The Metallicity of the Intracluster Medium Over Cosmic Time: Further Evidence for Early Enrichment

Adam B. Mantz,[1,2*] Steven W. Allen,[1,2,3] R. Glenn Morris,[1,3] Aurora Simionescu,[4] Ondrej Urban,[1,2] Norbert Werner,[5,6,7] Irina Zhuravleva[1,2]

[1] *Kavli Institute for Particle Astrophysics and Cosmology, Stanford University, 452 Lomita Mall, Stanford, CA 94305, USA*
[2] *Department of Physics, Stanford University, 382 Via Pueblo Mall, Stanford, CA 94305, USA*
[3] *SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, USA*
[4] *Institute of Space and Astronautical Science (ISAS), JAXA, 3-1-1 Yoshinodai, Chuo-ku, Sagamihara, Kanagawa, 252-5210, Japan*
[5] *MTA-Eötvös University Lendület Hot Universe Research Group, Pázmány Péter sétány 1/A, Budapest, 1117, Hungary*
[6] *Department of Theoretical Physics and Astrophysics, Faculty of Science, Masaryk University, Kotlářská 2, Brno, 611 37, Czech Republic*
[7] *School of Science, Hiroshima University, 1-3-1 Kagamiyama, Higashi-Hiroshima 739-8526, Japan*

arXiv:1706.01476v3 [astro-ph.CO] 14 Sep 2017

ABSTRACT

We use *Chandra* X-ray data to measure the metallicity of the intracluster medium (ICM) in 245 massive galaxy clusters selected from X-ray and Sunyaev-Zel'dovich (SZ) effect surveys, spanning redshifts $0 < z < 1.2$. Metallicities were measured in three different radial ranges, spanning cluster cores through their outskirts. We explore trends in these measurements as a function of cluster redshift, temperature, and surface brightness "peakiness" (a proxy for gas cooling efficiency in cluster centers). The data at large radii (0.5–$1\,r_{500}$) are consistent with a constant metallicity, while at intermediate radii (0.1–$0.5\,r_{500}$) we see a late-time increase in enrichment, consistent with the expected production and mixing of metals in cluster cores. In cluster centers, there are strong trends of metallicity with temperature and peakiness, reflecting enhanced metal production in the lowest-entropy gas. Within the cool-core/sharply peaked cluster population, there is a large intrinsic scatter in central metallicity and no overall evolution, indicating significant astrophysical variations in the efficiency of enrichment. The central metallicity in clusters with flat surface brightness profiles is lower, with a smaller intrinsic scatter, but increases towards lower redshifts. Our results are consistent with other recent measurements of ICM metallicity as a function of redshift. They reinforce the picture implied by observations of uniform metal distributions in the outskirts of nearby clusters, in which most of the enrichment of the ICM takes place before cluster formation, with significant later enrichment taking place only in cluster centers, as the stellar populations of the central galaxies evolve.

Key words: galaxies: clusters: intracluster medium – X-rays: galaxies: clusters

1 INTRODUCTION

The deep gravitational potential wells associated with clusters of galaxies retain nearly all of the baryonic matter involved in their formation or accreted later (e.g. Allen, Evrard, & Mantz 2011; Borgani & Kravtsov 2011), including the metals produced in stars and ejected from galaxies into the ICM (e.g. Böhringer & Werner 2010). The abundances and distribution of metals in the ICM thus encodes the history of star formation in cluster galaxies, as well as the processes that eject metals from the galaxies themselves and mix them with the surrounding gas.

* E-mail: amantz@slac.stanford.edu

It was shown early on that the ICM is enriched to about 0.3–0.5 of the Solar metallicity (Mushotzky et al. 1978; we adopt the Solar relative abundance values of Asplund et al. 2009 throughout this work). Later observations with higher spatial resolution determined that the metallicity in nearby clusters is often centrally peaked, declining to ~ 0.3 Solar outside of cluster centers (Allen & Fabian 1998; De Grandi et al. 2004; Leccardi & Molendi 2008). The most sensitive measurements of metallicity in the outskirts of galaxy clusters to date, extending to the virial radius, are from *Suzaku* observations of nearby, X-ray bright clusters. *Suzaku* Key Project observations of the outskirts of the Perseus cluster revealed a spatially homogenous metal distribution at $r > 0.2\,r_{200}$, with an iron abundance with respect to Solar

of $Z/Z_\odot = 0.314 \pm 0.012$ (Werner et al. 2013). Consistent metallicity measurements have been obtained from *Suzaku* observations of the outskirts of Coma (Simionescu et al. 2013), Virgo (Simionescu et al. 2015), and other nearby clusters (Urban et al. 2017) and groups (Thölken et al. 2016). The uniformity and universality of the metallicity in cluster outskirts argues for early enrichment of the ICM, with most of the metals present being produced and mixed prior to the development of steep entropy gradients in clusters.

A natural consequence of this model is that, beyond their inner regions ($r \gtrsim 0.5\,r_{500}$), we should see no significant evolution in the metal content of clusters, at least out to redshifts $z \gtrsim 2$. Directly verifying this hypothesis is challenging, however, because the current X-ray observatories with sufficient spatial resolution to measure core-excluded metallicities in distant clusters suffer from higher backgrounds than *Suzaku*, while using similar (relatively) low-energy-resolution CCD imaging spectrometers. Furthermore, cluster catalogs with available X-ray data that extend to high redshifts necessarily comprise the most massive clusters, with ICM temperatures $\gtrsim 3\,\mathrm{keV}$; continuum emission is consequently significant, and only the Fe-K emission lines are typically detectable at CCD resolution. Consequently, the metallicity outside of the cores of cosmologically distant clusters is a low signal-to-noise observable, and will remain so pending the launch of new facilities with much larger collecting areas, good spatial resolution, and high energy resolution, such as *Athena* or *Lynx*. Previous studies of evolution in the metallicity of the ICM have yielded mixed results (e.g. Balestra et al. 2007; Maughan et al. 2008; Anderson et al. 2009), although more recent results using XMM-*Newton* and *Chandra* are consistent with little or no evolution (Ettori et al. 2015; McDonald et al. 2016).

Here we revisit this question using the largest cluster sample to date, 245 objects selected from X-ray and SZ surveys and observed by *Chandra*, spanning redshifts $0 < z < 1.2$. We perform this analysis for three radial ranges, with the smallest radii ($r < 0.1\,r_{500}$) being dominated by cool cores when they exist, and the largest radii ($r > 0.5\,r_{500}$) comfortably excluding the central gradients in metallicity previously observed. The latter measurement is challenging, but yields the most direct test of the predictions from *Suzaku* regarding the metallicity in cluster outskirts. In Section 2, we introduce the *Chandra* data set and the analysis methods used in this work. Section 3 presents our results on the average cluster metallicity and its evolution as a function of cluster radius, temperature and morphology. We discuss the consequences of these results in Section 4 and conclude in Section 5. The appendices address the issue of fit statistics used for sparse X-ray spectra (the C statistic versus χ^2) and resulting biases (Appendix A), and the cross calibration of *Chandra* and *Suzaku* metallicity measurements (Appendix B).

2 DATA AND ANALYSIS

The sample of clusters employed here consists of several differently selected subsamples, and contains most of the known massive clusters that have been observed with *Chandra*. Specifically, it contains clusters with *Chandra* data that are either
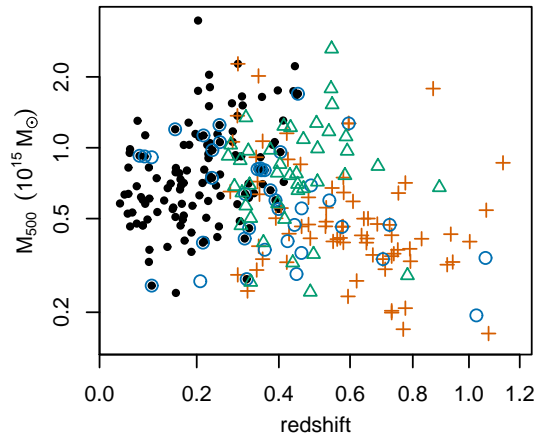

Figure 1. Mass–redshift distribution of our cluster data set, where we neglect error bars for clarity. Symbols are as follows: black, filled circles are selected from the X-ray flux limited BCS, REFLEX and bright MACS samples, with the further requirement that their 0.1–2.4 keV RASS luminosities exceed $2.5 \times 10^{44}\,\mathrm{erg\,s^{-1}}$ (see Mantz et al. 2016); blue, empty circles correspond to massive, dynamically relaxed clusters identified from the *Chandra* archive; red crosses are clusters detected by SPT; and green triangles are X-ray selected clusters that do not fall into any of the other categories.

(i) used in studies of cluster cosmology and scaling relations by Mantz et al. (2015b, 2016). This subsample is formed from the ROSAT All-Sky Survey (RASS) based BCS (Ebeling et al. 1998), REFLEX (Böhringer et al. 2004), and Bright MACS (Ebeling et al. 2010) catalogs, with a minimum 0.1–2.4 keV luminosity (as measured from RASS) of $2.5 \times 10^{44}\,\mathrm{erg\,s^{-1}}$.

(ii) identified as massive and dynamically relaxed based on a morphological search of the *Chandra* archive (see Mantz et al. 2014, 2015a).

(iii) detected through their SZ effect by the South Pole Telescope (SPT; Bleem et al. 2015). Note that *Chandra* follow-up preferentially targeted the most massive SPT clusters.

(iv) X-ray selected from the Faint MACS catalog (Ebeling et al. 2007; Mann & Ebeling 2012), the BCS (extending to a lower flux limit than was used by Mantz et al. 2015b), and from earlier samples of relaxed clusters (in particular, that of Allen et al. 2008).

In addition, we require the temperature measured from our analysis, below, at intermediate radii (0.1–$0.5\,r_{500}$) to be $\geqslant 5\,\mathrm{keV}$, in order to ensure that the sample consists of genuinely massive clusters (this cut only eliminates 14 clusters from the sample, given the initial selection). The *Chandra* data for all clusters in the sample have been analyzed uniformly, as described below. Figure 1 shows the distribution in redshift and mass of the different subsamples.

Our reduction of the *Chandra* data follows the procedure detailed by Mantz et al. (2015a), with the exception that this work employs a more recent version of the *Chandra* calibration files (specifically, CALDB[1] version 4.6.7). In brief, the raw data were reprocessed to produce level 2 event files,

---

[1] `http://cxc.harvard.edu/caldb/`

and were filtered to eliminate periods of high background. For each observation, a corresponding quiescent background data set was produced using the *Chandra* blank-sky data,[2] rescaled according to the measured count rate in the 9.5–12 keV band. Each cluster field was then tested for the presence of a soft Galactic foreground, which is absent from the blank-sky data sets, as described in Mantz et al. (2014). When present, this foreground component was constrained simultaneously with the cluster model in our subsequent spectral analysis.

A cluster's mass and its associated characteristic radius are jointly defined in terms of the critical density at its redshift, $M_\Delta = (4/3)\pi\Delta\rho_{\rm cr}(z)r_\Delta^3$, for the conventional "overdensity" of $\Delta = 500$. In this work, we are only interested in obtaining estimates of r_{500} for each cluster so that constraints on the metallicity in comparable regions can be extracted. We obtained these r_{500} estimates following the procedure described by Mantz et al. (2016), which is summarized below. After defining the center of the X-ray emission, spectra were extracted in concentric annuli, binned to have at least 1 count per channel, and were used to fit in XSPEC[3] a non-parametric, spherically symmetric model for the three-dimensional gas density in a given cluster. Cluster emission was modeled using the APEC plasma model (ATOMDB version 2.0.2). Relative metal abundances were fixed to the solar ratios of Asplund et al. (2009), with the overall metallicity allowed to vary. Photoelectric absorption by Galactic gas was accounted for using the PHABS model, employing the cross sections of Balucinska-Church & McCammon (1992). For each cluster field, the equivalent absorbing hydrogen column densities, $N_{\rm H}$, were fixed to the values from the H I survey of Kalberla et al. (2005), provided that the published values were $< 10^{21}$ cm^{-2}. When the published column densities were $\geqslant 10^{21}$ cm^{-2}, we included $N_{\rm H}$ as a free parameter fitted to the X-ray data. The likelihood of spectral models was evaluated using the Cash (1979) statistic, as modified by Arnaud et al.[4] (the C-statistic) to account for the use of an empirical background model (i.e. the blank-sky data). Confidence regions were determined by Markov Chain Monte Carlo (MCMC) explorations of the parameter space using the LMC code.[5] From the resulting constraints on the enclosed gas mass as a function of radius, we can obtain an estimate of r_{500}, based on the integrated gas mass fraction of $f_{\rm gas}(r_{500}) = 0.125$ measured by Mantz et al. (2016), by finding the radius that satisfies

$$M_{\rm gas}(r) = \frac{4\pi}{3}\, 500\rho_{\rm cr}(z)f_{\rm gas}(r_{500})r^3. \qquad (1)$$

Given an estimate of r_{500} for each cluster, we extract new spectra in each of three annular apertures, spanning radii of 0.0–0.1 r_{500}, 0.1–0.5 r_{500} and 0.5–1.0 r_{500}. Each spectrum is fitted independently using the spectral analysis described above. In this case, however, the cluster emission in each region is modeled by a single APEC emission component with free temperature, metallicity and normalization parameters; Galactic absorption and foregrounds (when present)

are handled as before. We allow the temperature, kT, and metallicity, Z, of the cluster model to vary over the ranges 0.1–64 keV and 0.0–5.0 $Z_\odot$, respectively. Because the apertures used in these fits are defined to cover comparable regions of each cluster rather than being based on signal-to-noise considerations, some cluster spectra are inadequate to constrain the model. Such unconstrained fits could potentially bias our results, given the uniform and asymmetric nature of the priors on kT and Z relative to typical values of ~ 6 keV and $\sim 0.3\, Z_\odot$, even though clusters with large uncertainties would naturally be down-weighted in our subsequent analysis. In practice, we find that unconstrained spectral fits can be automatically identified by applying generous cuts on the recovered parameters, specifically by requiring values of $kT > 30$ keV and $Z > 2\, Z_\odot$ to be ruled out at the 95 per cent confidence level.[6] Reasonable adjustments of these cuts have no impact on our results, due to the fact that the few measurements greater or fewer that are excluded have little statistical leverage. At the lowest redshifts, the limited *Chandra* field of view also sometimes prevents the larger apertures from being included in our analysis at all.

In addition to redshift and X-ray spectral measurements, we characterize the clusters in our sample morphologically, using measurements of the surface brightness 'peakiness' from Mantz et al. (2015a). This metric is a proxy for the 'cool cores' in clusters, i.e. the presence of significant density enhancements and temperature decrements that appear in the centers of some clusters. The upper-right panel of Figure 2 explicitly verifies this relationship, showing peakiness against the ratio of the temperatures measured in the 0.0–0.1 r_{500} and 0.1–0.5 r_{500} apertures. The lower-left triangle portion of the figure shows the distribution of our cluster sample in redshift, peakiness and the temperature at intermediate radii (0.1–0.5 r_{500}), along with histograms of these parameters.

In Section 3, we report the results of fitting a model for the mean cluster metallicity as a function of redshift and temperature. Simple curve fitting techniques such as least squares are inadequate for this task, due to the non-zero measurement covariance between kT and Z measured from the same spectrum, and the often non-Gaussian shape of the posterior distribution of Z (resulting from the prior cut at $Z = 0$). We therefore make use of the full set of MCMC samples for each cluster when fitting these higher-level (hierarchical) models. In detail, this means that the samples for each cluster are importance weighted according to the hierarchical model (e.g. a mean Z and log-normal intrinsic scatter) and then numerically integrated to provide a final posterior probability. The sampling itself was performed using RGW,[7] which implements the algorithm of Goodman & Weare (2010) in the R environment.[8]

---

[2] http://cxc.cfa.harvard.edu/ciao/threads/acisbackground/
[3] http://heasarc.gsfc.nasa.gov/docs/xanadu/xspec/
[4] XSPEC Users' Guide, Appendix B: https://heasarc.gsfc.nasa.gov/docs/xanadu/xspec/manual/manual.html
[5] https://github.com/abmantz/lmc

[6] The only exception to this rule is CL J1415+3612, which in the 0.0–0.1 r_{500} region has $kT = (6.1 \pm 0.4)$ keV and $Z = (1.9 \pm 0.4)\, Z_\odot$ (see also Santos et al. 2012). Even given the large cluster-to-cluster scatter in this central region (Section 3.1), this value is a sufficiently extreme outlier that we remove it from subsequent analysis.
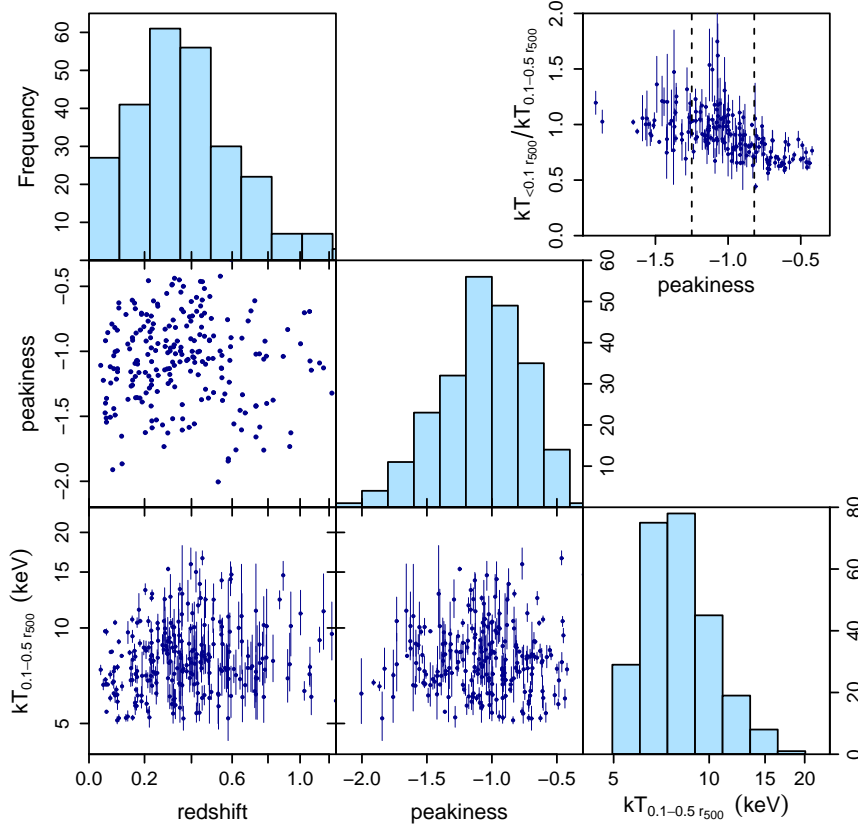[7] https://github.com/abmantz/rgw/
[8] https://www.r-project.org

Figure 2. Properties of our cluster sample. The lower-left triangle shows how the clusters are distributed in terms of redshift, temperature at intermediate radii, and surface brightness peakiness, including histograms of these parameters. The upper-right panel shows peakiness against the temperature in cluster centers, displaying the expected trend wherein cool-core clusters tend to have sharper surface brightness peaks. The vertical, dashed lines divide the sample into 3 peakiness ranges that we refer to in this analysis.

3 RESULTS

Our measurements of metallicity in each aperture are plotted against redshift, temperature in the corresponding region, and peakiness in Figure 3. By eye, the clearest trends are between the central $(0.0$–$0.1\, r_{500})$ metallicity, temperature and peakiness. However, these trends are not independent, due to the relationship between peakiness and the central temperature (top-right panel of Figure 2, top-center panel of Figure 3). Consequently, we consider models where metallicity varies with redshift and temperature, but do not explicitly include peakiness as an additional explanatory variable. However, in subsequent discussion, we will contrast subsamples of clusters in 3 categories: those with peakiness values of $p \geqslant -0.82$, $-1.25 \leqslant p < -0.82$, and $p < -1.25$. Henceforth, we refer to these groups as high (H), medium (M), and low-peakiness (L) clusters.

To be precise, we fit the model

$$Z = Z_0 \left(\frac{1+z}{1+z_{\rm piv}} \right)^{\beta_{1+z}} \left(\frac{kT}{kT_{\rm piv}} \right)^{\beta_{kT}}, \qquad (2)$$

including a log-normal intrinsic scatter, $\sigma_{\ln Z}$, where Z and kT apply to a particular aperture. The pivot redshift and temperature, $z_{\rm piv}$ and $kT_{\rm piv}$, can be chosen in each case to minimize the posterior correlation of Z_0 with β_{1+z} and β_{kT}. Table 1 shows the parameter constraints obtained by fitting

the *Chandra* measurements corresponding to each aperture (without selecting based on peakiness); 68.3 per cent confidence bands for the mean metallicity trends are shown in Figure 3. Results from fitting the 3 peakiness-based subsets of clusters in each aperture appear in Table 2. In the following subsections, we discuss in more detail the results for each aperture.

3.1 Small Radii $(0.0$–$0.1\, r_{500})$

In cluster centers, a trend towards higher metallicity in cool/peaky cores is clear (β_{kT} is almost 6σ different from zero in Table 1). Controlling for this trend, our measurements are consistent with a constant value with redshift, although with a significant intrinsic scatter (0.18 ± 0.02 in $\ln Z$). Breaking these features down in terms of peakiness, we see that the intrinsic scatter and temperature-dependence are significantly non-zero only for the H and M subsamples (Table 2). For the L subsample, the scatter and temperature dependence are both consistent with zero. As a function of redshift, the H sample is consistent with a constant, while the L sample evolves strongly, with the M sample presenting an intermediate case (a $\sim 1.5\sigma$ departure from zero).

The high metallicity and lack of evolution in the centers of cool-core clusters (H), and the correlation with peakiness, suggest substantial early enrichment of the coolest,
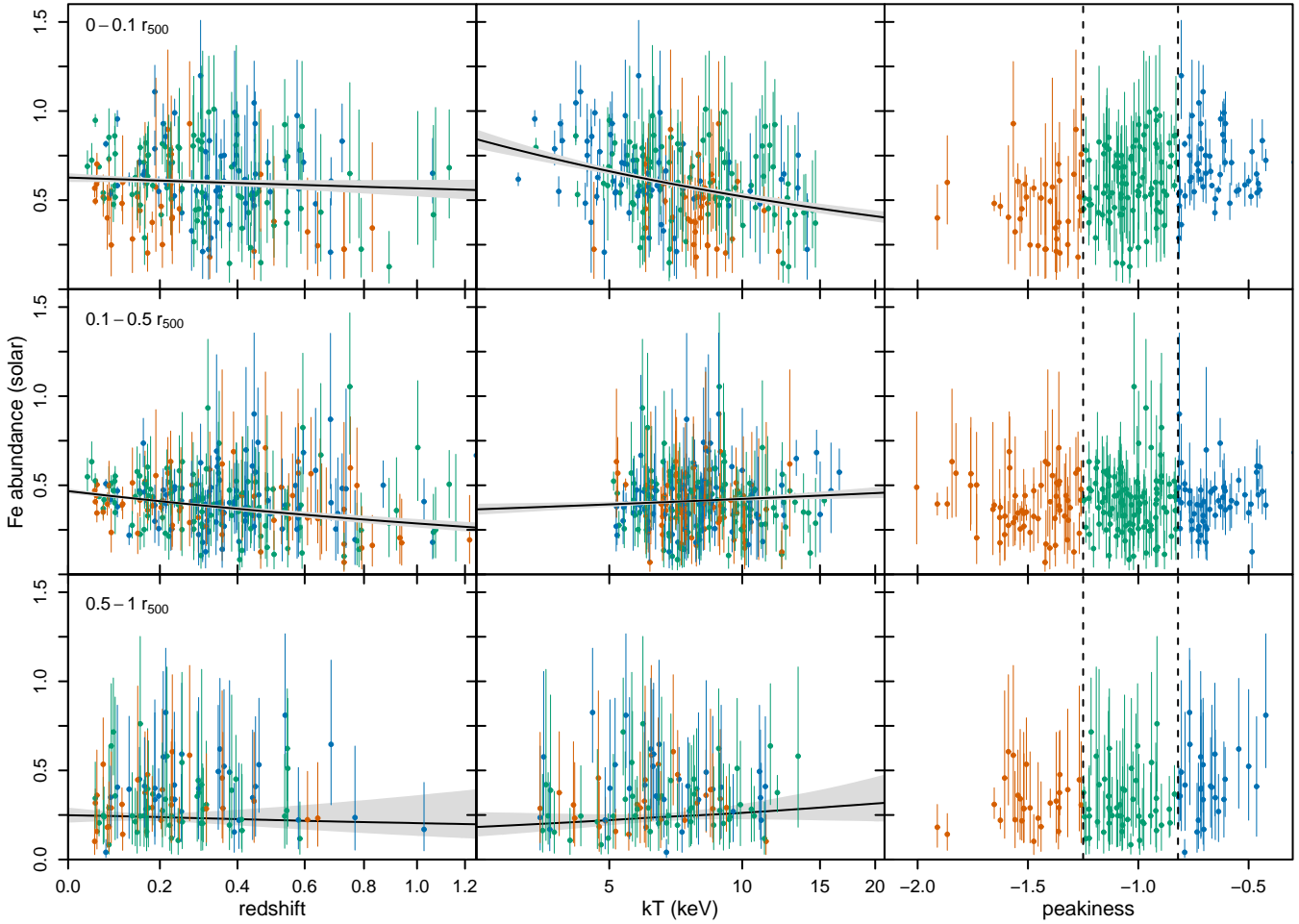
Figure 3. Measured metallicities in three apertures (rows) are plotted as a function of redshift, temperature, and peakiness (columns). Points are color coded according to peakiness (right column) and correspond to the posterior median for each cluster; error bars span the 15.85th–84.15th percentiles (68.3 per cent confidence). In the left and center columns, black lines and gray shading similarly show the median and 68.3 per cent confidence region for the mean metallicity power-laws with $1 + z$ or kT (Equation 2), evaluated at the pivot temperature or redshift. In the right column, vertical, dashed lines show the thresholds used to define subsamples of clusters based on peakiness.

Table 1. Constraints on the model of Equation 2 in three apertures from our *Chandra* data. [1] Radial range in units of r_{500}; [2–3] pivot redshift and temperature (in keV) corresponding to the normalization, Z_0; [4] number of clusters contributing to the fit for a given aperture; [5–8] best-fitting values (posterior modes) and 68.3 per cent maximum-probability confidence intervals for the model parameters.

Aperture	z_{piv}	kT_{piv}	N_{cl}	$Z_0/Z_\odot$	β_{1+z}	β_{kT}	$\sigma_{\ln Z}$
0.0–0.1	0.23	6.4	186	0.607 ± 0.012	-0.14 ± 0.17	-0.35 ± 0.06	0.18 ± 0.02
0.1–0.5	0.19	8.0	245	0.413 ± 0.007	-0.71 ± 0.15	0.10 ± 0.07	0.08 ± 0.02
0.5–1.0	0.17	6.7	86	0.240 ± 0.022	-0.30 ± 0.91	0.22 ± 0.34	$0.00^{+0.17}_{-0.00}$

lowest-entropy gas through vigorous star formation, with subsequent enrichment by aging stellar populations playing a relatively minor role. The large scatter within this population implies that properties of the clusters beyond their redshifts and temperatures, perhaps related to accretion rates and the strength of AGN feedback, can also have a significant influence on central metallicities. In clusters without cool cores (L), the scatter in central metallicity and its dependence on central temperature are minimal. At the same time, this cluster population steadily accumulates metals in cluster centers, presumably due to the continuing evolution and mass loss of stellar populations in the central galaxies. Clusters with medium-strength cores (M) lie between these two extremes.

Table 2. Constraints on the model of Equation 2 in three apertures and for peakiness-based subsets of the cluster sample from our *Chandra* data. Columns are as in Table 1, except that the first column here indicates whether the constraints apply to high (H: $p \geqslant -0.82$), medium (M: $-1.25 \leqslant p < -0.82$), or low (L: $p < -1.25$) peakiness clusters.

Peakiness	Aperture	$z_{\rm piv}$	$kT_{\rm piv}$	$N_{\rm cl}$	$Z_0/Z_\odot$	β_{1+z}	β_{kT}	$\sigma_{\ln Z}$
H	0.0–0.1	0.31	5.5	53	0.652 ± 0.022	-0.08 ± 0.28	-0.24 ± 0.13	0.19 ± 0.03
H	0.1–0.5	0.24	8.2	56	0.421 ± 0.013	-0.44 ± 0.26	0.40 ± 0.14	$0.00^{+0.10}_{-0.00}$
H	0.5–1.0	0.21	5.9	23	0.370 ± 0.080	$-0.38^{+1.31}_{-2.05}$	$-0.81^{+0.56}_{-1.07}$	$0.00^{+0.19}_{-0.00}$
M	0.0–0.1	0.19	7.1	90	0.614 ± 0.018	-0.45 ± 0.30	-0.30 ± 0.09	0.15 ± 0.03
M	0.1–0.5	0.18	8.1	107	0.426 ± 0.009	$-0.89^{+0.15}_{-0.47}$	0.08 ± 0.10	0.07 ± 0.03
M	0.5–1.0	0.16	7.2	35	0.199 ± 0.037	0.66 ± 2.19	0.73 ± 0.67	$0.00^{+0.21}_{-0.00}$
L	0.0–0.1	0.11	7.6	33	0.489 ± 0.024	-2.07 ± 0.79	-0.41 ± 0.34	0.09 ± 0.05
L	0.1–0.5	0.14	7.6	58	0.388 ± 0.012	-0.97 ± 0.33	-0.25 ± 0.19	$0.00^{+0.08}_{-0.00}$
L	0.5–1.0	0.13	6.8	25	0.245 ± 0.037	-1.58 ± 1.64	0.67 ± 0.68	$0.00^{+0.16}_{-0.00}$

Table 3. Constraints on the model of Equation 2 in the 0.1–0.5 r_{500} aperture from our *Chandra* data, for different cluster selections. The first column indicates the selection, which is either a redshift range or 'PSZ' (see text). Otherwise, columns are as in Table 1.

Selection	$z_{\rm piv}$	$kT_{\rm piv}$	$N_{\rm cl}$	$Z_0/Z_\odot$	β_{1+z}	β_{kT}	$\sigma_{\ln Z}$
$z < 0.4$	0.16	7.9	149	0.423 ± 0.007	-0.53 ± 0.23	0.08 ± 0.08	0.08 ± 0.02
$z \geqslant 0.4$	0.55	8.7	99	0.327 ± 0.017	-0.23 ± 0.56	0.20 ± 0.21	$0.00^{+0.14}_{-0.00}$
PSZ	0.13	8.3	79	0.428 ± 0.008	-0.70 ± 0.39	0.10 ± 0.11	0.07 ± 0.02

3.2 Intermediate Radii (0.1–0.5 r_{500})

Metallicities measured at these radii lack the clear dependence on the local temperature that is present in cluster centers. They do, however, strongly prefer evolving models, irrespective of peakiness ($\beta_{1+z} = -0.71 \pm 0.15$ overall). Similarly, the intrinsic scatter is small, $\lesssim 10$ per cent, for all three peakiness-based subsamples.

However, the evidence for evolution does not come uniformly from the data at all redshifts. Splitting the sample at $z = 0.4$ (approximately the mean), we find that the $z < 0.4$ sample prefers evolution at $\sim 2\sigma$ significance, while the $z \geqslant 0.4$ sample is consistent with a constant (Table 3). The latter, however, has larger statistical uncertainties, and does not rule out the evolution implied by the lower-redshift sample. Figure 4 visualizes these fit results.

Referring to Figure 1, we see that clusters in our sample at $z < 0.4$ are predominantly X-ray selected (from RASS), while the majority at $z > 0.4$ are SZ selected (from SPT). The possibility therefore exists that the different evolutionary behavior seen at $z < 0.4$ and $z > 0.4$ is due to selection effects. To test this, we take advantage of the fact that the *Planck* SZ-selected cluster catalog overlaps substantially with both the RASS and SPT catalogs (Planck Collaboration 2016). Table 3 shows results from fitting a subset of our data set from the *Planck*-SZ catalog (PSZ), with SZ signal-to-noise > 10 (highlighted in Figure 4). This signal-to-noise threshold was chosen to provide a sample of comparable size to the others considered here, but unfortunately comes at the expense of relatively low completeness (65 per cent). The fit to this PSZ subsample, which spans $0 < z < 0.87$, is consistent with the full data set, with $\beta_{1+z} = -0.70 \pm 0.39$.

Thus, while the data are not entirely conclusive, the indications point towards a constant metallicity at radii of 0.1–0.5 r_{500} at high redshift, with metallicities beginning to rise
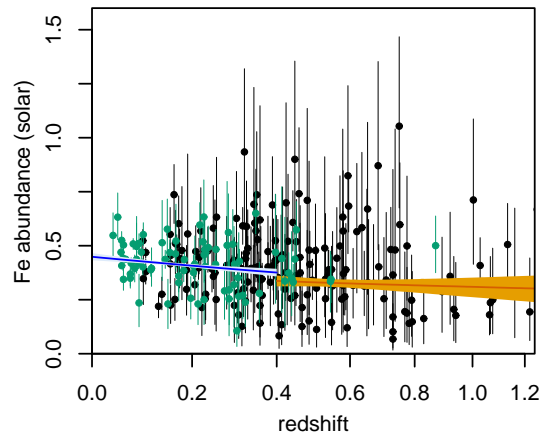

Figure 4. Measured metallicities in the 0.1–0.5 r_{500} aperture, with green points indicating those detected by *Planck* with signal-to-noise > 10. Blue and orange shading show the 68.3 per cent confidence regions for power-law evolution models fit to the data (irrespective of *Planck* selection) at $z < 0.4$ and $z \geqslant 0.4$, respectively.

at late times, independent of selection effects. Interestingly, the metallicity evolution at these radii as a function of peakiness is qualitatively similar to that seen in cluster centers (i.e. stronger evidence of evolution in less peaky clusters). In situ enrichment from stellar evolution in cluster galaxies provides a plausible mechanism in both cases, although the gas at radii of 0.1–0.5 r_{500} may also be enriched through mixing with gas from cluster centers, driven by mergers or AGN outbursts.

Table 4. Redshifts and metallicities measured from *Suzaku* data by Werner et al. (2013), Simionescu et al. (2013) and Urban et al. (2017) for hot, low-redshift clusters ($kT > 5$ keV, $z < 0.2$). The values listed here are relative to the Asplund et al. (2009) Solar abundances, and represent the average of the available measurements at cluster radii $\gtrsim 0.5\,r_{500}$. The *Suzaku* metallicity profiles typically extend to $r > r_{500}$, and are consistent with a constant value over the range averaged.

Cluster	z	$Z^{\mathrm{Suz}}/Z_{\odot}$
Perseus	0.018	0.314 ± 0.012
Coma	0.023	0.291 ± 0.037
Abell 1795	0.063	0.307 ± 0.025
Abell 2029	0.078	0.326 ± 0.065
Abell 2142	0.089	0.357 ± 0.035
Abell 2204	0.152	0.396 ± 0.069
Abell 1689	0.183	0.353 ± 0.131

3.3 Large Radii (0.5–$1.0\,r_{500}$)

Only about 35 per cent of clusters in our data set provide reliable metallicity constraints at these radii, and the statistical uncertainties are relatively large; consequently, the model parameters are much less well constrained here than in the other apertures. The *Chandra* data are consistent with a constant metallicity as a function of redshift and temperature ($\beta_{1+z} = -0.30 \pm 0.91$ and $\beta_{kT} = 0.22 \pm 0.34$), with no intrinsic scatter ($\sigma_{\ln Z} = 0.00^{+0.17}_{-0.00}$). Note that the a posteriori correlation of the scatter with the slope parameters is small, such that the upper limit on the scatter does not change if the metallicity is assumed to be constant with redshift and temperature.

Our results for this aperture can be compared with *Suzaku* measurements of metallicity at similar radii in nearby clusters. The low particle background of *Suzaku* makes it ideal for probing the faint ICM in cluster outskirts, while its large point spread function and concerns about scattered light effectively restrict its usage to radii $\gtrsim 0.5\,r_{500}$ and cluster redshifts $\lesssim 0.2$. Here we make use of measurements made from very deep observations of the Coma and Perseus clusters (Simionescu et al. 2013; Werner et al. 2013; redshifts of 0.023 and 0.018, respectively), as well as metallicity measurements for 5 clusters at redshifts $0.063 < z < 0.183$ by Urban et al. (2017), all of which satisfy the same temperature threshold as our *Chandra* data set ($kT > 5$ keV). Table 4 lists these *Suzaku* metallicity measurements.

While incorporating the *Suzaku* measurements into our evolution analysis can in principle yield tighter constraints, an overall offset cross-calibrating Fe abundance measurements with the two telescopes must be accounted for. The need for such an offset is clear from a comparison of the average metallicity measured at large radii from the two data sets: 0.240 ± 0.022 from *Chandra* and 0.314 ± 0.012 (Perseus; Werner et al. 2013) or 0.316 ± 0.012 (other clusters; Urban et al. 2017) from *Suzaku*. Note that the pivot redshift for our *Chandra* data in this aperture is relatively low (0.17), making it unlikely that this disagreement reflects true evolution of the type we are aiming to measure (see also Figure 5). A comparable offset is found by directly comparing *Chandra* and *Suzaku* measurements at large radii in the Coma cluster (see Section 3.5 and Appendix B). We include an overall

Table 5. Constraints on the model of Equation 2 (with β_{kT} fixed to zero) in the 0.5–$1.0\,r_{500}$ aperture from the combination of our *Chandra* and *Suzaku* data. The Z_0 parameter is given relative to both the *Chandra* and *Suzaku* references.

z_{piv}	0.15
N_{cl}	89
$Z_0^{\mathrm{Cha}}/Z_{\odot}$	0.234 ± 0.016
$Z_0^{\mathrm{Suz}}/Z_{\odot}$	0.342 ± 0.025
β_{1+z}	$0.35^{+1.18}_{-0.36}$
$\sigma_{\ln Z}$	$0.00^{+0.06}_{-0.00}$
$\ln(Z^{\mathrm{Suz}}/Z^{\mathrm{Cha}})$	0.375 ± 0.096

cross-calibration factor as a free parameter of the model, adopting a Gaussian prior of $\ln(Z^{\mathrm{Suz}}/Z^{\mathrm{Cha}}) = 0.41 \pm 0.14$ based on the Coma data.

Given the consistency of the *Chandra* data with a temperature-independent metallicity at these radii, we simplify the joint *Chandra*+*Suzaku* analysis by assuming $\beta_{kT} = 0$. Where measurements are available from both *Chandra* and *Suzaku*, we use the *Suzaku* data preferentially (Coma is not explicitly included in the fit, as it provides the basis of the cross-calibration prior above). The constraints obtained from this combined data set are summarized in Table 5, which lists Z_0 with and without the cross-calibration applied. The constraints on evolution, $\beta_{1+z} = 0.35^{+1.18}_{-0.36}$, are slightly tighter than those from *Chandra* alone, and are consistent with a constant value; negative values of β_{1+z} in particular are less favored. Despite the high precision of the *Suzaku* measurements, the intrinsic scatter remains consistent with zero. Our a posteriori constraint on the cross-calibration parameter is $\ln(Z^{\mathrm{Suz}}/Z^{\mathrm{Cha}}) = 0.375 \pm 0.096$, somewhat tighter than, and consistent with, the prior from Coma.

Figure 5 shows the *Suzaku* data set, along with the 68.3 per cent confidence limits for the model. For visualization, the *Chandra* data are represented by their 68.3 and 95.4 per cent confidence intervals for the Z_0 parameter (corrected to the *Suzaku* reference) when the model in Equation 2 is fit to clusters in the redshift bins shown (bounded by $z = 0.05$, 0.1, 0.2, 0.25 and 1.05) independently (fixing $\beta_{1+z} = 0$).

3.4 Synthesis

The results in Table 2 are summarized visually in Figure 6, which shows metallicity constraints in each of the apertures considered here, for H, M and L peakiness clusters, at three different redshifts spanning the range that is well sampled by our data set ($z = 0.1$, 0.2 and 0.4). Here we can see the broad trends with redshift and morphology identified in the previous sections. The profiles for H and M clusters are relatively constant with redshift, excepting the intermediate-radius aperture of M clusters, where the metallicity does increase with time. In general, the M clusters agree well with the full-sample average profile, with H clusters being marginally higher in metallicity. In contrast, at radii $< 0.5\,r_{500}$, the L clusters have significantly lower metallicities than the average, particularly in cluster centers ($r < 0.1\,r_{500}$). However, the $r < 0.5\,r_{500}$ metallicity in L clusters evolves relatively strongly as well, approaching the profiles of the other subsamples at low redshifts. At large radii, $r > 0.5\,r_{500}$, the
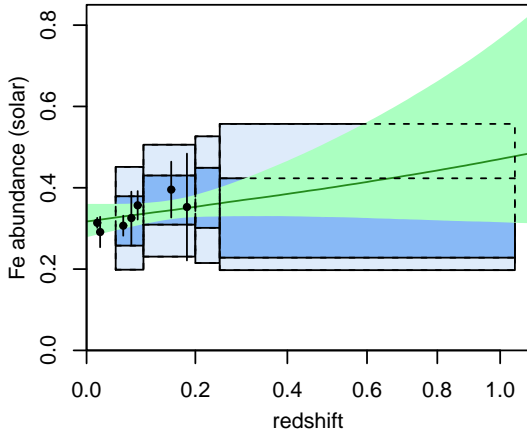
Figure 5. Measured metallicities at $r > 0.5\,r_{500}$ from *Suzaku* (black points). For visualization purposes, *Chandra* measurements in the $0.5–1.0\,r_{500}$ aperture are represented by blue-shaded boxes; these show the 68.3 and 95.4 per cent confidence intervals for the Z_0 parameter when fitting clusters in each redshift bin (fixing $\beta_{1+z} = 0$), and are corrected by the best-fitting *Suzaku/Chandra* cross calibration factor (37.5 per cent). Green shading shows the 68.3 per cent confidence regions for power-law evolution models fit to the combined data set.

three subsamples of clusters are consistent with one another, and with a constant value (full-sample intrinsic scatter of $\sigma_{\ln Z} = 0.00^{+0.17}_{-0.00}$).

3.5 Systematic Checks

In this section, we consider the impact of a number of potential systematic issues on our measurements. To begin with, we tested on simulated data the ability of our spectral fitting procedure to recover unbiased posterior distributions for temperature and metallicity as a function of signal-to-noise. Specifically, these tests apply to the use of the modified C statistic when spectra are grouped to $\geqslant 1$ count per channel, using a realistic *Chandra* background and for redshifts and temperatures typical of our data set. Details are presented in Appendix A. To summarize, we find that our approach produces accurate estimates and confidence intervals. This is in contrast to the still widely used practice of binning spectra more heavily and using the χ^2 fit statistic, which generally produces biased temperature and metallicity estimates over a range of typical exposure times (see also Leccardi & Molendi 2007; Humphrey et al. 2009). Within the *Chandra* data themselves, there is no overall trend of measured metallicity with signal-to-noise (measured in an energy band that excludes the Fe-K line).

Several more tests yielded null results. These were: shifting the normalizations of the blank-sky backgrounds by ±10 per cent; fitting two-temperature models, with common metallicity and the second temperature fixed to 0.6 keV; and using the MEKAL thermal emission model instead of APEC. Each of these adjustments resulted in shifts in the metallicity measurements for individual clusters, but the shifts were generally within the 1σ statistical uncertainties, and none of these tests impact our conclusions as a function of redshift. Our *Chandra* results thus appear to be internally robust against reasonable uncertainties in the X-ray background,
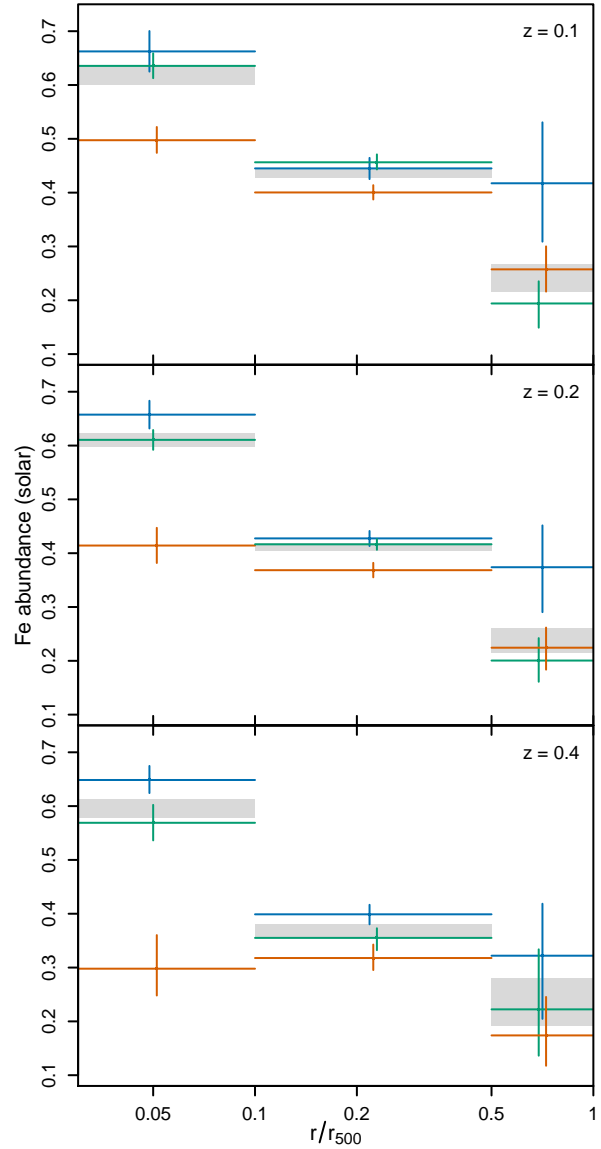

Figure 6. Metallicity profiles our constraints on the model in Equation 2 at redshifts 0.1, 0.2 and 0.4. Gray shading shows the 68.3 per cent confidence region at each redshift and radial aperture from fitting the full cluster sample, and blue, green and red crosses respectively show the same for the H, M and L subsamples. For each cluster sample and aperture, the pivot temperature (i.e. a typical temperature for the clusters in the fit) is adopted when evaluating the model.

the temperature structure of the ICM, and modelling uncertainties.

The remaining, and significant, question is whether there may be an overall bias in metallicities measured with *Chandra*, in particular with respect to the *Suzaku* measurements that we compare to in Section 3.3. Comparison of Tables 1 and 4 suggests that the average metallicity at $r > 0.5\,r_{500}$ measured from the two telescopes for clusters at the same redshifts differs substantially (i.e. a substantial correction is needed to produce the agreement shown in Figure 5). Ideally, any cross-calibration factor could be constrained by measuring the metallicity of the same cluster(s) with both instruments, at radii where scattered light is not

a concern for *Suzaku*. The best test case for such a direct comparison is the Coma cluster. Uniquely, Coma is a nearby, bright, non-centrally-peaked cluster that has been imaged out to large radii along several arms by *Suzaku* (Simionescu et al. 2013) and has a very deep ($\sim 1\,\mathrm{Ms}$) *Chandra* observation targeting radii close to $0.5\,r_{500}$.

Our analysis of this Coma data is detailed in Appendix B. In summary, the metallicity measured from *Chandra* is lower than that measured by *Suzaku* at the same radii in this cluster by ~ 40 per cent, similar to the offset implied by the sample average metallicities in the 0.5–$1.0\,r_{500}$ aperture. This analysis motivates the prior $\ln(Z^{\mathrm{Suz}}/Z^{\mathrm{Cha}}) = 0.41 \pm 0.14$ marginalized over in our joint *Chandra+Suzaku* analysis in Section 3.3. Although not definitive, the data are consistent with the possibility that charge transfer inefficiency (CTI) due to damage to the ACIS detectors aboard *Chandra* reduces the flux observed in the Fe emission line, biasing the metallicity measurements low.

4 DISCUSSION

4.1 Comparison with Previous Work

Since metallicity measurements for high-redshift clusters first became available, a number of authors have investigated the evolution of metallicity in the ICM. Table 6 lists results from the past ~ 10 years where a power-law in $1 + z$ was fitted to the data. There is a wide variation in the precise aperture adopted, the distribution of cluster redshifts, the analysis methods and the X-ray observatories used among these works, and correspondingly a range of results, with some indicating strong evolution in metallicity and some being consistent with a constant. Additional contributions beyond those listed in the table have been made by, e.g., Leccardi & Molendi (2008, qualitatively consistent with a constant), Maughan et al. (2008, qualitatively strong evolution), and Anderson et al. (2009, consistent with no evolution internally, but requiring strong evolution when combined with Maughan et al. 2008). The most recent works are by Ettori et al. (2015, XMM data), who find evidence for evolution in the 0.0–$0.15\,r_{500}$ (0.15–$0.4\,r_{500}$) apertures at $\sim 7\sigma$ ($\sim 2\sigma$) significance, and consistency with a constant at radii $> 0.4\,r_{500}$; and McDonald et al. (2016, *Chandra* data), who find consistency with a constant value at $\sim 1.5\sigma$ significance for a full 0.0–$1.0\,r_{500}$ aperture.

Ettori et al. (2015) find significant evolution in cluster centers ($r < 0.15\,r_{500}$), driven by cool core clusters. When considering only non-cool core clusters, they find no evidence for evolution. Both of these observations are ostensibly at odds with our results; to compare them in greater depth, we would need to investigate in detail the correspondence between our H, M and L subsamples and the cool-core/non-cool-core division employed by Ettori et al. (2015). At intermediate radii, our 0.1–$0.5\,r_{500}$ evolution constraint is in good agreement with the 0.15–$0.4\,r_{500}$ result from Ettori et al. (2015, $\beta_{1+z} = -0.71 \pm 0.15$ vs. -0.70 ± 0.32); in the outermost annulus considered, both works are consistent with a constant metallicity.

McDonald et al. (2016) used *Chandra* data for SPT selected clusters, most of which are also in our data set; indeed, nearly all of our clusters at $z > 0.6$ are from SPT. Apart from the constraint listed in Table 6, McDonald et al. (2016) present their results on evolution (in 0.0–0.15 and 0.15–$1\,r_{500}$ apertures) in terms of a linear-in-redshift model which is not directly comparable to our fits. Their constraints on this linear model are compatible with a constant metallicity in both apertures. Given that measurements in the 0.15–$1\,r_{500}$ aperture should be most directly comparable to our 0.1–$0.5\,r_{500}$ results (due to the strong weighting of surface brightness towards smaller radii), the McDonald et al. (2016) constraint at these radii is in conflict with our measurement of non-zero evolution in this aperture. However, as noted in Section 3.2, our measurement of evolution is driven by the data at redshifts $z < 0.4$, which are not well represented in an SPT-only data set (Figure 1). A better comparison is with our data set at $z > 0.4$ (or, almost equivalently, with only the SPT clusters in our data set), for which we find no evidence for evolution ($\beta_{1+z} = -0.23 \pm 0.56$; see Figure 4).

4.2 The History of ICM Enrichment

Measurements of metallicity in the Perseus and Virgo clusters revealed a uniform concentration of Fe outside of their cores and extending to their virial radii (Werner et al. 2013; Simionescu et al. 2017). In Virgo, the Si/Fe, S/Fe and Mg/Fe abundance ratios are also uniform at these radii (Simionescu et al. 2015). Since enrichment of the ICM at late times by ram-pressure stripping or galactic outflows generically predicts a non-uniform metal distribution over this radial range (e.g. Gunn & Gott 1972; Domainko et al. 2006; Matsushita et al. 2013), and since large-scale mixing is prohibited by the steep entropy gradients in clusters, these observations have been taken as evidence that most of the metals in the ICM are produced prior to cluster formation. In this scenario, metals would need to be expelled from galaxies and mixed throughout the intergalactic medium in the proto-cluster environment; once accreted onto a cluster and virialized, significant further enrichment of the ICM can only occur in regions where the stellar density is relatively high (cluster cores). This early enrichment could have been driven by galactic winds (De Young 1978) powered by AGN and supernovae (Madau et al. 1996; Brandt & Hasinger 2005). This picture is supported by the recent simulations of Fabjan et al. (2010) and Biffi et al. (2017), who find that AGN feedback at $z \gtrsim 2$, when cluster potentials are still relatively shallow, is particularly effective at mixing metals throughout the ICM, producing similar radial trends to those observed, and minimal evolution.

The lack of metallicity evolution at radii $\gtrsim 0.5\,r_{500}$ measured by Ettori et al. (2015) and in this work is consistent with early enrichment, as is the constant metallicity we find at somewhat smaller radii (0.1–$0.5\,r_{500}$) for redshifts $z > 0.4$ (see also McDonald et al. 2016). The late-time increase in metallicity that we measure in this intermediate radial range can be plausibly explained by small-scale mixing of the gas from cluster cores, where stellar evolution continues to enrich the ICM locally and where AGN feedback and cluster mergers provide a mechanism for small-scale mixing. While these observations produce a consistent picture of the history of enrichment, the precision of our metallicity constraints at high redshifts, especially at large radii, is extremely limited compared to what has been achieved in the nearby Universe with *Suzaku*. A truly definitive test of early

Table 6. Summary of metallicity evolution constraints from the literature, where a power-law in $1 + z$ is used to parametrized the evolution, for comparison to our results in Table 1. [1] Reference; [2] redshift range of the cluster sample employed; [3] radial range used to fit the metallicity evolution in units of r_{500}; [4] reported constraint on the power-law index of metallicity as a function of $1 + z$.

Paper	Redshifts	Aperture	β_{1+z}
Balestra et al. (2007)	0.3–1.3	0.15–0.30	-1.25 ± 0.15
Baldi et al. (2012)	0.4–1.4	0.00–0.15	$0.19^{+0.72}_{-0.68}$
Baldi et al. (2012)	0.4–1.4	0.15–0.40	$-0.44^{+0.78}_{-0.72}$
Ettori et al. (2015)	0.09–1.4	0.00–0.15	-1.60 ± 0.22
Ettori et al. (2015)	0.09–1.4	0.15–0.40	-0.70 ± 0.32
Ettori et al. (2015)	0.09–1.4	>0.4	-0.26 ± 0.61
McDonald et al. (2016)	0.25–1.5	0.0–1.0	-0.41 ± 0.25

enrichment will require much more precise measurements of metallicity in the diffuse ICM, ideally at the highest redshifts possible. In the near term, one could envision doing this with deep XMM observations of high-z clusters where the point source population has been constrained by *Chandra*. In the future, observatories such as *Athena* or *Lynx*, with large collecting areas and high energy resolution, will revolutionize studies of the enrichment of the ICM at high redshifts.

5   CONCLUSION

We have presented *Chandra* measurements of the metallicity of the ICM in three radial ranges, 0.0–0.1, 0.1–0.5 and 0.5–1.0 r_{500}, from a sample of 245 massive galaxy clusters. In the outermost radial range, we supplement this data set with recent, very precise measurements of nearby clusters from *Suzaku* data. We fit these data with a model that includes power-law dependences on redshift and temperature, and distinguish between three classes of clusters (H, M and L) corresponding to different ranges of surface brightness peakiness. Our results can be summarized as follows.

• In cluster centers (0.0–0.1 r_{500}), metallicity clearly correlates with temperature and peakiness, in the sense of cooler and peakier clusters being more enriched. For the peakiest clusters (subsample H), we find a large intrinsic scatter and no overall trend with redshift, indicating that the lowest-entropy gas was enriched at early times, with variations in astrophysical processes such as feedback potentially driving the scatter. For the least-peaky (L) clusters, the scatter is small, and there is a trend of increasing metallicity with time. Medium-peakiness (M) clusters form an intermediate case.

• At intermediate radii (0.1–0.5 r_{500}), we see evidence for evolution in the ICM metallicity, across morphological groups. However, this evolution appears to be limited to low redshifts, $z \lesssim 0.4$. A plausible explanation is that, given enough time, this region can become contaminated by metals produced in cluster centers, aided by small-scale mixing due to gas sloshing or AGN activity.

• At larger radii (0.5–1.0 r_{500}), the *Chandra* data are consistent with a constant metallicity value for all clusters ($\beta_{1+z} = -0.30 \pm 0.91$). We combine these data with *Suzaku* measurements of nearby clusters, accounting and fitting for a systematic offset between metallicities measured with the two telescopes. We find an average metallicity of 0.234 ± 0.016 relative to the solar abundances of Asplund

et al. (2009, or 0.342 ± 0.025, correcting to the *Suzaku* reference), and an intrinsic scatter consistent with zero (< 7 per cent at 68.3 per cent confidence). Our constraint on the power-law dependence of metallicity on $1 + z$ from these data, $\beta_{1+z} = 0.35^{+1.18}_{-0.36}$, is consistent with zero and disfavors an increasing metallicity (ongoing enrichment) at these radii.

• Our results regarding the lack of evolution in metallicity at large radii are consistent with recent work by Ettori et al. (2015) and McDonald et al. (2016). They are also consistent with the implication of the uniform metallicity in cluster outskirts measured in Perseus and Virgo (Werner et al. 2013; Simionescu et al. 2015, 2017), namely that the ICM in these regions must have been enriched to its current level of ~ 0.3 solar prior to cluster formation. Thereafter, the increases in metallicity that we find at smaller radii can be ascribed to stellar evolution products and the locally high stellar density in cluster centers. While this scenario is consistent with the current data, a definitive test of early enrichment would be best achieved through precise, spatially resolved measurements of metallicity in high-redshift clusters – ideally at the highest redshifts where such observations are feasible.

REFERENCES

Allen S. W., Evrard A. E., Mantz A. B., 2011, ARA&A, 49, 409
Allen S. W., Fabian A. C., 1998, MNRAS, 297, L63
Allen S. W., Rapetti D. A., Schmidt R. W., Ebeling H., Morris R. G., Fabian A. C., 2008, MNRAS, 383, 879
Anderson M. E., Bregman J. N., Butler S. C., Mullis C. R., 2009, ApJ, 698, 317

Asplund M., Grevesse N., Sauval A. J., Scott P., 2009, ARA&A, 47, 481

Baldi A., Ettori S., Molendi S., Balestra I., Gastaldello F., Tozzi P., 2012, A&A, 537, A142

Balestra I., Tozzi P., Ettori S., Rosati P., Borgani S., Mainieri V., Norman C., Viola M., 2007, A&A, 462, 429

Balucinska-Church M., McCammon D., 1992, ApJ, 400, 699

Biffi V. et al., 2017, MNRAS, 468, 531

Bleem L. E. et al., 2015, ApJS, 216, 27

Böhringer H. et al., 2004, A&A, 425, 367

Böhringer H., Werner N., 2010, A&AR, 18, 127

Borgani S., Kravtsov A., 2011, Adv. Sci. Lett., 4, 204

Brandt W. N., Hasinger G., 2005, ARA&A, 43, 827

Cash W., 1979, ApJ, 228, 939

De Grandi S., Ettori S., Longhetti M., Molendi S., 2004, A&A, 419, 7

De Young D. S., 1978, ApJ, 223, 47

Domainko W. et al., 2006, A&A, 452, 795

Ebeling H., Barrett E., Donovan D., Ma C.-J., Edge A. C., van Speybroeck L., 2007, ApJ, 661, L33

Ebeling H., Edge A. C., Bohringer H., Allen S. W., Crawford C. S., Fabian A. C., Voges W., Huchra J. P., 1998, MNRAS, 301, 881

Ebeling H., Edge A. C., Mantz A., Barrett E., Henry J. P., Ma C. J., van Speybroeck L., 2010, MNRAS, 407, 83

Ettori S., Baldi A., Balestra I., Gastaldello F., Molendi S., Tozzi P., 2015, A&A, 578, A46

Fabjan D., Borgani S., Tornatore L., Saro A., Murante G., Dolag K., 2010, MNRAS, 401, 1670

Goodman J., Weare J., 2010, Communications in Applied Mathematics and Computational Science, 5, 65

Gunn J. E., Gott J. R., III, 1972, ApJ, 176, 1

Humphrey P. J., Liu W., Buote D. A., 2009, ApJ, 693, 822

Kalberla P. M. W., Burton W. B., Hartmann D., Arnal E. M., Bajaja E., Morras R., Pöppel W. G. L., 2005, A&A, 440, 775

Leccardi A., Molendi S., 2007, A&A, 472, 21

Leccardi A., Molendi S., 2008, A&A, 487, 461

Madau P., Ferguson H. C., Dickinson M. E., Giavalisco M., Steidel C. C., Fruchter A., 1996, MNRAS, 283, 1388

Mann A. W., Ebeling H., 2012, MNRAS, 420, 2120

Mantz A. B., Allen S. W., Morris R. G., Rapetti D. A., Applegate D. E., Kelly P. L., von der Linden A., Schmidt R. W., 2014, MNRAS, 440, 2077

Mantz A. B., Allen S. W., Morris R. G., Schmidt R. W., von der Linden A., Urban O., 2015a, MNRAS, 449, 199

Mantz A. B. et al., 2016, MNRAS, 463, 3582

Mantz A. B. et al., 2015b, MNRAS, 446, 2205

Matsushita K., Sakuma E., Sasaki T., Sato K., Simionescu A., 2013, ApJ, 764, 147

Maughan B. J., Jones C., Forman W., Van Speybroeck L., 2008, ApJS, 174, 117

McDonald M. et al., 2016, ApJ, 826, 124

Mushotzky R. F., Serlemitsos P. J., Boldt E. A., Holt S. S., Smith B. W., 1978, ApJ, 225, 21

Planck Collaboration, 2016, A&A, 594, A27

Santos J. S., Tozzi P., Rosati P., Nonino M., Giovannini G., 2012, A&A, 539, A105

Simionescu A., Werner N., Mantz A., Allen S. W., Urban O., 2017, MNRAS, 469, 1476

Simionescu A. et al., 2013, ApJ, 775, 4

Simionescu A., Werner N., Urban O., Allen S. W., Ichinohe Y., Zhuravleva I., 2015, ApJ, 811, L25

Thölken S., Lovisari L., Reiprich T. H., Hasenbusch J., 2016, A&A, 592, A37

Urban O., Werner N., Allen S. W., Simionescu A., Mantz A., 2017, MNRAS, 470, 4583

Werner N., Urban O., Simionescu A., Allen S. W., 2013, Nat, 502, 656

APPENDIX A: FITTING SIMULATED SPECTRA

Here we show the results of testing our spectral fitting approach on simulated spectra. In particular, we contrast the use of the modified C statistic with minimally binned spectra (grouped to $\geqslant 1$ count per channel) with the practice of binning spectra more heavily ($\geqslant 20$ counts per channel, to be concrete) and using the χ^2 fit statistic (with "standard" weighting in XSPEC, i.e. with the variance set equal to the number of observed counts in a bin). Leccardi & Molendi (2007) and Humphrey et al. (2009) present more detailed discussions of the relative performance of the two statistics; however, neither specifically addresses the case of *Chandra* data and the associated practice of using blank-sky spectra to empirically model the background. In this appendix, we therefore consider that scenario for a cluster whose redshift and temperature are typical of our sample.

The starting point for our simulations is one of the intermediate-radius spectra for MACS J1115.8+0129. The redshift and absorbing column density, $z = 0.355$ and $N_\mathrm{H} = 4.34 \times 10^{20}\,\mathrm{cm}^{-2}$, are kept fixed in our analysis.[9] We adjust the best-fitting PHABS∗APEC model for this spectrum slightly, in the interest of working with round numbers, so that it has a temperature of 8 keV, a metallicity of 0.3 solar, and a normalization that corresponds to 3000 net counts on average for an exposure time of 35 ks (approximately the clean exposure of the original spectrum). Each realization consists of a Poisson draw from this model, along with the background, that is then grouped appropriately for the fit statistic being tested.

We analyzed 10^4 Poisson realizations of the science and blank-sky spectra, for science exposures ranging from 4–70 ks (nominally $\sim$ 350–7000 net counts). The simulated blank sky exposures were kept fixed at the actual exposure time of 400 ks. For each realization, we found the model parameters that minimize the fit statistic, S (where S is either C or χ^2), and determined 1-dimensional 68.3 per cent confidence intervals on both temperature and metallicity based on the range of parameter values satisfying $\Delta S = S - S_\mathrm{min} \leqslant 1$.

Figure A1 summarizes the results of these simulations. The left and center columns show smoothed histograms of the best fitting values for various science exposure times, for temperature and metallicity (columns), and for each fit

[9] Our results would certainly differ substantially if the cluster redshift needed to be simultaneously constrained from the X-ray data.

statistic (C in the first row and χ^2 in the second). The simulation input values are shown with vertical lines. The top-right panel shows the fractional bias of the median of each of those histograms as a function of simulated exposure time. The bottom-right panel shows the frequentist coverage of the $\Delta S \leqslant 1$ confidence interval for each method, i.e. the fraction of simulations for a given method and exposure where the $\Delta S \leqslant 1$ interval contains the true value, which should equal ≈ 0.683.[10]

We find that biases are minimal in the case of the C-statistic analysis, and that the coverage of the 1σ confidence intervals, constructed by requiring $\Delta C \leqslant 1$, is accurate. In agreement with the authors cited above, we find that the same cannot be said of the χ^2 analysis, even when the spectra are heavily binned and contain many thousand net counts. This is particularly the case for the temperature, for which the χ^2 best fit is biased at the ~ 5 per cent level for all the exposure times considered here, and for which the $\Delta\chi^2 \leqslant 1$ confidence intervals have significantly lower coverage than they should at the longest exposures.

Independent of the fit statistic, we also see the "piling-up" of probability at very low metallicities for low signal-to-noise spectra, as noted by Leccardi & Molendi (2007), resulting from the requirement that metallicity be non-negative. Our method for fitting Equation 2, which uses the full shape of the Bayesian posterior from each cluster, in principle accounts for both this effect and the generally asymmetric shape of many of the curves in Figure A1 (so long as the shape of the posterior is known sufficiently accurately). Methods which assume Gaussian or log-normal uncertainties, or indeed any fixed distribution *centered about the best fit*, will clearly perform badly in this situation.

APPENDIX B: COMPARING CHANDRA AND SUZAKU MEASUREMENTS OF THE COMA CLUSTER

In this appendix, we compare *Suzaku* metallicity measurements of the Coma Cluster with those from a deep (963 ks clean exposure), offset *Chandra* observation (proposal ID 17800479). The data were reduced and analyzed in the same way as the rest of the *Chandra* data used in this work.

The field targeted by this observation lies west of the center of Coma, with the four ACIS-I CCDs typically covering cluster-centric radii of $\sim 10'$–$25'$. This range spans the transition from a decreasing metallicity profile to an approximately flat one, as measured by Simionescu et al. (2013) with *Suzaku*. To conservatively exclude the decreasing part of the profile, we initially consider only the data at $r > 13'$. Given the possibility that CTI may influence the metallicity measurements, we additionally restrict the analysis to CCDs which lie entirely at radii $> 12'$, i.e. we do not use spectra extracted over partial CCDs. We note that the majority of

the data (21/26 OBSIDs, or 77 per cent of the total clean exposure) was taken at roll angles similar to that shown in the left panel of Figure B1, in which readout is approximately along the radial direction.

We fit the usual PHABS∗APEC model to the data identified above, linking the temperature and metallicity across all of the spectra, but allowing a different normalization for each. This results in a temperature of 8.6 ± 0.07 keV, in good agreement with the *Suzaku* measurements, and a metallicity of $0.189 \pm 0.024\, Z_\odot$. Compared with the value of $0.283 \pm 0.018\, Z_\odot$ measured from *Suzaku* data at the same radii, this implies an offset of $\ln(Z^{\mathrm{Suz}}/Z^{\mathrm{Cha}}) = 0.41 \pm 0.14$, which we adopt as a prior in Section 3.3.

If CTI is responsible for the reduced metallicity measured by *Chandra*, we should see a dependence of the measurements on detector position: the metallicity measured from a region adjacent to the readout strip should be higher than that measured from the opposite side of the same CCD. To test this, we fit spectra extracted from the readout sides and the opposite sides of the two CCDs at larger cluster radii in the left panel of Figure B1. Here we use only OBSIDs with roll angles similar to that shown (as noted above, this represents most of the data). At this roll angle, the readout strip for these CCDs lies at the far edge of ACIS-I from the cluster center. The signature of CTI, a higher metallicity measurement on the readout (large radius) side compared with the opposite (small radius) side, runs counter to our expectation that the true metallicity profile be flat (or, if anything, decreasing) at these radii. As shown in the right panel of Figure B1, we do measure a higher metallicity on the readout side than on the opposite side ($0.15 \pm 0.07\, Z_\odot$ compared with $0.04^{+0.04}_{-0.03}\, Z_\odot$), although the uncertainties are relatively large. While we cannot necessarily conclude on this basis that CTI is responsible for the observed *Chandra-Suzaku* metallicity offset, the data are consistent with this hypothesis.

Since the effects of CTI have varied over the course of the *Chandra* mission, and impact the front-illuminated (FI) and back-illuminated (BI) ACIS CCDs differently, metallicities measured from the archival data set could in principle also be used to test this hypothesis. Simple checks along these lines, splitting the spectra used in our archival analysis by detector (ACIS-S vs. ACIS-I) and by epoch of observation, were unable to detect trends within the significant statistical uncertainties. While it is beyond the scope of this work, it remains possible that a more intricate analysis, comparing spectra for the same regions of many multiply observed clusters while accounting for detector position and orientation, may shed more light on the issue. We note, however, that such a study would necessarily be very complex, given the complicated relationship that exists in practice between cluster mass and redshift, the date of observation (low-z, massive clusters preferentially earlier), the CCDs used (BI significantly more common early in the mission), and the placement of the cluster outskirts with respect to the active CCDs. In contrast, the Coma analysis above used exactly the same regions of the same CCDs, pointed at the same cluster gas with similar detector orientations, from observations spanning just over one year; in this case, even nearly 1 Ms of exposure of one of the brightest clusters in the sky lacks the statistical power to clearly distinguish the detector regions studied.
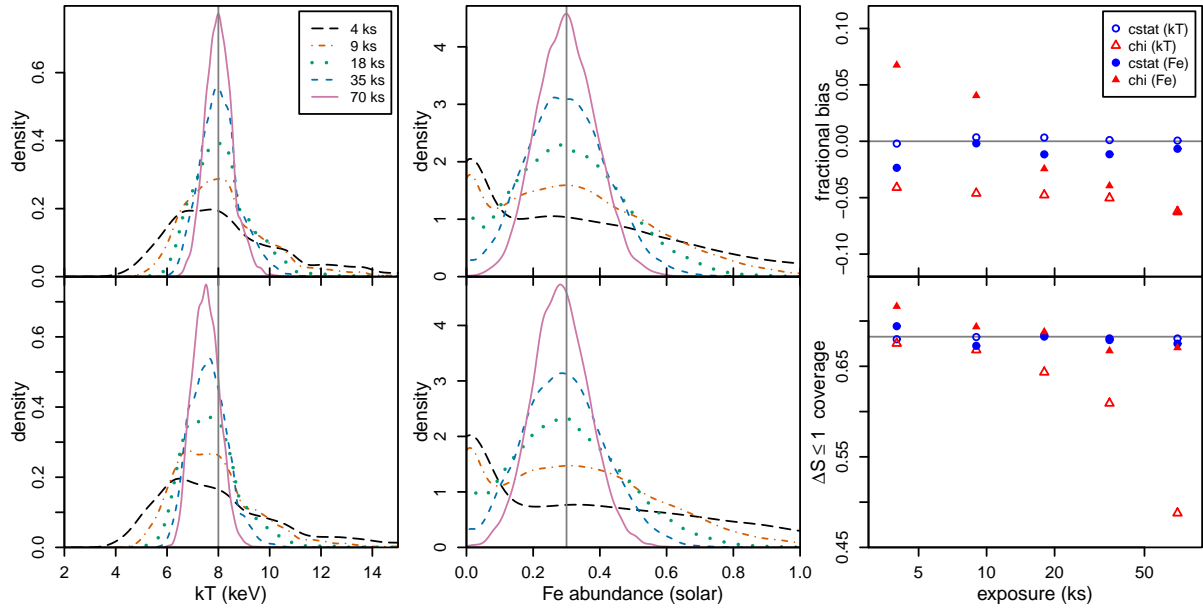
---

[10] We have separately verified that the histograms in the figure agree well with the average Bayesian posterior distribution across realizations, as they should, and similarly that our conclusions about confidence interval coverage obtained from ΔS agree with those derived from Bayesian posteriors. Note that the same priors, most significantly the requirement of non-negative metallicity, apply in both cases.

Figure A1. Left and center columns: smoothed histograms of the best-fitting temperature and metallicity values from many Poisson realizations of a model spectrum, fit using either the *C*-statistic (top row) or χ^2 (bottom row), for various exposure times. The longest exposure of 70 ks corresponds to 6000 source counts, on average. Right column: the fractional bias of the median of each histogram (top), and the corresponding frequentist coverages of the 1σ confidence intervals, defined by ΔC or $\Delta \chi^2 \leqslant 1$, as a function of exposure time.
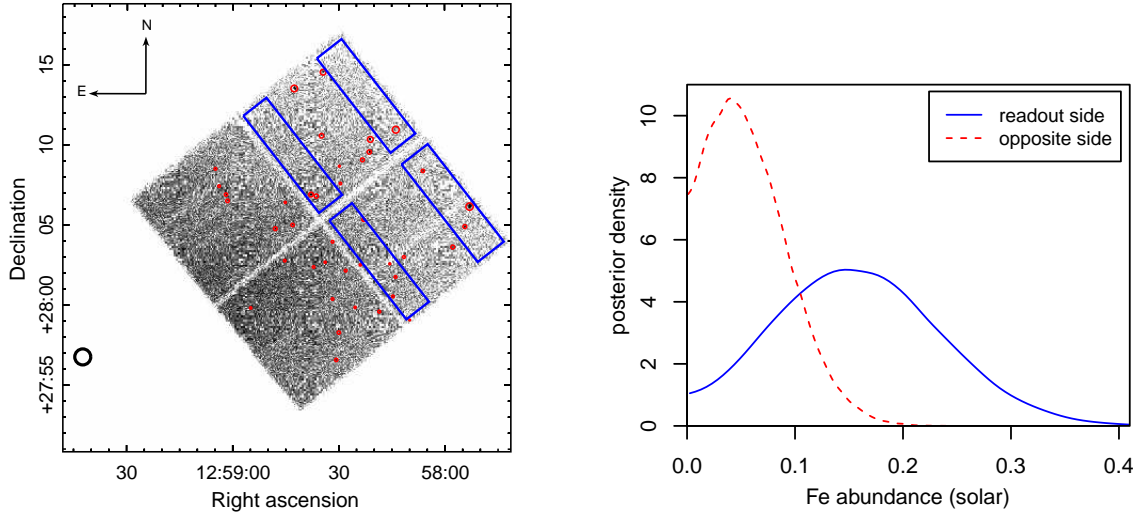

Figure B1. Left: representative *Chandra* event map from the deep Coma data set (OBSID 18271; 53 ks clean exposure). The $1'$ diameter black circle marks the cluster center adopted by **Simionescu et al. (2013)**, while red circles indicate masked point sources. The blue rectangles overlaying the CCDs at larger cluster radii show regions adjacent to (upper right) and opposite (lower left) the readout strip. Within these CCDs, charge is transfered from the lower-left edge (closer to the cluster center) towards the upper-right edge during readout. Right: constraints on the metallicity of Coma from a subset of the *Chandra* data. Blue solid and red dashed curves, respectively, correspond to readout-side and opposite-side regions of the CCDs farthest from the cluster center, as outlined in the left panel. Only observations with roll angles close to that shown in the left panel are used in these fits.