*
[6]*International Centre for Radio Astronomy Research, University of Western Australia, 7 Fairway, Crawley, 6009 WA, Australia*
[7]*4-183 CCIS, University of Alberta, Edmonton, AB T6G 2E1, Canada*
[8]*Argelander-Institut für Astronomie, Universität Bonn, Auf dem Hügel 71, D-53121 Bonn, Germany*
[9]*Centro de Astronomìa (CITEVA), Universidad de Antofagasta, Avenida Angamos 601, Antofagasta, Chile*
[10]*Department of Physics and Astronomy, University of Wyoming, Laramie, WY 82071, USA*
[11]*IPAC, California Institute of Technology, Pasadena, CA 91125, USA*
[12]*European Southern Observatory, Karl-Schwarzschild-Straße 2, D-85748 Garching bei München, Germany*
[13]*ENS de Lyon, CNRS, Centre de Recherche Astrophysique de Lyon, Univ. Lyon, Univ. Lyon1, UMR5574, F-69230 Saint-Genis-Laval, France*
[14]*Research School of Astronomy and Astrophysics, Australian National University, Weston Creek 2611, Australia*
[15]*Interdisziplinäres Zentrum für Wissenschaftliches Rechnen, Universität Heidelberg, Im Neuenheimer Feld 205, D-69120 Heidelberg, Germany*
[16]*Department of Astronomy, The Ohio State University, 140 West 18th Ave, Columbus, OH 43210, USA*
[17]*Sydney Institute for Astronomy, School of Physics, A28, The University of Sydney, NSW 2006, Australia*
[18]*Sterrenkundig Observatorium, Universiteit Gent, Krijgslaan 281 S9, B-9000 Gent, Belgium*
[19]*Departamento de Física Teórica, Universidad Autónoma de Madrid, Cantoblanco, E-28049 Madrid, Spain*
[20]*Center for Astrophysics and Space Sciences, Department of Physics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA*
[21]*Max-Planck Institut für Extraterrestrische Physik, Giessenbachstraße 1, D-85748 Garching, Germany*

This paper has been typeset from a TEX/LATEX file prepared by the author.