Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **539,** 1723–1737 (2025) https://doi.org/10.1093/mnras/staf593
Advance Access publication 2025 April 21

QUOKKA-based understanding of outflows (QED) – II. X-ray metallicity gradients as a signature of galactic wind metal loading

Rongjun Huang(黄镕钧) [1,2]★ Aditi Vijayan [1,2]★ and Mark R. Krumholz [1,2]★

[1]*Research School of Astronomy and Astrophysics, Australian National University, Cotter Road, Weston Creek, ACT 2611, Australia*
[2]*ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Australia*

Accepted 2025 April 8. Received 2025 April 7; in original form 2024 September 29

ABSTRACT

Supernova-driven galactic outflows play a vital but still poorly understood role in galactic chemical evolution, and one of the largest uncertainties about them is the extent to which they consist of supernova ejecta that are unmixed, or only poorly mixed, with the remainder of the interstellar medium (ISM). Simulations of wind launching make a range of predictions about the extent of mixing between the wind and the ISM, but thus far these have proven challenging to test observationally. In this study, we post-process high-resolution simulations of outflows from the QED simulation suite to generate synthetic X-ray spectra from galactic winds, which we then analyse using standard observational procedures, in order to search for detectable markers of wind mixing. Our synthetic observations reveal that partially mixed winds show significant and detectable metallicity gradients when viewed edge-on, with metallicity decreasing away from the central galactic disc. We explore how this signature results from imperfect mixing and the extent to which measurements of it can be used to diagnose the level of mixing in winds. We determine the signal-to-noise ratio (SNR) requirements for such measurements to be reliable, and provide a simple quantitative model that can be used to connect metallicity gradients to mixing between the hot ($T > 10^6$ K) and cold ($T < 10^4$ K) phases in observations that reach the required SNR, providing a framework to interpret current and future observations.

Key words: methods: data analysis – methods: numerical – X-rays: galaxies.

1 INTRODUCTION

Metals are synthesized in the cores of stars that, at least in spiral galaxies such as the Milky Way, reside in a thin disc. However, observations reveal that these metals are not confined to the disc. Instead, ∼50 per cent of the mass of metals like oxygen is found in galactic haloes, at distances as large as ∼100 kpc from the central galaxy (Tumlinson et al. 2011; Peeples et al. 2014). This raises the intriguing question of how metals are transported from the disc to the halo. The most plausible explanation is that they are carried by galactic outflows, powerful winds driven by supernova (SN) feedback that can transport material, momentum, and energy from the disc to the halo.

Physically, these outflows are driven by SN thermal energy injection that heats up the surrounding gas, driving it to expand and break out of the disc. Once hot bubbles break out, the momentum-carrying gas expands freely and eventually ends up in the circumgalactic medium (CGM) around the galaxy. Because this hot gas is generated as a direct result of SN activity, it is expected to be loaded with metals and therefore enriches the CGM with metals. However, the balance in the gas escaping the galaxy between unmixed SN ejecta and swept-up ambient gas from the galactic interstellar medium (ISM) – which is also metal-enriched but much less so than the fresh supernova ejecta – is highly uncertain. It is also highly important: the fraction of SN-produced metals that are promptly lost to the outflow rather than being retained in the disc is a crucial free parameter in models of galactic chemical evolution (e.g. Peeples & Shankar 2011; Forbes, Krumholz & Speagle 2019; Telford et al. 2019; Johnson & Weinberg 2020; Sharda et al. 2021; Kravtsov & Manwadkar 2022).

One potential way to quantify the fraction of SN-injected metals carried by the outflowing gas is by studying the diffuse X-ray emission around star-forming galaxies. Such emission is linked to star formation processes occurring within the disc (Strickland et al. 2004; Grimes et al. 2005; Tüllmann et al. 2006; Owen & Warwick 2009; Li & Wang 2013; Wang et al. 2016), and is useful for studying metallicity due to the large number of abundance-diagnostic lines that appear in soft X-rays. In particular, the starburst M82 serves as a textbook case of SN-driven outflows and much attention has been given to measuring the metallicity of the outflows launched from the central region in the X-ray (e.g. Read & Stevens 2002; Origlia et al. 2004; Ranalli et al. 2008; Konami et al. 2011; Lopez et al. 2020; Fukushima, Kobayashi & Matsushita 2024). All these studies have revealed that the outflows M82 have non-uniform metallicities. Read & Stevens (2002) and Origlia et al. (2004) found that the outflows are enhanced in α/Fe relative to the disc, while Lopez et al. (2020) measured the gradients of metallicity along the minor axis of M82 and reported a clear trend of declining abundances and gas temperatures with distance away from the central galaxy for several α elements. They hypothesize that the outflow profile of M82 is a result of mixing between different temperature phases and not simply because of the adiabatic expansion of hot wind. However, the nature of X-ray emission poses challenges for interpretation. For example

★ E-mail: U6569836@anu.edu.au (RH); Aditi.Vijayan@anu.edu.au (AV); Mark.Krumholz@anu.edu.au (MRK)

spectral fitting relies on simplifying assumptions about temperature profiles of X-ray emitting gas (Vijayan & Li 2022). One way to overcome such limitations is to provide a theoretical basis for the interpretation of metallicity gradients using simulations.

While M82 is the best-studied case, thanks to its proximity and nearly edge-on geometry, there is also substantial evidence for metallicity variations in the winds of other galaxies, and in phases other than the hot, X-ray-emitting one. Lopez et al. (2023) find a metallicity gradient in the hot phase of the wind of NGC 253 that is similar to that in M82. Cameron et al. (2021) identify a clear negative metallicity gradient along the minor axis of the wind of edge-on starburst galaxy Markarian 1486 using direct (electron temperature-based) measurements of metallicity in the warm ionized phase. Martin, Kobulnicky & Heckman (2002) report significant α/Fe enhancement in the hot wind phase of the dwarf starburst NGC 1569, but Hamel-Bravo et al. (2024) find no comparable enhancement in the warm ionized gas, suggesting a composition difference between these two components. Xu et al. (2022) also conclude that there is likely a significant difference between the metal content of the warm ionized and hot components of the wind in their absorption spectroscopic study of 45 low-redshifts starbursts drawn from the Cosmic Origins Spectrograph Legacy Spectroscopic Survey (CLASSY) sample (Berg et al. 2022). Chisholm, Tremonti & Leitherer (2018) carry out absorption spectroscopy in a sample of seven local galaxies and find evidence that metallicities in the warm ionized wind phase are significantly enhanced compared to the interstellar media of the galaxies driving those winds. Significantly, their sample contains ordinary star-forming galaxies in addition to starbursts, and there is no significant difference in the metal enhancement between the two groups. This suggests that, while most studies of metal-enriched winds to date have focused on starbursts simply because their winds are bright and easy to observe, metal-enrichment is a phenomenon common to all galactic winds, not just the winds of starbursts.

There have also been numerous attempts to use simulations to study the metal content of galactic winds. This problem is generally impossible to address in cosmological simulations, which lack the resolution to capture the phase structure of winds without relying on sub-grid models (e.g. Smith et al. 2024). However, even excluding cosmological simulations, there are a range of approaches that represent different trade-offs between volume, resolution, and physical complexity. At the finest resolution but lowest volume end are 'tall-box' simulations (e.g. Kim et al. 2020; Rathjen et al. 2021) that simulate only a portion of a galactic disc, while at the larger volume, lower resolution end are simulations of entire isolated dwarf galaxies (e.g. Melioli, de Gouveia Dal Pino & Geraissate 2013; Emerick et al. 2018; Schneider, Robertson & Thompson 2018; Emerick, Bryan & Mac Low 2019, 2020; Schneider et al. 2020; Andersson et al. 2023; Rey et al. 2024; Steinwandel et al. 2024a,b). Similarly, there are a range of levels of physical complexity, with some simulations using relatively simple treatments of ISM physics in order to maximize resolution (e.g. Melioli et al. 2013; Schneider et al. 2018, 2020) while others include live treatments of physical processes such as radiative transfer (e.g. Rey et al. 2024) and cosmic rays (e.g. Rathjen et al. 2021) or including multiple elements with different nucleosynthetic origins (e.g. Emerick et al. 2018, 2019, 2020). Moreover, recent high-resolution dwarf galaxy simulations have incorporated multiple feedback channels (e.g. Gutcke et al. 2021; Smith et al. 2021; Deng et al. 2024) or focused on star-by-star physics in isolated systems (e.g. Hu et al. 2016, 2017, 2023a; Hu 2019; Hu, Sternberg & van Dishoeck 2023b), highlighting the importance of resolving SN remnants and properly modelling radiation in shaping multiphase, metal-enriched outflows. However, there have thus far been limited efforts to carry out detailed observational comparisons between these simulations and X-ray observations; the primary exception is the work of Schneider & Mao (2024), who generate detailed synthetic X-ray maps from their simulations. However, those simulations do not include metals, and thus are not directly usable for comparing to observations of metallicity variation in winds such as those summarized above.

Part of the reason for this relative paucity of studies of the metal content of winds including detailed observational comparisons is the numerical difficulty of the problem. Simulations targeting these observations require very high resolution because mixing between the hot gas that emits in X-rays and the cooler gas that can be studied in optical or radio is a critical process in determining both the temperature and the metal content of the hot phase. The rate of mixing between the hot and cooler phases is easily affected by poor resolution, with Lagrangian methods tending to undermix without an explicit sub-grid diffusion model (Shen, Wadsley & Stinson 2010; Escala et al. 2018), and Eulerian methods tending to overmix due to numerical diffusion. At a minimum, no simulation that has to rely on sub-grid models of SN explosions to avoid overcooling can reliably study metal mixing, a condition that rules out even zoom-in cosmological simulations, and even properly resolving the Sedov–Taylor phase may be insufficient to yield converged results for clustered SNe (Gentry et al. 2019). At the same time, observations of wind metal content generally target regions extending to several kpc off galactic discs, and thus simulations attempting to compare to these observations must have similar extents. Few existing simulations meet these simultaneous requirements.

In this work, we use QED[1] simulations, introduced in Vijayan, Krumholz & Wibking (2024, hereafter QED I), to study metals in galactic winds with mock X-ray observations. QED is a suite of 3D hydrodynamic simulations of galactic winds generated by the new code QUOKKA (Wibking & Krumholz 2022; He, Wibking & Krumholz 2024), a state-of-the-art adaptive mesh refinement (AMR) radiation-hydrodynamics (RHD) tool optimized for graphics processing units (GPUs). In terms of the taxonomy of simulation approaches discussed above, the QED suite lies at the maximum resolution end of the spectrum: tall box rather than whole-galaxy, and including a relatively simple treatment of ISM physics and feedback that includes only driving by randomly distributed SNe. QED's primary virtue is the very large dynamic range afforded by GPU acceleration: the simulations feature a uniformly high resolution of 2 pc that allows us to accurately follow the metal exchanges taking place in the multiphase outflows. QED I demonstrates that, at the 2 pc resolution achieved in the QED suite, the wind mass flux, metal flux, and metallicity as a function of height off the mid-plane are all numerically converged. At the same time, the simulation domain extends to ± 4 kpc above and below the galactic mid-plane, enabling us to follow trends up to a few kpc above the disc, capturing the off-disc regions typically targeted in observations. Of course QED – like all tall box simulations – still captures only a small fraction of the volume of a real galactic wind in a galaxy like the Milky Way, which has an area much larger than 1 kpc^2 area of the QED tall box. However, it is reasonable to hypothesize that the QED simulation box is statistically representative of the full wind,

[1]QUOKKA-based Understanding of Outflows Derived from Extensive, Repeated, Accurate, Thorough, Demanding, Expensive, Memory-consuming, Ongoing Numerical Simulations of Transport, Removal, Accretion, Nucleosynthesis, Deposition, and Uplifting of Metals (QUOD ERAT DEMONSTRANDUM, or QED).

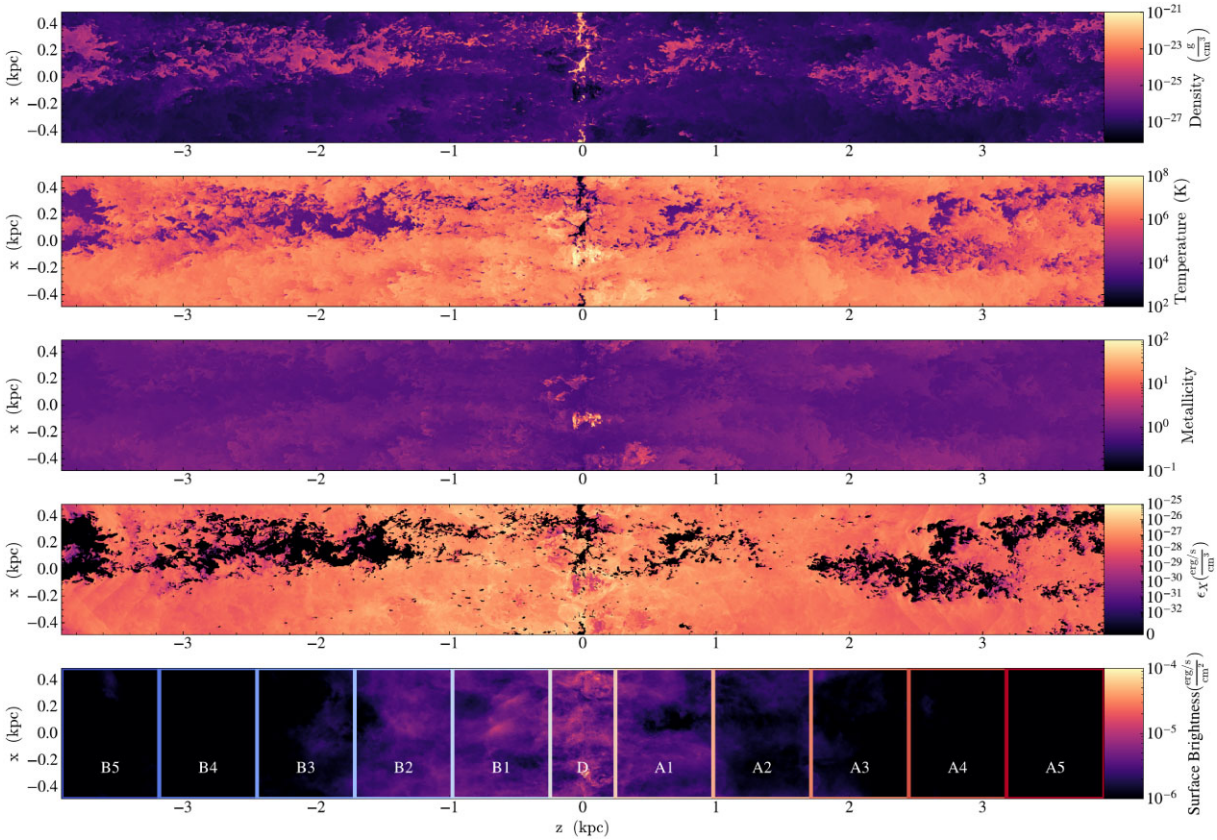

Figure 1. Slices in the xz plane from a snapshot of the QED simulations (QED I) at 110 Myr. From the top to bottom we show gas density, temperature, metallicity, and X-ray emissivity field per unit volume in the 0.4–2 keV band in the first four panels. The bottom panel shows the X-ray surface brightness, overlaid with rectangles highlighting the 11 regions over which we average to produce the spectra used for analysis. The metallicity and X-ray luminosity panels are computed using $Z_{\rm bg} = 0.5 Z_{\odot}$.

and is therefore suitable for use in simulated observations provided that we compensate for the smaller volume. Finally, the QUOKKA code on which the QED simulations are based uses a very high resolution hydrodynamics scheme based on a PPM solver with a monotonized-central (MC) limiter, using a combination of a dual energy formalism and first-order flux correction to maintain stability in problematic cells (Wibking & Krumholz 2022); this scheme should be significantly less dissipative, and thus better able to follow phase and metal mixing, than the more dissipative methods used in some earlier simulations.

This paper is organized as follows: Section 2 provides a brief summary of the set-up of the QED simulations, and the method we use to generate synthetic observations from them and then analyse them using the same techniques that are routinely applied to real observations. Section 3 presents our main results related to fits of the mock observations, and Section 4 discusses the interpretation of the results and their implications. We outline our major conclusions in Section 5.

2 METHODS AND DATA

In this section, we first introduce the QED simulations (Section 2.1) and then explain our methods for generating synthetic X-ray data (Section 2.2), and analysing those synthetic data to extract physical parameters (Section 2.3).

2.1 The QED simulations

QED is a suite of fixed-grid simulations of a $\sim 1 \times 1$ kpc^2 patch of the Milky Way disc; QED I presents results and a convergence study for Solar neighbourhood conditions, while Vijayan, Krumholz & Wibking (2025) extend the model grid to a wide range of galactic environments. In this paper, we use the highest-resolution case presented in QED I, and we refer readers to that paper for full details of the numerical set-up, which we only summarize here. The simulation has a simulation box centred on the disc mid-plane that extends to ± 4 kpc and a uniform resolution of $512 \times 512 \times 4096$, giving a grid spacing of 2 pc, and is run for 116 Myr, by which point the outflow has reached statistical steady-state. We show mass outflow rates and loading factors as a function of time and simulation resolution to demonstrate convergence in Appendix A. The gas is confined by the gravity of the stellar disc and dark matter halo and driven by SNe, which occur at a constant rate calibrated to the Kennicutt (1998) relation and distributed exponentially in height around the disc with a scale height of 150 pc. SNe are treated as injections of 10^{51} erg of pure thermal energy; the resolution is high enough that the Sedov–Taylor phase is well-resolved, and no sub-grid models are required. Gas cooling is implemented through a tabulated cooling function derived from the GRACKLE cooling model (Smith et al. 2017). After nearly 100 Myr of SN feedback, the simulation reaches quasi-steady-state outflow rates. We show $y = 0$ density and temperature slices from a snapshot in 110 Myr in the top two panels of Fig. 1. The outflows clearly possess a multiphase structure

characterized by a broad range of temperatures from <10 K to $> 10^7$ K, with high (low) temperature gas being associated with low (high) densities.

Metals in the simulations can be either injected by SNe or present in the initial conditions. For the former, each SN injects a fixed mass $\Delta M_{\rm SN} = 1$ M$_\odot$ of oxygen into the same cell in which it deposits thermal energy, which is subsequently advected as a passive scalar. The second source of metals is the initial background metallicity of the ISM present at the start of the simulation, which we parametrize by $Z_{\rm bg}$. We scale it to Solar abundances, so the initial oxygen mass density in a cell of total mass density ρ is $\rho_{\rm O} = \rho(Z_{\rm bg}/Z_\odot)Z_{\rm O,\odot}$, where $Z_\odot$ denotes Solar metallicity and we set the Solar oxygen abundance to $Z_{\rm O,\odot} = 8.6 \times 10^{-3}$. We treat $Z_{\rm bg}$ as an independent parameter that we can alter in the post-processing of the simulations.[2] The total metal density of a cell is the sum of the background and SN-added contributions – see section 2.3.1 of QED I for full details of the computation. Once we have computed the oxygen metallicity in a cell, we assume Solar-scaled abundances for all other elements. This assumption is not entirely realistic, since we are modeling only Type II SNe that produce predominantly α elements rather than Type Ia SN that produce iron peak elements, and thus our added metal field should have a different α/Fe ratio than our background field; we discuss the implications of this in more detail below.

In the third panel from the top of Fig. 1, we show the metallicity of the same slice whose density and temperature are shown in the top two panels. We note here that higher metallicity is correlated with higher temperatures and lower densities tracing the SN-processed gas. Lower metallicity mainly follows the high-density warm gas.

2.2 Generation of synthetic X-ray data

We use the X-ray post-processing packages PYXSIM and SOXS (ZuHone & Hallman 2016; ZuHone et al. 2023) based on the YT analysis and visualization suite (Turk et al. 2011) to generate synthetic observations from a simulation snapshot via a procedure we describe in this Section.

2.2.1 *Generating photons using PYXSIM*

The PYXSIM package generates a set of X-ray photons from each cell in the simulation under the assumption of a collisionally ionized and optically thin plasma, following the ASTROPHYSICAL PLASMA EMISSION CODE (APEC; Smith et al. 2001). Because we are interested in the diffuse X-ray emission from galactic winds, we limit the photon energies to the soft X-ray band, with energies 0.4–2.0 keV; this is the energy range over which *Chandra* is the most sensitive, contains most (but not all) of the important diagnostic lines for metallicity, and in practice is usually the only usable band for spatial studies of galactic winds because emission from the relatively harder X-ray band (2–5 keV) is significant compared to the soft X-ray contribution only in regions close to the disc, and drops sharply away from it (Vijayan et al. 2018). One complication in generating synthetic observations from our simulation is that we simulate only a 1 kpc^2 patch that represents a small fraction of an actual Milky Way-sized galaxy (typical area closer to 100 kpc^2).[3] In order to compensate for this, we tile 100 copies of our simulation box together in order to generate our synthetic spectra – that is, we generate 100 times as many X-ray photons from each cell per unit time as would be expected, on the basis that a realistically large galaxy would contain 100 times as many cells with similar physical conditions.

Table 1. Parameters for the PYXSIM package. From top to bottom, these are the emission model, energy range, exposure time, area to collect the photons, distance of the object, RA and Dec. of the pointing in degrees, and absorption column.

Parameter	Value
Emission model	apec
Energy range (keV)	0.4–2.0
Exposure time (Ms)	10
Collecting area (cm^2)	$\pi \times 60^2$
Distance (kpc)	500
Sky position (RA, Dec.)	[30., 45.]
$N_{\rm H}$ (cm^{-2})	None

In addition to gas density and temperature, PYXSIM also requires metallicity. As discussed in Section 2.1, this depends on the choice of initial background metallicity $Z_{\rm bg}$; we carry out our analysis for the cases $Z_{\rm bg} = 0$, 0.2, 0.5, 1, and 2, providing a wide baseline. We summarize our choices for the other parameters required by PYXSIM, including the exposure time, collecting area, and distance to source, in Table 1. Our choice of collecting area is typical of *Chandra*, and our distance is roughly the distance to M82, and is similar to the distances of other nearby galaxies whose wind metallicities have been measured in X-rays (e.g. NGC 1569 – Martin et al. 2002; NGC 253 – Lopez et al. 2023). In order to get a 'pure' unadulterated spectrum, we disable absorption, backgrounds, and foregrounds, and we use a large exposure time of 10 Ms. We will consider the effects of smaller and more realistic signal-to-noise ratios (SNRs) in Section 3.4. We do not seek to explore the effects of confusion due to absorption, backgrounds, and foregrounds because these are all highly variable across the sky, and thus it is not possible to make general statements about the extent to which they can be corrected and the level of observational uncertainty introduced by the need to make such corrections.

Fig. 1 shows the 0.4–2.0 keV X-ray luminosity per unit volume (ϵ_X, second from the bottom panel) and the surface brightness (bottom panel) as computed by PYXSIM for the same 110 Myr snapshot shown in the upper panels, and for $Z_{\rm bg} = 0.5Z_\odot$. As expected, the X-ray luminosity is spatially correlated with gas at temperatures $\gtrsim 10^5$ K, and for gas at these temperatures is positively correlated with density.

2.2.2 *Characteristics of the X-ray emitting gas*

Before proceeding to simulated observations, it is important to examine the characteristics of the X-ray emitting gas, both to ensure that they are physically reasonable and to ensure that we adequately resolve the emitting regions in our simulations. The reason that resolution is a potential concern is that, thanks to the strong density and temperature dependence of X-ray emissivity, in

[2]We note here that in QED I all gas cools at Solar metallicity, irrespective of the local metallicity. This is the downside to the flexibility we gain by our purely passive treatment of metals.

[3]Conversely, our simulation volume would be more realistic for a circumnuclear starburst such as M82. However, in such a galaxy the wind would be much denser and more powerful than in our Solar neighbourhood set-up, and thus much brighter in the X-rays.

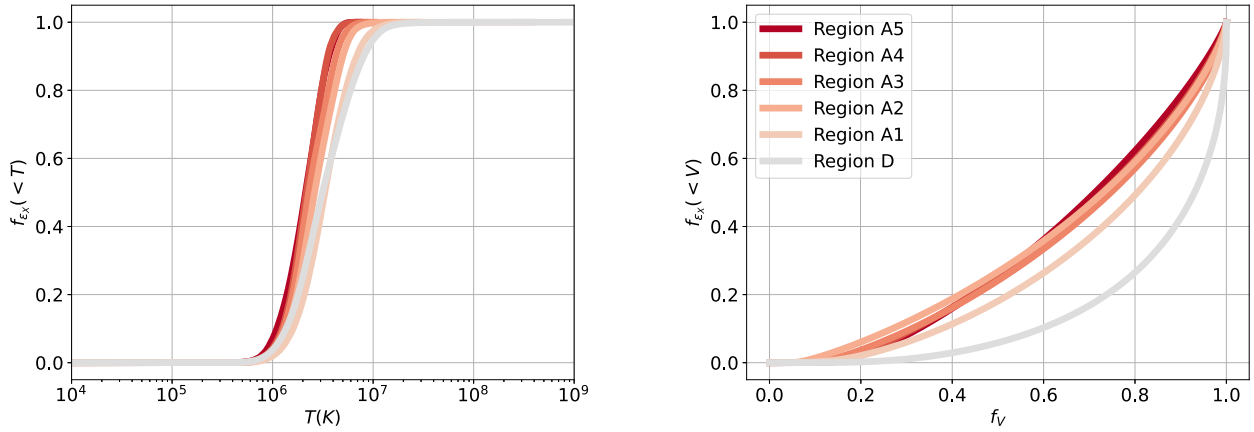

Figure 2. Cumulative distribution of X-ray emission in the 0.4–2.0 keV band with respect to gas temperature (left) and volume (right). Different colours correspond to different regions within the simulation, as explained in the main text; we show only the 'A' regions at positive z and omit the 'B' regions at negative z to reduce clutter in the plot, but the results for the 'A' and 'B' regions are qualitatively identical. We see that the X-ray emitting gas lies within a narrow temperature range ($\approx 10^6 - 10^7$ K), but that the emitting regions are reasonably well-resolved, with 50 per cent of the emission coming from $\approx$ 30 per cent of the volume in all regions but D, where the emission arises from $\approx$10 per cent of the volume.

a multiphase medium such as our simulated wind, X-ray emission can be dominated by relatively thin interface layers between the hot volume-filling phase and the cooler entrained gas that can be hard to resolve (Toalá & Arthur 2018; Vijayan et al. 2018).

To this end, we examine the cumulative distribution functions (CDFs) of X-ray emission with respect to gas temperature T and volume V for the same snapshot (and using the same $Z_{\rm bg} = 0.5Z_\odot$ as shown in Fig. 1. For the purpose of analysis discussed here and in subsequent Sections, we partition the simulation domain into 11 regions, mirroring the approach used by Lopez et al. (2020); we mark our regions in the bottom panel of Fig. 1. The central region, labelled 'D', corresponds to the area where SNe are injected and extends up to a height of 489.03 pc (i.e. 256 cells), while the remaining 10 regions ('A/B1' to 'A/B5') cover the outflow zone, each spanning 733.55 pc (i.e. 384 cells) in height.

We show the CDFs we obtain for each region in Fig. 2. The two panels of this figure show the fraction of total X-ray emission in the 0.4–2.0 keV band arising from gas with temperature $<T$ (left panel) and from a given fraction of total volume (right). As expected based on the models of Toalá & Arthur (2018), the left panel of this figure shows that X-ray emission in our simulation is dominated by moderately (∼few 10^6 K) rather than extremely ($\gtrsim 10^7$ K) hot gas. The right panel, however, shows that the volume occupied by this gas is substantial, and is thus well-resolved within our simulations. Quantitatively, except in region 'D', we find that $\approx$ 50 per cent of the emission emerges from 30 per cent of the volume. Given that our simulation contains a total of $\approx 10^9$ computational cells, even 30 per cent of this number constitutes very good resolution: quantitatively, 30 per cent of the volume corresponds to a resolution of roughly 685^3 cells. Resolution is slightly worse in region 'D', where 50 per cent of the emission emerges from just under 10 per cent of the volume, but even 10 per cent of our simulation volume corresponds to 475^3 cells. The lower volume filling fraction in region 'D' is not surprising since this region is dominated by fresh SNe ejecta with very high metallicities. We therefore conclude that the emitting regions in our simulations are relatively well-resolved.

2.2.3 *Generating mock observations using* SOXS

Once the X-ray photons are generated, we use SOXS to produce mock images and spectra as they would be observed by the *Chandra X-ray Observatory*. SOXS uses a subset of these photons to project on the plane of the instrument and convolve with the instrument response to produce an events file. We use the response matrix (specified by Auxiliary Response File and Redistribution Matrix File) of the ACIS-I instrument which has a field of view (FoV) of 16.9 × 16.9 arcmin^2 and a spectral resolution of 130 eV at 1.49 keV. The mock SOXS observation also needs an exposure time; we carry out our post-processing for exposure times of 10, 1, 0.1, and 0.01 Ms in order to generate results with a range of SNRs. We summarize the parameters we use for the SOXS post-processing in Table 2.

Table 2. Parameters for the SOXS package. From top to bottom, these are the energy range, exposure time, RA and Dec. of the pointing in degrees, and instrument specification.

Parameter	Value
Energy range (keV)	0.4–2.0
Exposure time (Ms)	10, 1, 0.1, and 0.01
Sky position (RA, Dec.)	[30., 45.]
Instrument	chandra_acisi_cy22

Fig. 3 shows a synthetic X-ray image for the same snapshot as Fig. 1 as it would be observed by *Chandra* using the ACIS-I configuration. The axes represent the RA and Dec. for a galaxy arbitrarily located at [30°, 45°], while the colour scale indicates photon counts per pixel integrated over energies from 0.4 to 2.0 keV, computed using the 10 Ms results. We plot the spectra of the regions D and A1-5, again for the example case of the 110 Myr snapshot with $Z_{\rm bg} = 0.5Z_\odot$, in Fig. 4.

2.3 Fitting the X-ray spectrum

The final step in our post-processing pipeline is to fit the synthetic spectra from 0.4 to 2 keV using SHERPA, following the procedure commonly used for observations (e.g. Lopez et al. 2020). We carry out this fit assuming that the emitting medium can be described as a superposition of thermal plasmas, each characterized by a distinct temperature but with shared elemental abundances. The standard observational practice is to use one, two, or three temperature components in such fits (though of course, this is an oversimplification since the outflow is multiphase and follows a smooth distribution in

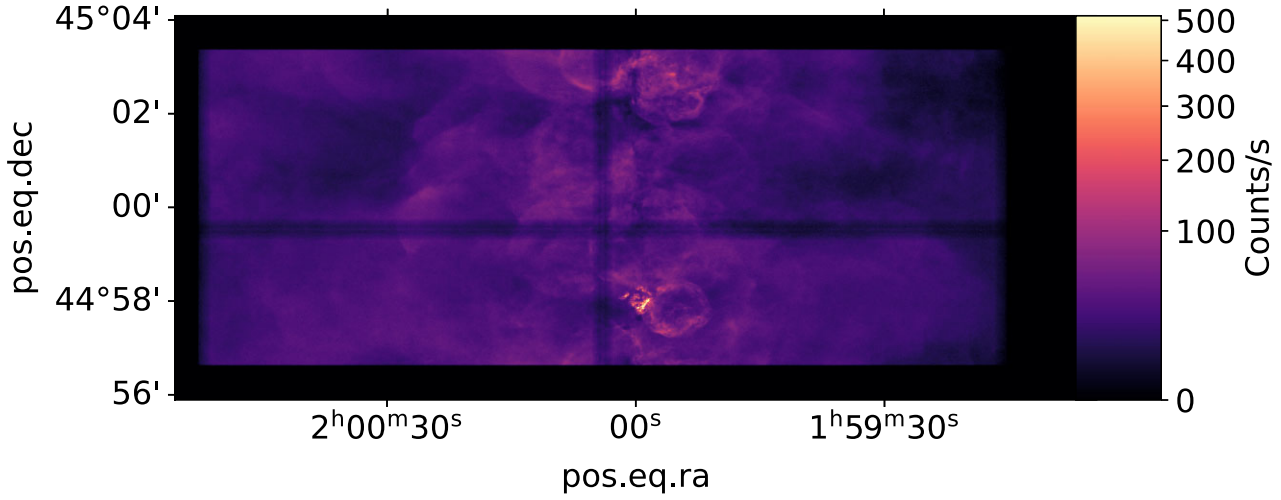

Figure 3. A simulated *Chandra*/ACIS-I image in the 0.4–2 keV band of the snapshot shown in Fig. 1; the colourbar shows photon count rate per pixel. We take $Z_{\rm bg} = 0.5 Z_\odot$ and use a 10 Ms exposure time.

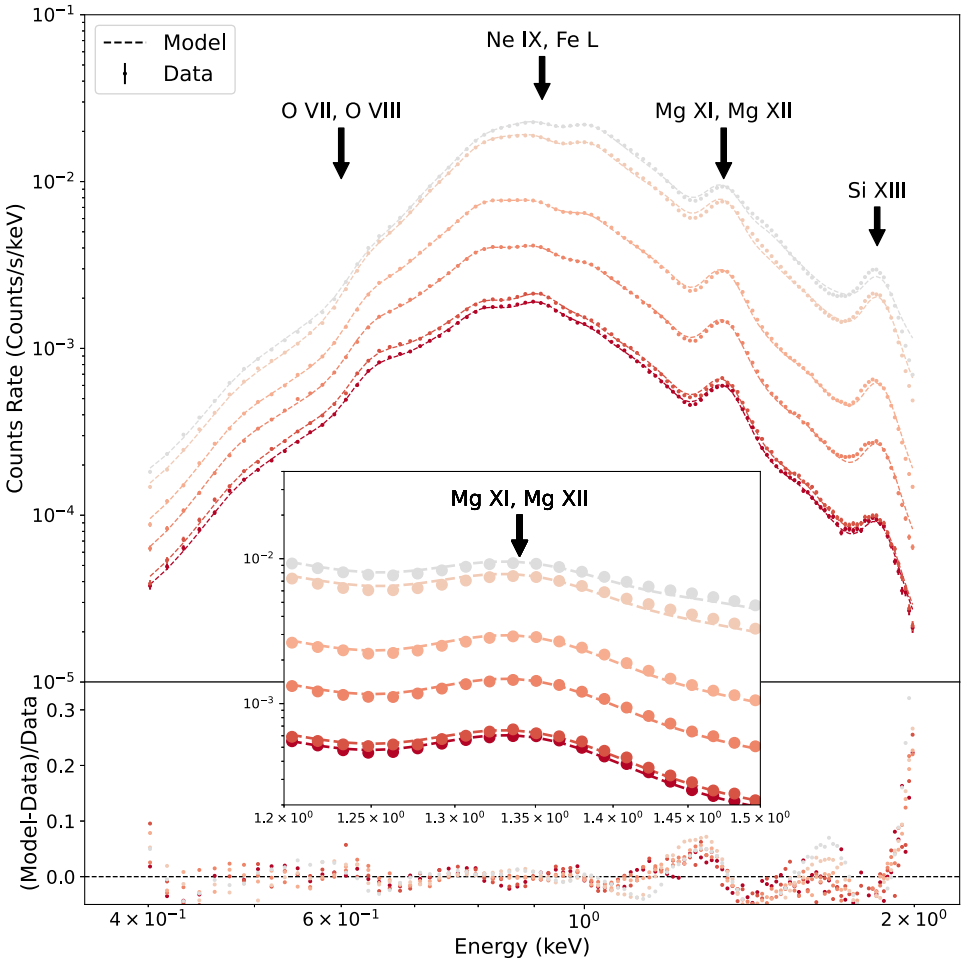

Figure 4. Mock spectra generated from PYXSIM and the fit from SHERPA. The top panel shows synthetic spectra (points with error bars) and best-fitting models (dashed lines) for the regions shown in Fig. 1. Colours indicate the region over which the spectrum was computed, starting with 'D' (grey, top) and continuing from 'A1' to 'A5' (top to bottom, lighter to darker red, matching the colours of the rectangles in the bottom panel of Fig. 1). We omit the results for regions 'B1' to 'B5' for aesthetics, but these are qualitatively similar to the corresponding 'A' regions. Labels indicate the locations of particular spectral features and the inset panel zooms in on the region from 1.2 to 1.5 keV. The bottom panel shows residuals between the data and the best-fitting model.

temperature – Vijayan & Li 2022). For each of these possibilities, we use SHERPA to find the model that best fits the spectrum in each region, as characterized by the minimum χ^2 value, using SHERPA's built-in Levenberg–Marquardt fitting algorithm. When carrying out these fits we treat the abundances of the elements O, Ne, Mg, Si, S, and Fe as free parameters; although we know that our model galaxies have Solar-scaled abundances, we do not add this as a constraint to the fits in order to make the fitting procedure as realistic as possible. This means that for an N temperature model, we have a total of $6 + 2N$ free parameters to be fit: the abundances of six elements plus the temperature and normalization of each component.

Due to the high-dimensional nature of the parameter space, we find that for certain values of the initial guess the fitter sometimes fails to find that true χ^2 minimum; this problem is particularly likely to occur for regions far from the disc where the metallicity and SNR are lower. To mitigate this we carry out fits using a range of guesses for the initial elemental abundances from $Z/Z_{\rm X,\odot} = \max(Z_{\rm bg} - 0.5, 0)$ to $Z/Z_{\rm X,\odot} = Z_{\rm bg} + 0.5$ in steps of 0.1, where $Z_{\rm X,\odot}$ is the abundance of each species X at Solar metallicity. We then accept whichever fit returns the lowest χ^2 value. Once we have these best-fitting values for each of the one-, two-, and three-plasma models, we accept as our final fit whichever of these produces a reduced χ^2 closest to unity for each region. For all the snapshots and values of $Z_{\rm bg}$ we consider in this paper, we find that for the central region 'D' the best fit is a three-plasma model, while all other regions are best fit by two-plasma models.

We present the best-fitting spectra for our example case as dotted lines in Fig. 4. The bottom panel shows the residuals (i.e. (Model − Data)/Data). The residuals are smaller for the lower end of the energy range (~ 0.5 keV) and for regions further away from the mid-plane. Overall, however, the fits reproduce most aspects of the mock spectra well; the inset in Fig. 4 demonstrates this by zooming in on the spectral region 1.2–1.5 keV.

3 RESULTS AND ANALYSIS

We are now in a position to discuss our main results. We use the procedures outlined in Section 2 to derive the metallicity and temperature values from the mock observations. As noted above, we carry out this analysis for $Z_{\rm bg} = 0, 0.2, 0.5, 1$, and 2; we also analyse a control run in which we leave the density and temperature structure unchanged, but set the metallicity to a *uniform* value $Z = Z_\odot$ everywhere in the simulation domain. Except where otherwise noted we use the simulated observations for a 10 Ms exposure so that we have essentially noise-free spectra.

3.1 Time variation

To ensure that our findings are not dependent on the specific simulation time we choose to analyse, we use the pipeline described in Section 2 to process four snapshots, corresponding to the state of the simulations at 100, 105, 110, and 115 Myr. Our goal here is simply to assess the level of fluctuations in the spectra once the wind has reached a statistically steady state. We choose these specific times because the outflows have reached statistical quasi-steady state after 100 Myr, and a 5 Myr interval between snapshots ensures that the results are not causally linked: the typical hot gas speed in the simulation is $\sim$1000 km s^{-1}, so 5 Myr interval is long enough hot gas generated at the mid-plane to exit the simulation domain at ± 4 kpc.

In Fig. 5, we present the spectra in regions 'D' and 'A5' from all four snapshots. The 'D' spectra are nearly identical across snapshots,

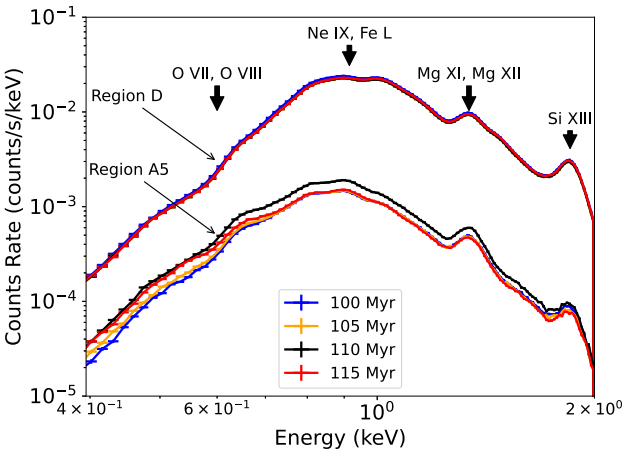

Figure 5. Synthetic spectra at region 'D' and 'A5' from the snapshots at 100, 105, 110, and 115 Myr, as indicated by different colours (see legend). The group of curves at the higher count rate corresponds to region 'D', while the group at the lower count rate corresponds to region 'A5'. Because all snapshots produce similar spectra, we focus on the at 110 Myr and generalize our results.

while the 'A5' spectra vary slightly in normalization but show a nearly constant shape. This indicates that the outflow properties and the resultant spectra are stable over time. Quantitatively, fitting spectra at the four times shown yields best-fitting estimates for elemental abundances that vary by an amount comparable to the formal uncertainties on the fits. Given this consistency, we focus our detailed analysis on the snapshot at 110 Myr for the remainder of this study, since the results are qualitatively – and mostly quantitatively – identical to other snapshots.

3.2 Outflow metallicity patterns

We first focus on the case $Z_{\rm bg} = 0.5$, as a baseline and to guide the discussion that follows. Fig. 6 shows the best-fitting temperature and abundances as a function of distance from the mid-plane for this case. We see that this fit includes one low-temperature component with $kT \approx 0.2$ keV almost independent of height, and one higher temperature component that goes from $kT \approx 0.5$ keV at $z = 0$ to ≈ 0.35 keV at $z \approx 3.5$ kpc; in the disc region only there is also evidence for a third even hotter component at $kT \approx 1.2$ keV. The general trend that the plasma temperature falls with distance off the mid-plane is consistent with the phase distribution of hot gas seen in QED I and shown in Fig. 1.

The abundances of O, Ne, Mg, Si, and Fe plotted in the bottom panel of Fig. 6 show similar patterns, which is expected given that we have assumed Solar-scaled abundances and thus the true abundance ratios are constant. This pattern is that the outflowing gas possesses significant metallicity gradients, with a peak in the central region D and a subsequent decline for A/B-1-5. Such a trend in metallicity is strikingly similar to that observed by Lopez et al. (2020) for S and Si in the wind of M82. We discuss the physical origin of this pattern and its implications in Section 4.

For comparison, we also show the 'true' hot gas metallicity, reminding readers that, because our simulation assumes Solar-scaled abundances, the true (Solar-normalized) metallicity is the same for all elements. There is some ambiguity about how to define this, since different choices of how to average over the metallicity distribution present in the real simulation data yield different outcomes. In order

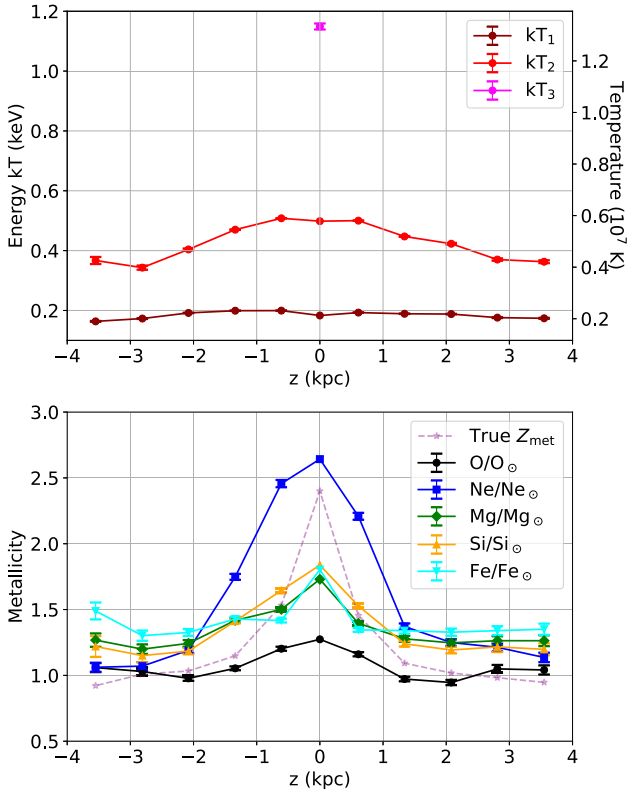

Figure 6. Top: best-fitting temperature of each plasma component as a function of distance from the mid-plane in the 110 Myr snapshot with $Z_{bg} = 0.5Z_\odot$. Mid-plane distances are measured to the centre of each slice, and note that a third temperature component (T_3) is present only in the 'D' region centred on $z = 0$. Bottom: best-fitting abundances of the elements O, Ne, Mg, Si, and Fe normalized to Solar (solid lines, as indicated in the legend) and the true metallicity (dashed line, see equation 1) as a function of height.

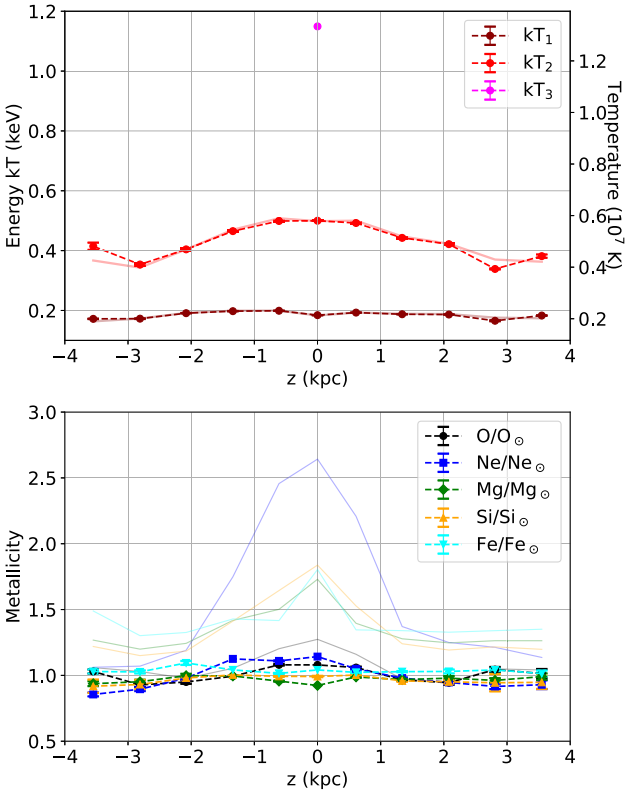

Figure 7. Same as Fig. 6, but now comparing the standard case $Z_{bg} = 0.5$ to a control run where we set $Z = Z_\odot$ in all cells. In the plots above, dash lines with points and error bars are the results for the control run, while faded lines are the $Z_{bg} = 0.5$ case for comparison; these lines are identical to those shown in Fig. 6.

to produce something as close as possible to the quantity returned by the observations, we define our true metallicity gradient to be the X-ray luminosity weighted mean metallicity. That is, we define the true metallicity as

$$Z_{met} = \frac{\int \epsilon_X(Z,T) Z \, dV}{\int \epsilon_X(Z,T) \, dV}, \quad (1)$$

where T and $Z = Z_{bg} + Z_{SN}$ are the temperature and total metallicity in each cell, Z_{SN} is the metallicity contributed by SNe as measured from the concentration of passive scalar in the simulation, $\epsilon_X(Z,T)$ is the luminosity per unit volume from 0.4 to 2 keV (as shown in the fourth row of Fig. 1), and the integral goes over all the cells that belong to a given slice. We see that none of the observationally inferred gradients quite reproduce the true metallicity structure, shown in the dashed line in Fig. 6, and that the fits show an asymmetry between the regions above and below the mid-plane that is not present in the underlying data. This is perhaps not surprising: the fits rely on the hugely simplifying assumption that there are only a small number of single-temperature plasma components, each with the same abundances. Neither of these assumptions are true, and given the very high SNR synthetic spectra we have used to carry out these fits, we expect these systematic errors to dominate the error budget. None the less, the qualitative trend we find in the simulation data – a maximum in the metallicity in the disc region and a decrease away from it – is reproduced for each fitted element.

As a further confirmation that the fits are responding to the real metallicity gradient present in the data, and are not simply an artefact introduced by the temperature gradient, we can compare to the control run which has $Z = Z_\odot$ everywhere. We show the results for this case superimposed on the results for $Z_{bg} = 0.5$ in Fig. 7. We see that in the control run the best-fitting temperatures are nearly identical, as expected, but we now recover no gradient in the abundances. Thus our fits correctly diagnose this case as having a wind of uniform metallicity. This gives us confidence that, while the inferred metal abundances do not precisely recover the true distribution, they do capture the major qualitative feature of the real distribution and are unlikely to be false positives.

3.3 Results for varying background metallicity

We now expand our results to other Z_{bg} values to understand how the observationally inferred metallicity gradient changes as we vary the strength of the background metallicity upon which the SN-injected metals are imposed. We show the result of this experiment for Ne in the top panel of Fig. 8; we choose Ne because it is the element that shows the steepest gradient in our fits with $Z_{bg} = 0.5$, but the qualitative result is the same for all elements. By contrast, we show the true X-ray luminosity weighted mean metallicities for these runs in the bottom panel.

We retrieve vertical gradients in metallicity for all values of Z_{bg}. The pristine case of $Z_{bg} = 0$ shows the strongest gradient in inferred Ne abundance, with a factor of ~ 8 difference between the

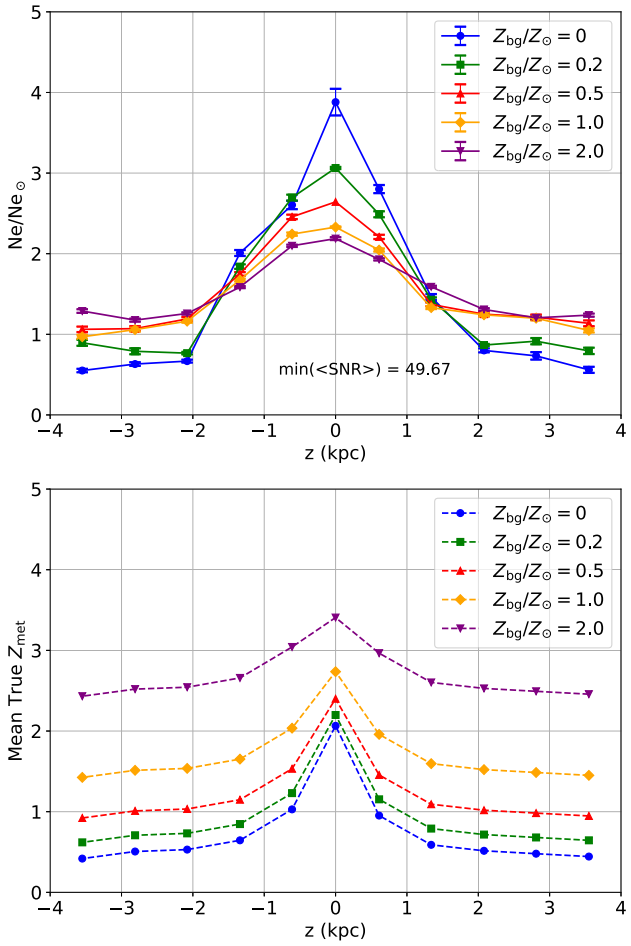

Figure 8. Top: best-fitting Ne abundance as a function of height off the mid-plane for varying levels of background metallicity Z_{bg}, as indicated in the legend. Bottom: true metallicity (equation 1) as a function of height estimated using different Z_{bg}.

abundances recovered at $z = 0$ and $z \approx |4|$ kpc. Gradients become successively flatter with increasing Z_{bg}, with the case $Z_{bg} = 2$ showing less than a factor of 2 difference between regions D and A5 or B5. Compared to the true gradients, we see that the numerical values do not match precisely, but the sense of variation does, i.e. as we reduce Z_{bg} both the true gradient and the gradient in inferred Ne abundance become steeper. However, we also note some surprising behaviour: the retrieved central abundance *decreases* with increasing Z_{bg}, opposite to the trend in the actual metallicity (cf. the top and bottom panels of Fig. 8).

We can understand this behaviour as arising from a fundamental limitation of the commonly used observational fitting procedure, which is that it assumes that the different temperature components share a common set of abundances. This assumption, while it is likely a practical necessity to avoid having the number of free parameters in the fit balloon out of control, is fundamentally incorrect, particularly near the disc, because the emitting medium here consists largely of bubbles of near-pure SN ejecta that are both very hot ($\sim 10^7$ K) and very metal-rich, and a slightly cooler (but still $\sim 10^6$ K) component of shocked ISM gas that is more metal-poor. Now let us consider the implications of this configuration for how the X-ray spectrum changes as we vary Z_{bg}. The bubbles of SN-ejecta produce emission that is not tremendously sensitive to Z_{bg}, because their metallicities are dominated by direct SN ejecta; they will be metal-rich even for $Z_{bg} = 0$. The shocked ISM component, on the other hand, has a luminosity that is highly sensitive to Z_{bg}; low Z_{bg} makes this component dimmer compared to the emission from the fresh SN ejecta, while higher Z_{bg} makes it brighter. We can therefore understand the effect of Z_{bg} as changing the weight of the two components, each of very different intrinsic metallicity, that contributes to the overall spectrum.

When we fit this composite spectrum with a model that assumes that there is a single metallicity, the fit naturally tries to find a compromise between the two components. However, it winds up favouring whichever one is brighter. Thus when Z_{bg} is small and emission is dominated by the fresh SN ejecta, the fit returns a very high metallicity characteristic of those ejecta. As we increase Z_{bg} and thus the relative contribution from the shocked ISM component, the fit moves toward the lower metallicity of this component, explaining why the inferred abundances are lower at higher Z_{bg}.

To confirm this hypothesis, we have experimented with generating spectra with emission from the most metal-rich gas, $Z > 4Z_\odot$, artificially set to zero. We find that, when we do this, the anomalous behaviour that inferred abundances decrease with increasing Z_{bg} vanishes. Instead, consistent with our proposed hypothesis, we recover abundances that increase with Z_{bg} as expected.

In regions further away from the mid-plane ($|z| > 2$ kpc) this effect vanishes and increasing background metallicity increases spectrum-predicted metallicity. This is because at these greater distances from the injection point, the superhot phase has now mixed and cooled down and the dominant X-ray emitting gas is at $\sim 10^6$ K. We no longer have emission coming from two phases at so starkly different metallicities and temperatures, but are closer to the state envisioned by the fit where there is a range of temperatures but only a single metallicity. The better match between the true and assumed gas metallicity and temperature distributions in turn means that the fit works better, and we no longer have the erroneous result that higher Z_{bg} pushes the fit in the wrong direction.

Finally, we note that neither the 'true' X-ray-weighted metallicity nor the elemental abundances we retrieve from the X-ray spectrum fully agree with the mass-weighted metallicities computed directly from the simulation data, even if we limit ourselves to the hot phase. In particular, fig. 9 of QED I shows that for the hot phase, the mass-weighted metallicity is near-identical to the background metallicity for $Z_{bg} \gtrsim 0.2$. Thus weighting by X-ray luminosity or the emission spectrum tends to accentuate the gradients compared to the mass-weighted average. For the true metallicity values, the averaging is biased towards regions of higher density, particularly because the luminosity is proportional to the square of the density. Therefore, even a small increase in emissivity due to increased background metallicity disproportionately biases the true metallicity average towards gas with a higher background, making the inferred metallicity appear higher than a mass-weighted value would suggest. We explore this issue further in Section 4.3.

3.4 Impact of finite signal-to-noise ratio

Thus far we have restricted our analysis to spectra generated for an exposure time of 10 Ms, so we are examining essentially noise-free data. It is therefore important to consider the more realistic case of observations with a finite SNR, in order to determine what quality data are required to retrieve the essential features of the outflow metallicity pattern.

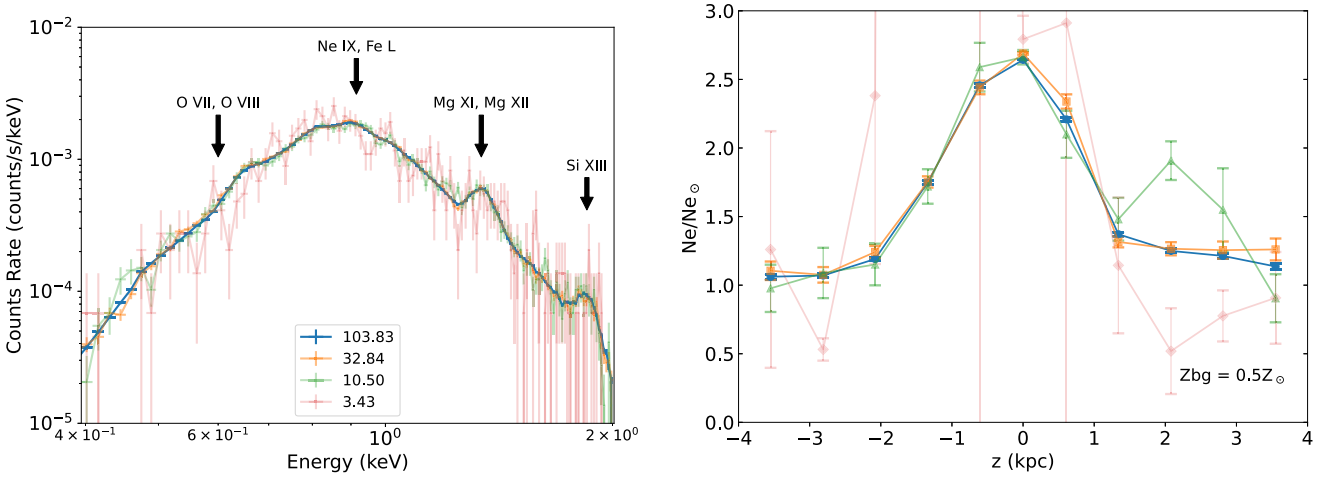

Figure 9. Left: spectra for region 'A5' for the case of $Z_{bg} = 0.5Z_\odot$ at 110 Myr measured at four different exposure times – 10, 1, 0.1, and 0.01 Ms. The legend indicates the value of SNR_{obs}, the overall SNR we estimate from the exposure time (equation 4). Right: Ne abundances as a function of height recovered from spectral fitting, for the same four SNRs shown in the left panel.

To quantify spectral quality, we define the mean photon count-weighted $\langle SNR \rangle$ to average over the SNRs in each energy bin:

$$\langle SNR \rangle = \frac{\sum_i c_i \cdot SNR_i}{\sum_i c_i}, \tag{2}$$

where the sum runs over all energy bins from 0.4 to 2 keV, c_i is the photon count in bin i (which is proportional to the exposure time), and SNR_i is the SNR of bin i. Assuming that we are dominated by Poisson noise, $SNR_i = c_i^{1/2}$, and this gives

$$\langle SNR \rangle = \frac{\sum_i c_i^{\frac{3}{2}}}{\sum_i c_i}. \tag{3}$$

Equation (3) provides us with a means of calculating the mean SNR of the spectrum in each region. To characterize the overall SNR of an observation, we then take the minimum over all regions, i.e. we define

$$SNR_{obs} = \min_j \left(\langle SNR \rangle_j \right), \tag{4}$$

where j runs over our observing regions, i.e. we let $j =$ A5, A4, . . . , D, B1, . . . B5.

To generate results for a range of values of SNR_{obs}, we vary the exposure time from the initial 10 Ms, decreasing it to 1, 0.1, and 0.01 Ms. For each of these cases we produce mock spectra with soxs and fit using SHERPA, and we plot the resulting spectra for the A5 region and the spectrum-derived Ne abundance for all regions in the right and left panels of Fig. 9, respectively; we again focus on Ne as an example, but the qualitative results are the same for other elements. In both the panels, the curves are labelled by SNR_{obs}. The blue curve ($SNR_{obs} = 103.83$) is the case for a 10 Ms exposure, and is identical to the red curve in Fig. 8; exposure times of 1, 0.1, and 0.01 Ms yield $SNR_{obs} = 32.84$, 10.50, and 3.43, respectively, as indicated in the legend. As one might expect, for poor SNRs, the deviation from the 'perfect' spectrum is smallest for mid-energy bins ($E \sim 1$ keV) where *Chandra* is most sensitive, and increases rapidly for lower and higher energy bins. This may be the result of the relatively poor sensitivity of *Chandra* at these energies.

From the right panel, we see that the retrieved abundances begin to deviate from the 'perfect data' case as the SNR degrades. At a SNR of 33, the results match the perfect case to within fit uncertainties, and even for a SNR of 10.5 the main feature of the metal distribution – in particular that there is a maximum near $z = 0$ and a decrease away from it – remains visible, albeit with an additional spurious peak appearing on the $+z$ side. Interestingly, this spurious peak looks quite similar to a feature that Lopez et al. (2020) detected for Ne in their analysis of the M82 wind – cf. their fig. 5. For a SNR of 3.4, even broad features of the abundance distribution are lost, and in fact the fit fails to converge at all for the bin at $z \approx -1$ kpc. This suggests that X-ray observations seeking to retrieve metal abundances in galactic outflows should aim for a minimum mean SNR of at least 10, and preferably closer to 30. With this level of SNR, the results will be dominated by the systematic errors coming from the simplified model of the emitting plasma rather than the random errors arising from photon counting statistics.

4 DISCUSSION

Here, we discuss two implications of our work: first that negative metallicity gradients in galactic outflows provide unambiguous observational evidence for metal loading (Section 4.1), and second that measurements of such gradients can be, at least approximately, turned into quantitative measurements of phase mixing in galactic outflows (Section 4.2).

4.1 Negative metallicity gradients as evidence for metal loading

Our findings offer a clear explanation for the observation that galactic winds show α abundances that are enhanced relative to those of the ISM, but that this enhancement declines away from the galactic plane. This pattern occurs due to the distribution of metals among different gas phases and their respective contributions to X-ray emission. In the disc of the galaxy, recent SN activity produces very hot gas that is highly enriched with α elements, leading to X-ray emission characteristic of elevated metallicities. At the same location there is a cold phase ($T < 10^4$ K) of gas with much lower metallicity (but which none the less carries most of the total metal mass), this gas does not contribute to X-ray emission because of its low temperature,

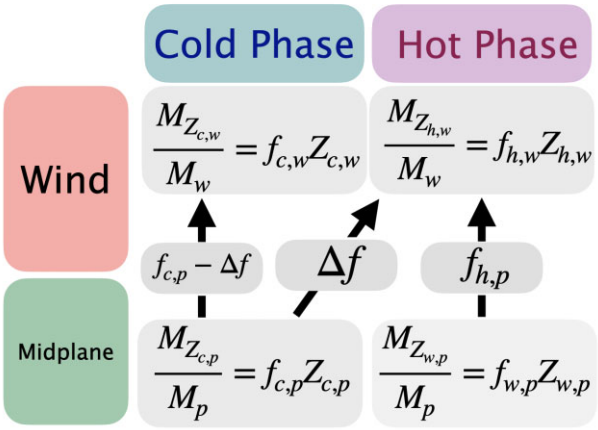

Figure 10. Visualization of the '4-grid model'. This schematic illustrates the division of gas into mid-plane and wind zones, each containing cold and hot phases. The model represents the transfer of mass fraction between these components, with Δf indicating the net increase in hot gas fraction from the mid-plane to the wind. Arrows represent potential gas flows and mixing processes between the components.

so the X-ray spectrum is dominated by the hot, very metal-rich gas from supernova ejecta. As the hot gas flows outward, however, it mixes with clouds of colder, metal-poorer gas from the surrounding ISM that have been entrained into the outflow. Over time and as the outflow moves away from the galactic plane, some of this cooler gas evaporates into the hot phase, diluting it and decreasing the mean metallicity of the X-ray emitting gas by an amount that increases with distance from the plane. This leads to a metallicity gradient where the X-ray emitting gas near the disc has higher metallicity than that further out in the wind. Although in our simulations we have assumed Solar-scaled abundances for simplicity, in reality, this trend should occur primarily for the α elements that are preferentially produced by core collapse SNe. This phenomenon explains the trends observed by Lopez et al. (2020).

The converse point is perhaps more interesting: as shown in Fig. 7, when we consider the control case of a wind that is *not* heavily metal loaded and has roughly the same metallicity as the ISM, the observed gradient vanishes. This means that the abundance pattern we observe in the wind of M82 requires that unmixed SN ejecta make a significant contribution to the total wind budget. In the language commonly used in galactic chemical evolution modelling, the observations of M82 require that the metal loading factor ζ, which measures the metallicity of the wind relative to that of the ISM (Peeples & Shankar 2011), be larger than unity, while the yield reduction factor ϕ, which parametrizes the fraction of newly produced metals from SNe that are retained in the disc rather than lost promptly to an outflow (Sharda et al. 2021), must be significantly below unity.

4.2 Analytical model of phase mixing

The explanation of abundance gradients that we have just presented raises an obvious question: can we use measured metallicity gradients to deduce the nature of phase mixing in galactic winds? To see that this is possible, it is helpful to introduce a toy '4-grid model' that quantifies the extent of mixing between the hot ($T > 10^6$ K) and cold ($T < 10^4$ K) phases of the outflows from observations of gradients; we illustrate this model schematically in Fig. 10.

We divide the gas into two zones – the mid-plane and the wind – with a boundary at $\sim \pm 1$ kpc, roughly the height at which the outflows are launched. Each zone contains both hot and cold components, giving us four 'boxes' of gas: cold gas in the plane (c,p), hot gas in the plane (h,p), cold gas in the wind (c,w), and hot gas in the wind (h,w). The respective metal masses in these boxes are denoted by $M_{Z_{\rm c/h,p/w}}$, and we can express these metal masses as the product of the total mass in each zone $M_{\rm p}$ and $M_{\rm w}$ times the fraction f of that mass in the hot or cold phases times the mean metallicity Z of each phase, i.e. $M_{Z,{\rm c,p}} = Z_{\rm c,p} f_{c,p} M_{\rm p}$ and similarly for the other three gas types shown in Fig. 10. Here, $f_{\rm c,p}$ for example is the fraction of in-plane gas that is in the cold phase, and clearly, we have $f_{\rm c,p} + f_{\rm h,p} = 1$ and $f_{\rm c,w} + f_{\rm h,w} = 1$.

Since we have seen that evaporation of cold gas into the hot phase is the key physical process driving metallicity gradients, we let Δf represent the net transfer of gas from the cold mid-plane to the hot wind, driven by the heating and evaporation of cold mid-plane gas into the hot wind phase. Hence, the value of Δf is

$$\Delta f = f_{\rm c,p} - f_{\rm c,w} = f_{\rm h,w} - f_{\rm h,p}. \tag{5}$$

The metal mass in the hot wind can therefore be written as a sum of two components: the metal in hot gas outflowing from the mid-plane and the metal added by evaporation of cold mid-plane gas into the hot wind phase,

$$\frac{M_{Z_{\rm h,w}}}{M_{\rm w}} = f_{\rm h,p} Z_{\rm h,p} + \Delta f Z_{\rm c,p}\,. \tag{6}$$

But since $M_{Z_{\rm h,w}} = f_{\rm h,w} Z_{\rm h,w} M_w$, this means that

$$f_{\rm h,w} Z_{\rm h,w} = f_{\rm h,p} Z_{\rm h,p} + \Delta f Z_{\rm c,p}\,. \tag{7}$$

If we now substitute for $f_{\rm h,w}$ in terms of Δf using equation (5), we can rewrite this relation as

$$\frac{\Delta f}{f_{\rm h,w}} = \frac{1 - Z_{\rm h,w}/Z_{\rm h,p}}{1 - Z_{\rm c,p}/Z_{\rm h,p}}. \tag{8}$$

The relation above is useful because it connects directly observable quantities, the metallicity ratio between the hot and cold gas in the central disc ($Z_{\rm h,p}$ and $Z_{\rm c,p}$), and that in the outflowing wind ($Z_{\rm h,w}$ and $Z_{\rm c,w}$), to something that is not directly observable but is of significant interest: the fractional increase in the fraction of outlowing material that is hot due to the evaporation of cool clouds into the wind ($\Delta f / f_{\rm h,p}$). Thus we can use the evolution of the metallicities with height as a direct probe of phase mixing.

In the limiting case where the background ISM has low metallicity, we can simplify this expression even further by assuming that $Z_{\rm c,p} \ll Z_{\rm h,p}$, i.e. that the hot gas freshly injected by SNe is much, much more metal-rich than the surrounding cool ISM. In that case, our expression simplifies even further to

$$\frac{\Delta f}{f_{\rm h,w}} \approx 1 - \frac{Z_{\rm h,w}}{Z_{\rm h,p}}, \tag{9}$$

and recalling that $\Delta f = f_{\rm c,p} - f_{\rm c,w} = f_{\rm h,w} - f_{\rm h,p}$ we then have

$$\frac{f_{\rm h,p}}{f_{\rm h,w}} \approx \frac{Z_{\rm h,w}}{Z_{\rm h,p}}. \tag{10}$$

In this case, the metallicities of the cold gas vanish entirely, and we can directly read off the increase in the hot gas mass fraction due to evaporation from the ratio of the hot phase metallicities in the plane versus the wind. Thus, for example, if observations of a galaxy reveal that the metallicity of the hot wind is half that of the hot mid-plane hot gas (i.e. $Z_{\rm h,w}/Z_{\rm h,p} = 0.5$, roughly what Lopez et al. (2020) find for Si in the wind of M82), the simplified relation implies that the

hot gas fraction in the mid-plane is approximately half that in the wind (i.e. $f_{\rm h,p}/f_{\rm h,w} \approx 0.5$), so the wind a few kpc from the galaxy consists roughly equally of gas that was heated at the mid-plane and escaped while hot, and cool gas that was entrained at the mid-plane but subsequently evaporated into the hot phase. However, we caution that this result will always be at least somewhat approximate, because we have seen that the elemental abundances inferred from X-ray measurements do not perfectly reflect the underlying abundances. For example, if we examine Fig. 6, we can see that carrying out this exercise for O versus Ne would lead to quite different numerical values for $f_{\rm h,p}/f_{\rm h,w}$. We explore the reasons for these errors further in Section 4.3.

4.3 Why do biases in derived abundances vary with element?

Following our toy model of phase mixing, we now examine why X-ray derived abundances can differ systematically from the underlying metallicities, and do so differently for different elements. The key drivers of these differences are that (1) there is no single characteristic temperature for the 'hot' ($T > 10^6$ K) phase, but rather a broad distribution of temperature, (2) that temperature and metallicity are highly correlated, and (3) due to the different lines that contribute to the X-ray spectrum from 0.4 and 2 keV, different elements are sensitive to gas at different temperatures, rather than simply reflecting some sort of mean X-ray weighted metallicity such as $Z_{\rm met}$ (equation 1).

We illustrate the first two points using the case $Z_{\rm bg} = 0.5 Z_\odot$ as an example; for each region of this case we divide the gas hotter than 10^6 K (the temperature range that drives the bulk of the X-ray emission; cf. Fig. 2) into temperature bins 0.25 dex wide, and within each bin we compute the mass-weighted mean metallicity. We show the result for regions A1, A3, and A5 in Fig. 11; other regions are qualitatively similar. A key point to take from these plots is that, except in region A5, 'hot' gas spans more than an order of magnitude in temperature, and metallicity is highly correlated with temperature, particularly close to the galactic plane. The origin of this correlation is simply that one of the most important cooling channels for the very hot gas produced by SNe, particularly at small z when there has been little time for adiabatic expansion or radiative cooling, is dilution by mixing with cooler gas, and dilution simultaneously cools and lowers metallicity.

To illustrate how this interacts with the third point, regarding the different temperature sensitivity of different elements, we also in Fig. 11 show the metallicities in each region inferred from the X-ray spectral fits for three sample elements, O, Ne, and Fe. We choose these three because, as shown in Fig. 6, O and Ne show systematically the lowest and highest inferred abundances near the mid-plane, while Fe is intermediate between the two. We can understand the differences between these elements in light of Fig. 11 as follows. First, note that the sensitivity of the spectrum to oxygen is largely through the prominent O VIII feature at ≈ 0.65 keV, and the strength of this feature is strongly influenced by the oxygen ionization balance, with the fraction of oxygen in the O VIII reaching a maximum of ≈ 50 per cent at $T \approx 2.5 \times 10^6$ K, and falling below 10 per cent for $T \lesssim 1.2 \times 10^6$ K or $T \gtrsim 6 \times 10^6$ K (Sutherland & Dopita 1993). Consequently, oxygen strongly reflects the abundances prevalent at the lower temperature end of 'hot gas', $T \approx (2-3) \times 10^6$ K. By contrast, the greatest sensitivity to Ne is to Ne IX and Ne X features at ≈ 1 keV, and the latter ionization state accounts for more than 10 per cent of the available neon only at temperatures $T \approx (2.5-10) \times 10^6$ K, with a maximum of ≈ 50 per cent abundance at $T \approx 5 \times 10^6$ K (Sutherland & Dopita 1993). This makes Ne sensitive to significantly hotter gas than O, and so when abundant very hot gas is present – as Fig. 11 shows it is in regions A1 and A3 – the metallicity inferred from Ne winds up reflecting the higher metallicity of this gas. By contrast iron is somewhat more complex, since the X-ray spectral feature most sensitive to iron is the complex of Fe L lines at ≈ 1 keV, which can be produced by Fe in ionization states from Fe XVIII to Fe XXIV. The lowest of these ionization states reaches $\gtrsim 10$ per cent abundance at $T \approx 3 \times 10^6$ K, leading to a behaviour that is somewhat intermediate between that of O and Ne, although the temperature dependence is more complex for Fe because the highest ionization state that produces Fe L emission remains abundant up to much higher temperatures, $T \approx 3 \times 10^7$ K. The net effect of this complex dependence is that in the outer regions (e.g. A3 and A5) the oxygen abundance is closer to the X-ray weighted metallicity $Z_{\rm met}$, while near the disc (e.g. A1) the iron abundance is closer.

In summary, we can understand qualitatively why different elements are biased by different amounts based on the fact that they are sensitive to different ranges of gas temperature, and that gas at different temperatures really does have systematically different metallicities due to the effects of cooling by dilution. Thus, while metallicity declines with distance from the mid-plane, and all elements properly capture this trend, the discrepancy between the abundances derived for any particular element and the 'true'

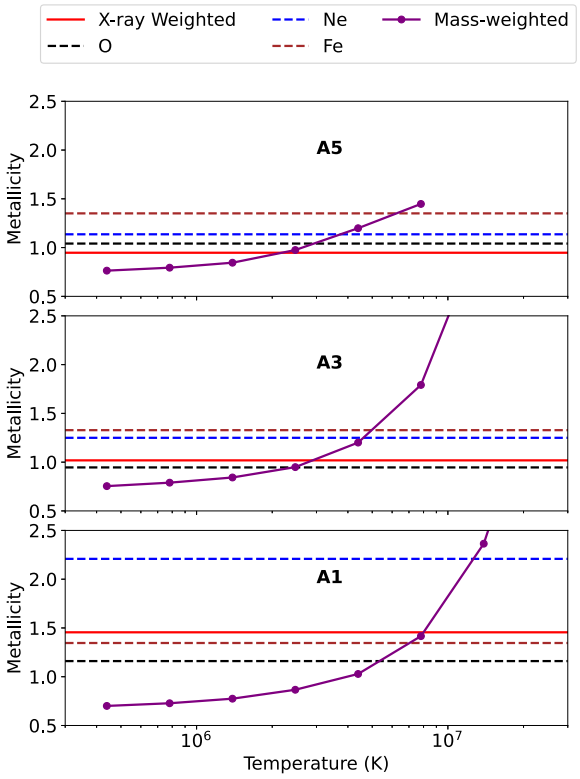

Figure 11. Mass-weighted metallicity in discrete temperature bins (purple curves), plotted together with horizontal lines showing the derived metallicities for oxygen (black dashed), neon (blue dashed), and iron (brown dashed). The red solid line shows the 'true' X-ray luminosity-weighted metallicity $Z_{\rm met}$. We display three representative wind regions ('A1', 'A3', 'A5') in stacked panels. O VII and O VIII lines are most sensitive to lower temperatures ($T \approx (2-3) \times 10^6$ K), while Ne and Fe trace hotter gas, which combined with the metallicity–temperature correlation lead metallicities derived based on the O abundance to be systematically smaller than those derived from Ne or Fe.

metallicity value arises naturally from these complex combinations of atomic physics and the presence of a metallicity temperature correlation in the gas. Given the complexity of this dependence, it seems unlikely that we will be able to acquire significantly more accurate results from spatially unresolved spectra.

5 CONCLUSION

We study mock X-ray emission from galactic winds generated in a high-resolution hydrodynamic simulation of the Solar neighbourhood taken from the QED simulation suite, with particular attention to spatial variations in the elemental abundances across the wind derived from analysis of the X-ray spectra. The simulations possess a significant gradient in abundances whereby the hot gas closer to the galactic plane, which is dominated by fresh SN ejecta, is more metal-rich than the hot gas further from the plane, which has been diluted down by mixing with more metal-poor cool gas entrained into the wind. Our analysis reveals that this effect should lead to a metallicity gradient that is detectable from X-ray spectra, with higher abundances of α elements like oxygen and silicon observed closer to the disc compared to the outer regions of the outflows. Such a pattern has already been observed in the wind of M82, and our simulations provide a natural explanation for it.

Our main conclusions are:

(i) Gradients in elemental abundance inferred from X-ray spectra of galactic winds, with a peak at the centre and decline at high altitudes, are a result of incomplete mixing between the hot and cold phases.

(ii) Conversely, the presence of such gradients in observations is strong evidence for preferential metal loading of winds – in the language of galactic chemical evolution models, they are evidence that the yield reduction factor ϕ, which measures the fraction of SN ejecta that are retained in the disc and available to enrich the next generation of stars, is < 1. Models that are not metal-loaded do not reproduce observations.

(iii) The strength of the gradient can be used as a diagnostic for the strength of evaporation of cold gas entrained by winds into the hot phase, with a steeper gradient implying an increasing contribution to the mass of the hot gas phase by evaporation of cool clouds.

(iv) Making these measurements in practice is quite sensitive to the SNR of the X-ray data; low SNR data can create artificial gradients or mask real ones. We find an SNR $\gtrsim 30$ is necessary to faithfully extract metallicity values from the spectrum, and that SNR $\gtrsim 10$ is sufficient to detect the presence of a gradient, but probably not to recover its shape faithfully.

(v) Even with high SNR data, however, there are likely to be non-negligible systematic errors in inferred gradients arising from the need to assume a single set of abundances for all of the gas in order to avoid the number of free parameters in the fit growing too large. In the most poorly mixed regions close to the disc, which contain bubbles of fresh SN ejecta with very high metallicity and temperature, this assumption breaks down.

SOFTWARE

This research made use of NUMPY (Harris et al. 2020, https://numpy.org), MATPLOTLIB (Hunter 2007, https://matplotlib.org/), YT (Turk et al. 2011, https://yt-project.org/), PYXSIM (ZuHone & Hallman 2016, http://www.ascl.net/1608.002), SOXS (ZuHone et al. 2023, http://ascl.net/2301.024), and QUOKKA (Wibking & Krumholz 2022; He et al. 2024, https://github.com/quokka-astro/quokka).

ACKNOWLEDGEMENTS

AV and MRK acknowledge support from the Australian Research Council through awards FL220100020 and DP230101055. The simulations suite QED and the mock data discussed in this paper were generated with the assistance of resources from the National Computational Infrastructure (NCI Australia), an NCRIS enabled capability supported by the Australian Government, and from the Pawsey Supercomputing Research Centre's Setonix Supercomputer (https://doi.org/10.48569/18sb-8s43), with funding from the Australian Government and the Government of Western Australia. RH thanks Andrew J. Battisti, Tianmu (Tim) Gao, Chong-Chong He, Yunhe Li, and Yusen Li for commenting on this work.

DATA AVAILABILITY

The software pipeline used in this analysis is available from https://github.com/Rongjun-ANU/Xpipeline. Due to their large size, the raw QED simulation outputs on which the analysis is performed are not available in this repository, but are available on reasonable request to the authors.

REFERENCES

Andersson E. P., Agertz O., Renaud F., Teyssier R., 2023, MNRAS, 521, 2196
Berg D. A. et al., 2022, ApJS, 261, 31
Cameron A. J. et al., 2021, ApJ, 918, L16
Chisholm J., Tremonti C., Leitherer C., 2018, MNRAS, 481, 1690
Deng Y., Li H., Liu B., Kannan R., Smith A., Bryan G. L., 2024, A&A, 691, A231
Emerick A., Bryan G. L., Mac Low M.-M., Côté B., Johnston K. V., O'Shea B. W., 2018, ApJ, 869, 94
Emerick A., Bryan G. L., Mac Low M.-M., 2019, MNRAS, 482, 1304
Emerick A., Bryan G. L., Mac Low M.-M., 2020, ApJ, 890, 155
Escala I. et al., 2018, MNRAS, 474, 2194
Forbes J. C., Krumholz M. R., Speagle J. S., 2019, MNRAS, 487, 3581
Fukushima K., Kobayashi S. B., Matsushita K., 2024, A&A, 686, 96
Gentry E. S., Krumholz M. R., Madau P., Lupi A., 2019, MNRAS, 483, 3647
Grimes J. P., Heckman T., Strickland D., Ptak A., 2005, ApJ, 628, 187
Gutcke T. A., Pakmor R., Naab T., Springel V., 2021, MNRAS, 501, 5597
Hamel-Bravo M. J. et al., 2024, MNRAS, 530, 3855
Harris C. R. et al., 2020, Nature, 585, 357
He C.-C., Wibking B. D., Krumholz M. R., 2024, MNRAS, 531, 1228
Hu C.-Y., 2019, MNRAS, 483, 3363
Hu C.-Y., Naab T., Walch S., Glover S. C. O., Clark P. C., 2016, MNRAS, 458, 3528
Hu C.-Y., Naab T., Glover S. C. O., Walch S., Clark P. C., 2017, MNRAS, 471, 2151
Hu C.-Y. et al., 2023a, ApJ, 950, 132
Hu C.-Y., Sternberg A., van Dishoeck E. F., 2023b, ApJ, 952, 140
Hunter J. D., 2007, Comput. Sci. Eng., 9, 90
Johnson J. W., Weinberg D. H., 2020, MNRAS, 498, 1364
Kennicutt R. C. Jr, 1998, ARA&A, 36, 189
Kim C.-G. et al., 2020, ApJ, 900, 61
Konami S., Matsushita K., Tsuru T. G., Gandhi P., Tamagawa T., 2011, PASJ, 63, S913
Kravtsov A., Manwadkar V., 2022, MNRAS, 514, 2667
Li J.-T., Wang Q. D., 2013, MNRAS, 435, 3071
Lopez L. A., Mathur S., Nguyen D. D., Thompson T. A., Olivier G. M., 2020, ApJ, 904, 152
Lopez S., Lopez L. A., Nguyen D. D., Thompson T. A., Mathur S., Bolatto A. D., Vulic N., Sardone A., 2023, ApJ, 942, 108
Martin C. L., Kobulnicky H. A., Heckman T. M., 2002, ApJ, 574, 663
Melioli C., de Gouveia Dal Pino E. M., Geraissate F. G., 2013, MNRAS, 430, 3235

Origlia L., Ranalli P., Comastri A., Maiolino R., 2004, ApJ, 606, 862
Owen R. A., Warwick R. S., 2009, MNRAS, 394, 1741
Peeples M. S., Shankar F., 2011, MNRAS, 417, 2962
Peeples M. S., Werk J. K., Tumlinson J., Oppenheimer B. D., Prochaska J. X., Katz N., Weinberg D. H., 2014, ApJ, 786, 54
Ranalli P., Comastri A., Origlia L., Maiolino R., 2008, MNRAS, 386, 1464
Rathjen T.-E. et al., 2021, MNRAS, 504, 1039
Read A. M., Stevens I. R., 2002, MNRAS, 335, L36
Rey M. P., Katz H. B., Cameron A. J., Devriendt J., Slyz A., 2024, MNRAS, 528, 5412
Schneider E. E., Mao S. A., 2024, ApJ, 966, 37
Schneider E. E., Robertson B. E., Thompson T. A., 2018, ApJ, 862, 56
Schneider E. E., Ostriker E. C., Robertson B. E., Thompson T. A., 2020, ApJ, 895, 43
Sharda P., Krumholz M. R., Wisnioski E., Forbes J. C., Federrath C., Acharyya A., 2021, MNRAS, 502, 5935
Shen S., Wadsley J., Stinson G., 2010, MNRAS, 407, 1581
Smith R. K., Brickhouse N. S., Liedahl D. A., Raymond J. C., 2001, ApJ, 556, L91
Smith B. D. et al., 2017, MNRAS, 466, 2217
Smith M. C., Bryan G. L., Somerville R. S., Hu C.-Y., Teyssier R., Burkhart B., Hernquist L., 2021, MNRAS, 506, 3882
Smith M. C. et al., 2024, MNRAS, 527, 1216
Steinwandel U. P., Rennehan D., Orr M. E., Fielding D. B., Kim C.-G., 2024a, preprint (arXiv:2407.14599)
Steinwandel U. P., Kim C.-G., Bryan G. L., Ostriker E. C., Somerville R. S., Fielding D. B., 2024b, ApJ, 960, 100
Strickland D. K., Heckman T. M., Colbert E. J. M., Hoopes C. G., Weaver K. A., 2004, ApJ, 606, 829
Sutherland R. S., Dopita M. A., 1993, ApJS, 88, 253
Telford O. G., Werk J. K., Dalcanton J. J., Williams B. F., 2019, ApJ, 877, 120
Toalá J. A., Arthur S. J., 2018, MNRAS, 478, 1218
Tüllmann R., Pietsch W., Rossa J., Breitschwerdt D., Dettmar R. J., 2006, A&A, 448, 43
Tumlinson J. et al., 2011, Science, 334, 948
Turk M. J., Smith B. D., Oishi J. S., Skory S., Skillman S. W., Abel T., Norman M. L., 2011, ApJS, 192, 9
Vijayan A., Li M., 2022, MNRAS, 510, 568
Vijayan A., Sarkar K. C., Nath B. B., Sharma P., Shchekinov Y., 2018, MNRAS, 475, 5513
Vijayan A., Krumholz M. R., Wibking B. D., 2024, MNRAS, 527, 10095 QED I
Vijayan A., Krumholz M. R., Wibking B. D., 2025, MNRAS, preprint (arXiv:2502.00929)
Wang Q. D., Li J., Jiang X., Fang T., 2016, MNRAS, 457, 1385
Wibking B. D., Krumholz M. R., 2022, MNRAS, 512, 1430
Xu X. et al., 2022, ApJ, 933, 222
ZuHone J. A., Hallman E. J., 2016, Astrophysics Source Code Library, record ascl:1608.002
ZuHone J. A., Vikhlinin A., Tremblay G. R., Randall S. W., Andrade-Santos F., Bourdin H., 2023, Astrophysics Source Code Library, record ascl:2301.024

APPENDIX A: CONVERGENCE IN TIME AND RESOLUTION

Here, we reproduce convergence results from QED I to demonstrate that our simulations have settled to steady state in time and with respect to the simulation resolution. To evaluate the former, in Fig. A1 we plot the mass and metal outflow rates as a function of the distance from the mid-plane (averaging the $+z$ and $-z$ directions) averaged over three time intervals: 90–105, 105–110, and 110–115 Myr. The figure shows that, while there are stochastic fluctuations, as expected given that SNe are injected stochastically, the outflow rates are fluctuating above the mean and are not systematically increasing or decreasing with time. This indicates that we have reached statistical steady state.

To check for convergence with respect to resolution, in Fig. A2 we show profiles of the mass and metal loading factors (defined by normalizing the outflow rates to the injection rates; see Vijayan et al. 2025) as a function of height for runs at resolutions of 32, 16, 8, 4, and 2 pc, all averaged over the same time interval used in QED I and again averaging the $+z$ and $-z$ directions together. In this plot, lines indicate the median and shaded regions mark the 16th to 84th percentile variation over time. We see that, although there are significant fluctuations (again as expected since SN events occur stochastically), the direction of change is no longer *monotonic* once the grid spacing reaches ~ 2 pc. Instead, the profiles oscillate around a well-defined mean, indicating that key outflow properties have converged to within modest fluctuations.

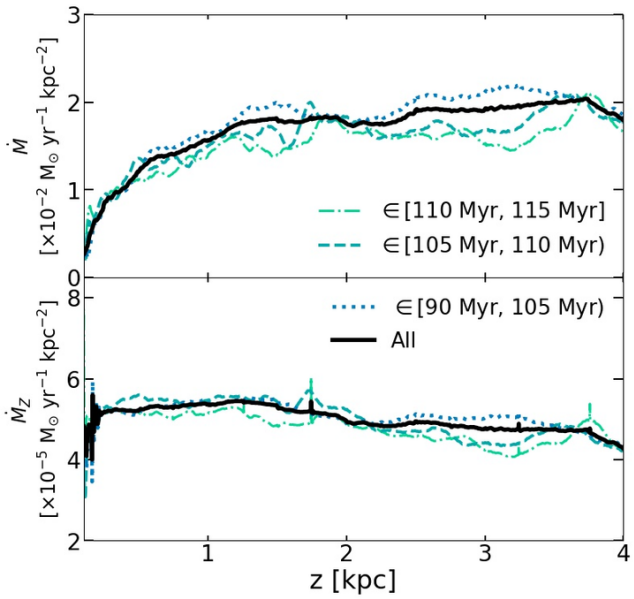

Figure A1. Time-averaged mass outflow rate (upper panel) and metal outflow rate (lower panel) versus height, for several time intervals between 90 and 115 Myr as indicated in the legend.

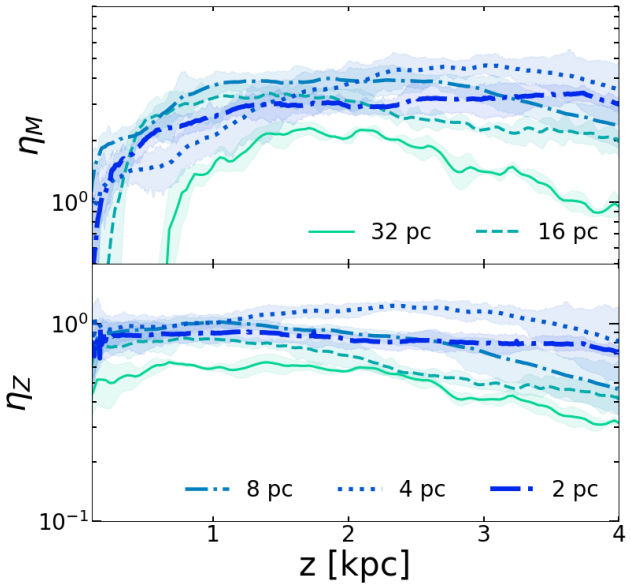

Figure A2. Mass (top panel) and metal (bottom panel) loading factors versus height for different spatial resolutions, measured at times and using averaging intervals identical to those in QED I. Lines represent mean profiles, while shaded bands show the 16th–84th percentile spread over time.

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