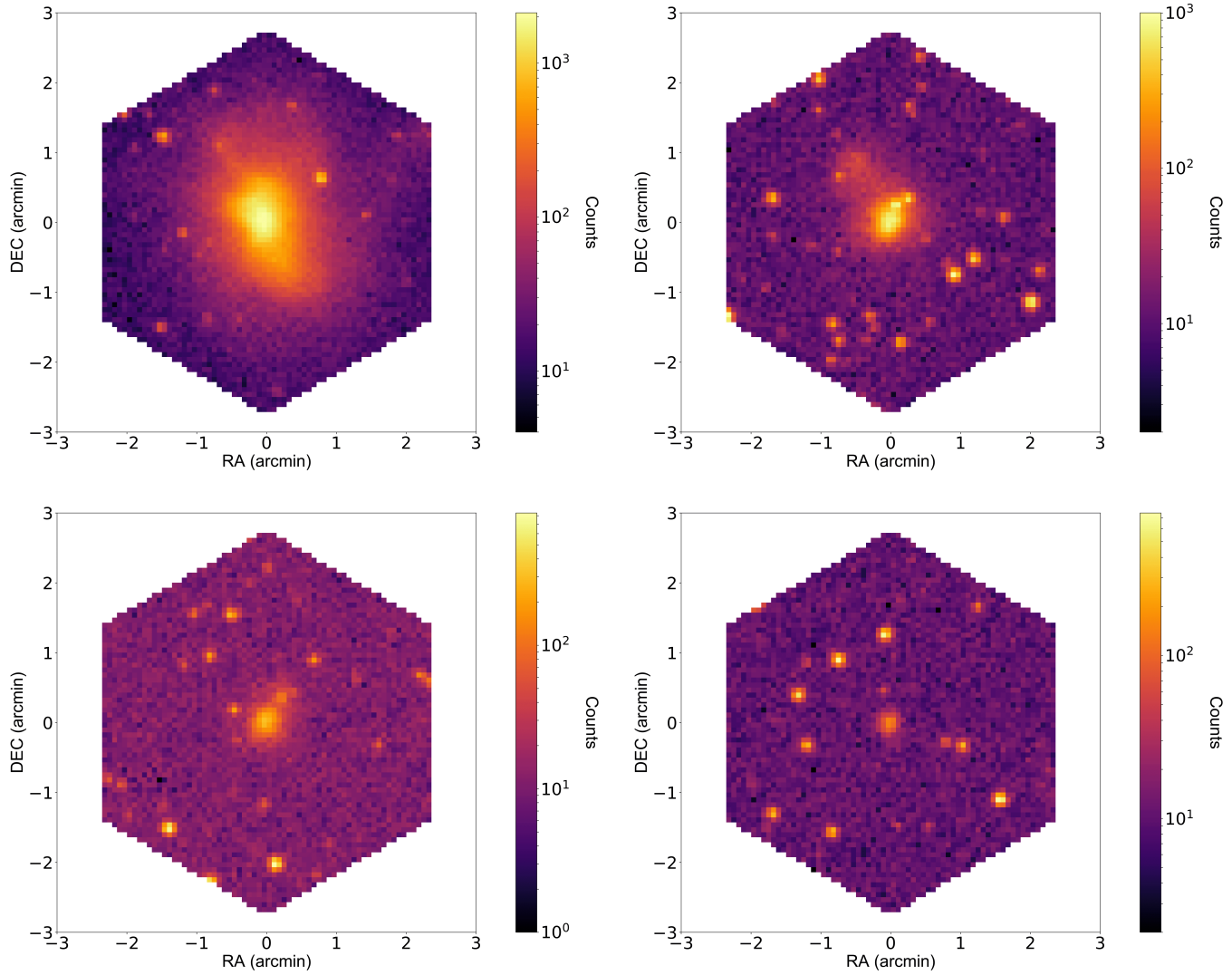

Fig. 7. *From top left to bottom right*: maps in number of counts per X-IFU pixel (249 μm pitch) for cluster C2 (see Table 1) simulated with the end-to-end simulator SIXTE for redshift $z \sim 0.5$ (*top left*), 1 (*top right*), 1.5 (*bottom left*) and 2 (*bottom right*), for an exposure time of 100 ks.

clusters we complied strictly to the definition of the *Athena* science case on chemical abundances and performed measurement in two different annuli from the cluster centre between 0–0.3 R_{500} and 0.3–1 R_{500}.

Figure 7 shows the mock surface brightness maps for C2 at the various redshift snapshots and illustrates the assembly history of halos through, for example, merging events. Figure 8 shows the evolution of the mean cluster abundance over our whole sample as a function of the redshift in the two aforementioned annular regions for O, Mg, Si, and Fe. Despite the lower source-to-background level of some objects, input metallicity values are recovered accurately within the statistical uncertainties of the measurements even for high-redshift clusters, although for $z \geq 1$, error bars start to be significant for elements such as O and Mg. In the case of low-mass elements, abundances are not measurable up to a redshift of $z = 2$, as lines are redshifted outside the instrument energy band (e.g. O and Mg for $z \geq 1.5$) or are too weak to be disentangled from the foreground and the background (e.g. Si at $z = 2$). As expected, measurements in the central parts of the cluster are more accurate, due to the higher level of background in the outskirts with respect to the cluster emission, especially for $z \geq 1.5$.

Through these measurements, we find that in the central parts of the cluster, metallicity hardly changes across time, even at a redshift of $z = 2$, once again consistently with the analysis by Biffi et al. (2017). This indicates that most of the enrichment occurs in the early days of the cluster. Interestingly, we notice that the iron abundance in the centre of the cluster slightly decreases with redshift, which could be explained for instance by an increase in time of iron production mechanisms (e.g. SN_{Ia}) or by the time delay with which long-lived SN_{Ia} release Fe. Abundances in the outskirts show a similar trend, with near-constant values up to local redshift values. Similar observational evidences, as for example, reported in Ettori et al. (2015) for iron abundance are consistent with these conclusions. The dynamic history of the clusters (mergers, shocks) is visible over time (data points are taken from the same cluster at different time steps), displaying local peaks of abundance, for example, C3 at $z = 1$ or C4 at $z = 0.5$. Given the sparse number of redshift points, turbulence or mixing within the cluster (whose eddy turn over timescale is of the order of a few Gyr over scales of ~1 Mpc for typical ~500 km s^{-1} velocities) create a more homogeneous distribution of metals in the structure, returning abundances to typical values after mergers.

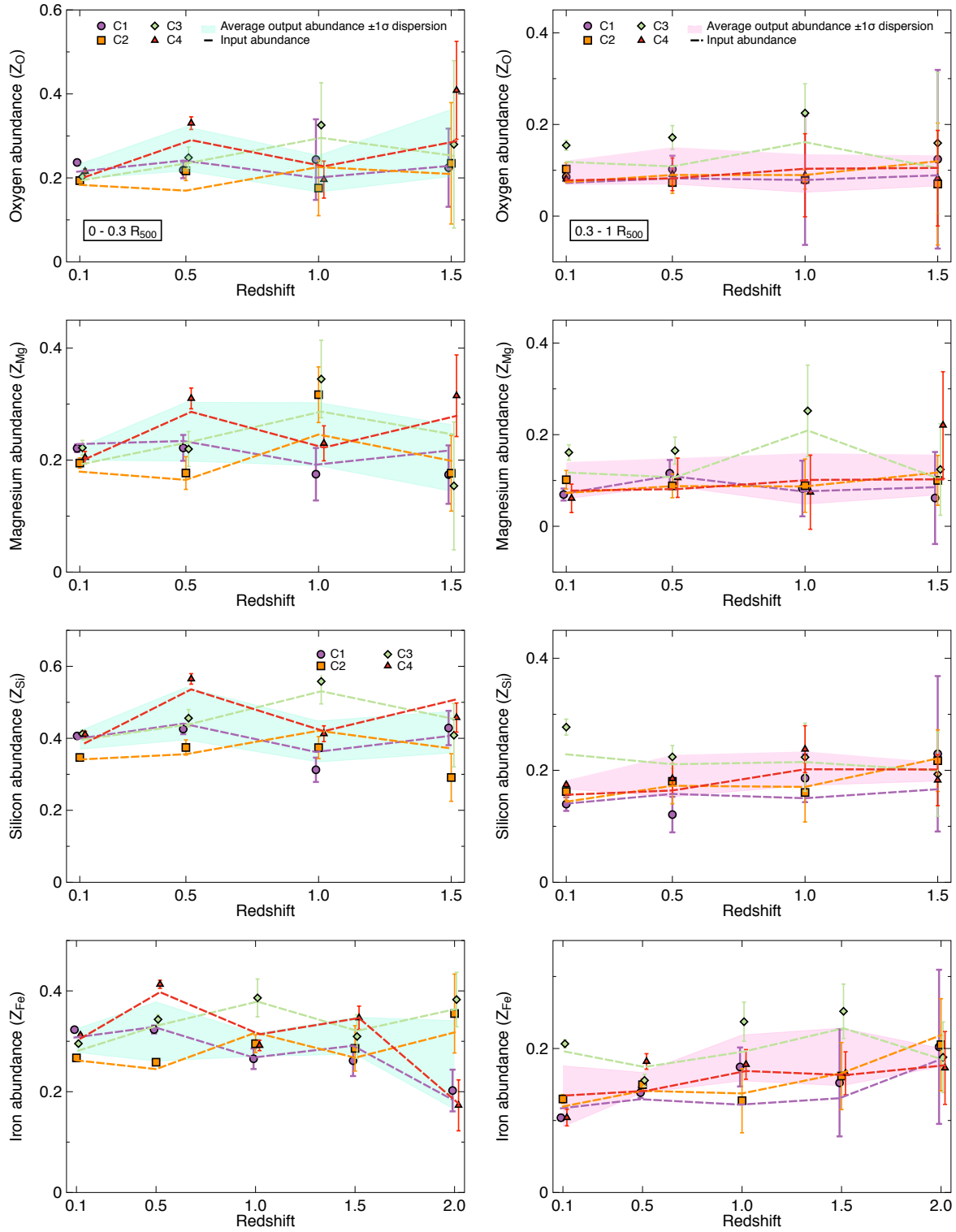

Fig. 8. Evolution of the average abundance of the cluster sample (C1 – purple dots, C2 – orange squares, C3 – green diamonds, C4 – red triangles) recovered via XSPEC as a function of the redshift between 0–0.3 R_{500} (*left*) and 0.3–1 R_{500} (*right*). *From top to bottom*: oxygen, magnesium, silicon and iron abundances with respect to solar. The dashed lines represent the profile of the emission-measure-weighted input abundances using the same colours. The cyan (resp. magenta) shaded envelope represents the $\pm 1\sigma$ dispersion of the output metallicity over the sample. Points are slightly shifted for clarity.

Using the accuracy of the X-IFU abundance measurements for high-redshift objects, we can also analyse changes in the metal enrichment mechanisms by performing a similar study as the one presented in Sect. 5.3. In the case of redshift $z = 1$ (see Fig. 6, right) for 100 ks observations, we find that the X-IFU will still be capable of accurately recovering abundance ratios within R_{500} with excellent accuracy. Most of the main elements have in fact no significant changes between both redshift values, consistently with our previous conclusions. In the case of high-redshift objects however, low-mass elements such as carbon and nitrogen

can no longer be detected (lines outside the energy bandpass) and rarer elements (e.g. Ne, Na, and Al) have large uncertainties due to the low S/N of the observations. Ni also tends to be underestimated (mostly in the outskirts) likely due to the low S/N of the line with respect to the high-energy background. This calls for better-adapted exposure strategies to optimise the results for distant objects and further investigate the chemical enrichment across cosmic time.

7. Summary and discussion

In this paper, we have addressed the feasibility of constraining the chemical enrichment of the Universe, which will be one of the key science objectives and a main driver of the performances of the future mission *Athena*. Notably, we investigated and quantified the capabilities of the X-IFU in accurately recovering metal abundances of the ICM across cosmic time. To this end, we developed a full end-to-end pipeline, which creates synthetic X-IFU observations using the instrument simulator SIXTE. We used as input of this pipeline a sample of four clusters generated using state-of-the-art cosmological simulations presented in Rasia et al. (2015) and Biffi et al. (2017) to create realistic event lists. All the relevant instrumental effects such as the convolution of spectra with the instrument spatial and spectral responses, realistic sources of foreground and background, and detector geometry were also included to obtain observations as realistic as possible.

The sample of four clusters was simulated at five different redshift values, for a fixed exposure time of 100 ks in order to achieve abundance measurements out to R_{500}. The accuracy of the pipeline was quantified by comparing our synthetic observations to weighted inputs quantities (e.g. spectroscopic temperature, emission-measure-weighted abundances). We find that a straightforward approach of a broad-band fit created systematic biases above 10% in a number of physical parameters. Rather a multi-band energy fitting procedure ensured more accurate recovery by optimising the extraction of the several chemical abundances and other physical parameters of interest (notably temperature). After post-processing, distributions are accurately recovered (almost always within the 3σ of the measurement error) with little to no systematic biases (of the order of a 5%, see Sect. 4.6) found mainly between the low-mass element abundances (e.g. O, Si) and the normalisation. The comparison of the relative distribution between outputs and inputs with respect to the XSPEC statistical error also showed reduced chi-squared values $\tilde{\chi}^2_{\rm red}$ close to 1 when a small fraction of outlier regions ($\leq$5%) is removed indicating a good accuracy in the fits. Remaining errors and biases can be linked to correlation between parameters (notably the normalisation), to the choice of the input weighting scheme, and to mixing effects along the line-of-sight in view of the single plasma temperature model used here. We also find that most of these errors decrease when statistics are strongly increased (biases below 2% at 1 Ms for the same spatial regions), suggesting that these effects may simply be related to statistics (see Appendix B). Studies conducted by decreasing the statistics of the runs (typically by decreasing the S/N of the regions to 50 or 100) provided equally encouraging results. Despite larger statistical errors (up to 10% higher), the main parameters (temperature, redshift, O, Si, or Fe abundance) were accurately recovered. Some fainter lines (e.g. Na, N, or Al) become however very difficult to constrain in this case.

For local clusters ($z \sim 0.1$), we demonstrate the power of the X-IFU in accurately recovering spatially resolved parameter maps, along with abundance profiles (Sect. 5.2) and abundance ratios (Sect. 5.3). The study was then extended at different redshift values, up to $z \sim 2$. By probing the chemical enrichment for very distant clusters and despite the lack of an optimised observation strategy (i.e. non-optimised exposure time), we also show the power of the X-IFU in investigating the ICM properties and the chemical enrichment of the distant Universe.

The binning and fitting procedures used here comprise "classical" approaches to X-ray data analysis, using S/N binning and fits through instrumental response matrices in XSPEC. Despite our efforts, the fitting procedures remain slightly biased ($\leq$5%) and small changes in the fitting approach can impact the overall results of the simulation (of the order of a few %). More accurate results may be achieved using, for example, Monte-Carlo (MC) fitting approaches, but unfortunately remain computationally cumbersome to be used on our large set of spatial regions. More optimised binning techniques could also be investigated for future applications (Kaastra & Bleeker 2016). The access to high-resolution spectra will provide new proxies to estimate quantities such as the temperature by using for example, line-ratio techniques. Eventually, hyper-spectral methods (e.g. blind source separation algorithms) or machine-learning-based fitting techniques (see, e.g. Ichinohe et al. 2018) could open new perspectives for the post-processing of high-resolution X-ray spectra. We would like to emphasise that, even though not applicable in our simulation case, the expected level of spectral resolution of the X-IFU will challenge our current knowledge accuracy of the spectral lines (centroid energies and intrinsic widths). This is critical to allow a meaningful interpretation of the results (as demonstrated in Hitomi Collaboration 2018d, for line ratios) and to disentangle fine spectroscopic effects (such as resonant scattering, Hitomi Collaboration 2018b). This emphasises the need for dedicated tools able to process and analyse future X-IFU high-resolution spectroscopy data-cube. In this regard, the *Athena* mission will certainly benefit from the advances expected in processing tools, fitting methods and atomic databases, from the future XRISM mission (Ishisaki et al. 2018).

Not only do these E2E simulations allow us to explore the capabilities of the future X-IFU instrument, but they also give crucial information on the effect of instrumental parameters in science observations. In this study for instance, the spectral shape of all the foreground and background components were assumed to be perfectly known. For the more local and massive clusters however, the field-of-view of the X-IFU will easily be encompassed within the angular extension of R_{500}. Cluster emission-free regions might thus be unavailable for local background calibration. The spectral resolution of the X-IFU will help mitigate this effect, by allowing us to disentangle various components through the characteristics of their spectral energy distribution. The instrument background may also contaminate the observation of faint sources, as the level of precision to which X-IFU is expected to perform requires its accurate and reproducible knowledge in flight. This may be achieved through, for example, in-flight cross-correlation with the WFI or the X-IFU cryogenic anti-coincidence detector (Cucchetti et al. 2018). Future developments could take advantage of this simulation pipeline to test other realistic instrumental effects (e.g. stray-light for galaxy cluster outskirts observations). More detailed studies of the abundance ratios recovered here will also be at the centre of a forthcoming study to highlight the capabilities of the X-IFU in constraining the ICM chemical enrichment, and notably to disentangle between the contributions of the various mechanisms of chemical enrichment (e.g. SN, AGB) throughout cosmic time.

Our study underlines the revolutionary capabilities brought by the X-IFU in future X-ray spectroscopy. With typical routine observations, the X-IFU will drastically change our

understanding of ICM mechanisms and provide a quantum leap forward in X-ray astronomy.

Acknowledgements. V. Biffi, S. Borgani and E. Rasia acknowledge financial contribution from the contract ASI-INAF n.2017-14-H.0. E. Rasia acknowledges the ExaNeSt and Euro Exa projects, funded by the European Union's Horizon 2020 research and innovation programme, under grant agreement No 754337. S. Borgani and L. Tornatore acknowledge support from the EU H2020 Research and Innovation Programme under the ExaNeSt project (Grant Agreement No. 671553). S. Borgani also acknowledge support from the INFN INDARK grant. S. Ettori acknowledges financial contribution from the contracts NARO15 ASI-INAF I/037/12/0, ASI 2015-046-R.0 and ASI-INAF n.2017-14-H.0. M. Gaspari is supported by NASA through Einstein Postdoctoral Fellowship Award Number PF5-160137 issued by the Chandra X-ray Observatory Center, which is operated by the SAO for and on behalf of NASA under contract NAS8-03060. Support for this work was also provided by Chandra grant GO7-18121X. The authors would like to extend the thanks to the anonymous referee for the suggestions and helpful comments.

References

Anders, E., & Grevesse, N. 1989, Geochim. Cosmochim. Acta, 53, 197
Arnaud, K. A. 1996, ASP Conf. Ser., 101, 17
Barret, D., Lam Trong, T., den Herder, J. W., et al. 2016, Space Telescopes and Instrumentation 2016: Ultraviolet to Gamma Ray, Proc. SPIE, 9905, 99052F
Beck, A. M., Murante, G., Arth, A., et al. 2016, MNRAS, 455, 2110
Biffi, V., Dolag, K., & Böhringer, H. 2013, MNRAS, 428, 1395
Biffi, V., Planelles, S., Borgani, S., et al. 2017, MNRAS, 468, 531
Biffi, V., Planelles, S., Borgani, S., et al. 2018, MNRAS, 476, 2689
Bonafede, A., Dolag, K., Stasyszyn, F., Murante, G., & Borgani, S. 2011, MNRAS, 418, 2234
Canizares, C. R., Clark, G. W., Markert, T. H., et al. 1979, ApJ, 234, L33
Canizares, C. R., Clark, G. W., Jernigan, J. G., & Markert, T. H. 1982, ApJ, 262, 33
Canizares, C. R., Davis, J. E., Dewey, D., et al. 2005, PASP, 117, 1144
Cappellari, M., & Copin, Y. 2003, MNRAS, 342, 345
Cash, W. 1979, ApJ, 228, 939
Chabrier, G. 2003, PASP, 115, 763
Churazov, E., Vikhlinin, A., Zhuravleva, I., et al. 2012, MNRAS, 421, 1123
Clerc, N., Ramos-Ceja, M. E., Ridl, J., et al. 2018, A&A, 617, A92
Cucchetti, E., Pointecouteau, E., Barret, D., et al. 2018, Space Telescopes and Instrumentation 2018: Ultraviolet to Gamma Ray, Proc. SPIE, 10699
de Grandi, S., & Molendi, S. 2009, A&A, 508, 565
de Plaa, J. 2013, Astron. Nachr., 334, 416
de Plaa, J., Werner, N., Bleeker, J. A. M., et al. 2007, A&A, 465, 345
den Hartog, R., Kirsch, C., de Vries, C. 2018, J. Low Temp. Phys., 193, 533
den Herder, J. W., Brinkman, A. C., Kahn, S. M., et al. 2001, A&A, 365, L7
Ettori, S., Pratt, G. W., de Plaa, J., et al. 2013, ArXiv e-prints [arXiv:1306.2322]
Ettori, S., Baldi, A., Balestra, I., et al. 2015, A&A, 578, A46
Ezer, C., Bulbul, E., Nihal Ercan, E., et al. 2017, ApJ, 836, 110
Gardini, A., Rasia, E., Mazzotta, P., et al. 2004, MNRAS, 351, 505
Gaspari, M., & Churazov, E. 2013, A&A, 559, A78
Gaspari, M., & Sdowski, A. 2017, ApJ, 837, 149
Gaspari, M., Brighenti, F., D'Ercole, A., & Melioli, C. 2011, MNRAS, 415, 1549
Gaspari, M., McDonald, M., Hamer, S. L., et al. 2018, ApJ, 854, 167
Gilli, R., Comastri, A., Brunetti, G., & Setti, G. 1999, New Astron., 4, 45
Gilli, R., Comastri, A., & Hasinger, G. 2007, A&A, 463, 79
Hasinger, G., Miyaji, T., & Schmidt, M. 2005, A&A, 441, 417
Hillebrandt, W., Kromer, M., Röpke, F. K., & Ruiter, A. J. 2013, Front. Phys., 8, 116
Hitomo Collaboration (Aharonian, F., et al.) 2016, Nature, 535, 117
Hitomi Collaboration (Aharonian, F., et al.) 2017, Nature, 551, 478
Hitomi Collaboration (Aharonian, F., et al.) 2018a, PASJ, 70, 9
Hitomi Collaboration (Aharonian, F., et al.) 2018b, PASJ, 70, 10
Hitomi Collaboration (Aharonian, F., et al.) 2018c, PASJ, 70, 11
Hitomi Collaboration (Aharonian, F., et al.) 2018d, PASJ, 70, 12
Hofmann, F., Sanders, J. S., Nandra, K., Clerc, N., & Gaspari, M. 2016, A&A, 585, A130
Iben, Jr., I., & Renzini, A. 1983, ARA&A, 21, 271
Ichinohe, Y., Yamada, S., Miyazaki, N., & Saito, S. 2018, MNRAS, 475, 4739
Ishisaki, Y., Ezoe, Y., Yamada, S., et al. 2018, J. Low Temp. Phys., submitted
Kaastra, J. S., & Bleeker, J. A. M. 2016, A&A, 587, A151
Kalberla, P. M. W., Burton, W. B., Hartmann, D., et al. 2005, A&A, 440, 775
Karakas, A. I. 2010, MNRAS, 403, 1413
Komatsu, E., Smith, K. M., Dunkley, J., et al. 2011, ApJS, 192, 18
Leccardi, A., Rossetti, M., & Molendi, S. 2010, A&A, 510, A82
Lehmer, B. D., Xue, Y. Q., Brandt, W. N., et al. 2012, ApJ, 752, 46
Lotti, S., Cea, D., Macculi, C., et al. 2014, A&A, 569, A54
Lotti, S., Mineo, T., Jacquey, C., et al. 2017, Exp. Astron., 44, 371
Macculi, C., Argan, A., D'Andrea, M., et al. 2016, Space Telescopes and Instrumentation 2016: Ultraviolet to Gamma Ray, Proc. SPIE, 9905, 99052K
Mantz, A. B., Allen, S. W., Morris, R. G., et al. 2017, MNRAS, 472, 2877
Maoz, D., Mannucci, F., & Nelemans, G. 2014, ARA&A, 52, 107
Mazzotta, P., Rasia, E., Moscardini, L., & Tormen, G. 2004, MNRAS, 354, 10
McCammon, D., Almy, R., Apodaca, E., et al. 2002, ApJ, 576, 188
Mernier, F., de Plaa, J., Pinto, C., et al. 2016a, A&A, 592, A157
Mernier, F., de Plaa, J., Pinto, C., et al. 2016b, A&A, 595, A126
Mernier, F., de Plaa, J., Kaastra, J. S., et al. 2017, A&A, 603, A80
Molendi, S., Eckert, D., De Grandi, S., et al. 2016, A&A, 586, A32
Moretti, A., Campana, S., Lazzati, D., & Tagliaferri, G. 2003, ApJ, 588, 696
Morrison, R., & McCammon, D. 1983, ApJ, 270, 119
Nandra, K., Barret, D., Barcons, X., et al. 2013, ArXiv e-prints [arXiv:1306.2307]
Nomoto, K., Kobayashi, C., & Tominaga, N. 2013, ARA&A, 51, 457
Padovani, P., & Matteucci, F. 1993, ApJ, 416, 26
Pajot, F., Lam Trong, T., den Herder, J. W., Piro, L., & Cappi, M. 2018, J. Low Temp. Phys., submitted
Peille, P., Dauser, T., Kirsch, C., et al. 2018, J. Low Temp. Phys., submitted
Peterson, J. R., & Fabian, A. C. 2006, Phys. Rep., 427, 1
Pointecouteau, E., Reiprich, T. H., Adami, C., et al. 2013, ArXiv e-prints [arXiv:1306.2319]
Rasia, E., Ettori, S., Moscardini, L., et al. 2006, MNRAS, 369, 2013
Rasia, E., Mazzotta, P., Bourdin, H., et al. 2008, ApJ, 674, 728
Rasia, E., Borgani, S., Murante, G., et al. 2015, ApJ, 813, L17
Romano, D., Karakas, A. I., Tosi, M., & Matteucci, F. 2010, A&A, 522, A32
Roncarelli, M., Gaspari, M., Ettori, S., et al. 2018, A&A, 618, A39
Sanders, J. S. 2006, MNRAS, 371, 829
Schmid, C., Smith, R., & Wilms, J. 2013, SIMPUT - A File Format for Simulation Input (HEASARC, Cambridge (MA): Tech. report)
Seta, H., Tashiro, M. S., Ishisaki, Y., et al. 2012, IEEE Trans. Nucl. Sci., 59, 366
Simionescu, A., Werner, N., Mantz, A., Allen, S. W., & Urban, O. 2017, MNRAS, 469, 1476
Simionescu, A., Nakashima, S., Yamaguchi, H., et al. 2018, MNRAS, submitted [arXiv:1806.00932]
Smith, R. K., Brickhouse, N. S., Liedahl, D. A., & Raymond, J. C. 2001, ApJ, 556, L91
Smith, S. J., Adams, J. S., Bandler, S. R., et al. 2016, Space Telescopes and Instrumentation 2016: Ultraviolet to Gamma Ray, Proc. SPIE, 9905, 99052H
Springel, V. 2005, MNRAS, 364, 1105
Springel, V., & Hernquist, L. 2003, MNRAS, 339, 289
Steinborn, L. K., Dolag, K., Hirschmann, M., Prieto, M. A., & Remus, R.-S. 2015, MNRAS, 448, 1504
Takahashi, T., Kokubun, M., Mitsuda, K., et al. 2016, Space Telescopes and Instrumentation 2016: Ultraviolet to Gamma Ray, Proc. SPIE, 9905, 99050U
Thielemann, F. K., Argast, D., Brachwitz, F., et al. 2003, in From Twilight to Highlight: The Physics of Supernovae, eds. W. Hillebrandt, & B. Leibundgut, 331
Tornatore, L., Borgani, S., Matteucci, F., Recchi, S., & Tozzi, P. 2004, MNRAS, 349, L19
Tornatore, L., Borgani, S., Dolag, K., & Matteucci, F. 2007, MNRAS, 382, 1050
Truong, N., Rasia, E., Mazzotta, P., et al. 2018, MNRAS, 474, 4089
Urban, O., Werner, N., Allen, S. W., Simionescu, A., & Mantz, A. 2017, MNRAS, 470, 4583
Verner, D. A., Ferland, G. J., Korista, K. T., & Yakovlev, D. G. 1996, ApJ, 465, 487
Vikhlinin, A., Kravtsov, A., Forman, W., et al. 2006, ApJ, 640, 691
Vogelsberger, M., Marinacci, F., Torrey, P., et al. 2018, MNRAS, 474, 2073
Webbink, R. F. 1984, ApJ, 277, 355
Werner, N., Böhringer, H., Kaastra, J. S., et al. 2007, in Heating versus Cooling in Galaxies and Clusters of Galaxies, eds. H. Böhringer, G. W. Pratt, A. Finoguenov, & P. Schuecker, 309
Werner, N., Durret, F., Ohashi, T., Schindler, S., & Wiersma, R. P. C. 2008, Space Sci. Rev., 134, 337
Werner, N., Urban, O., Simionescu, A., & Allen, S. W. 2013, Nature, 502, 656
Whelan, J., & Iben, Jr., I. 1973, ApJ, 186, 1007
White, S. D. M., Navarro, J. F., Evrard, A. E., & Frenk, C. S. 1993, Nature, 366, 429
Wiersma, R. P. C., Schaye, J., & Smith, B. D. 2009, MNRAS, 393, 99
Willingale, R., Pareschi, G., Christensen, F., et al. 2014, Space Telescopes and Instrumentation 2014: Ultraviolet to Gamma Ray, Proc. SPIE, 9144, 91442E
Wilms, J., Allen, A., & McCray, R. 2000, ApJ, 542, 914
Wilms, J., Brand, T., Barret, D., et al. 2014, Space Telescopes and Instrumentation 2014: Ultraviolet to Gamma Ray, Proc. SPIE, 9144, 91445X
Woosley, S. E., & Weaver, T. A. 1995, ApJS, 101, 181

Appendix A: Spatial binning algorithm comparison

Spatial binning of the data is used to increase the S/N of the considered regions, to have higher significance spectra. This is of particular interest whenever fine structures (e.g. line ratios, line doublets, absorption features) need to be observed in a spectrum. Multiple methods can be used to bin spatially, in this study two of them were considered:

– `contbin` tool developed by J. Sanders (Sanders 2006). The contbin scheme was run for a constrain fill value of two, which represents the maximum ratio length/width of a region.
– Voronoï tessellation, as defined by (Cappellari & Copin 2003)

Both these methods were tested on the same surface brightness map to estimate their performances. Various criteria were compared, such as their ability to reproduce spatial features of the cluster or the mean S/N ratio of the regions created. Figure A.1 shows a comparison of the cluster 2 regions at $z = 0.105$ without any foreground/background component (to investigate purely the binning effects) computed using either of these methods. Visually, we notice that `contbin` provides very similar results to Voronoï whenever the aspect ratio of the region is constrained to $C_{\rm fill} = 1$ (Fig. A.1, right). When a slightly higher value of the aspect ratio is allowed (Fig. A.1, bottom left), we notice that `contbin` is able to reproduce more accurately the radial contours of the cluster, notably the cold arc visible in the south-east corner of input data or the hot bubble rising west of the cluster (Fig. A.1, top left). Further, no difference is found on the average S/N of the regions, which is always above the required level. Finally, the methods give very close results in terms of number of regions (85 for Voronoï vs 87 for `contbin` for a S/N of 300), thus being equivalent computationally. For our purposes, `contbin` tool provides a more suitable binning algorithm than Voronoï. More ample tests also show no significant difference in the recovery of the physical parameters between both techniques.

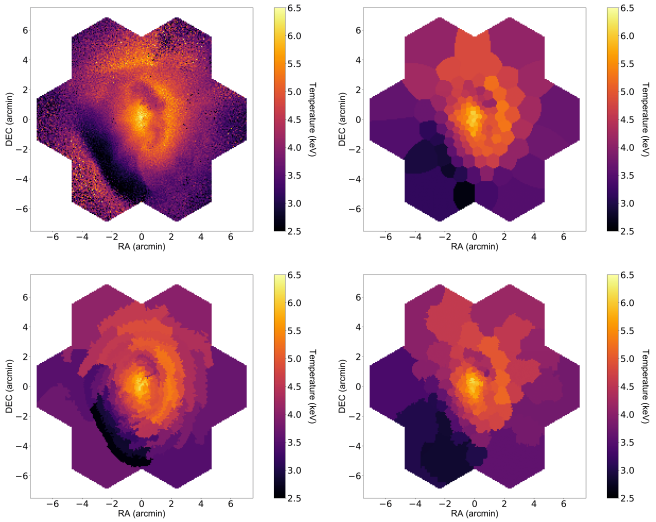

Fig. A.1. Spatial binning scheme comparison for cluster 4 spectroscopic temperature map (keV) without background, at redshift $z = 0.105$. *Top left*: unbinned raw input map from the hydrodynamical simulation. *Top right*: voronoi tessellation map using the algorithms described in (Cappellari & Copin 2003). *Bottom left*: contour map using `contbin` tool (Sanders 2006) with an aspect ratio constraint, $C_{\rm fill} = 2$. *Bottom right*: same as *bottom left*, with $C_{\rm fill} = 1$.

Appendix B: Validation of the simulation pipeline

Test hypothesis

The accuracy of the simulation pipeline needs to be verified using the cluster inputs provided. These inputs are 3D cubes of data, which needs to be projected along the line-of-sight of the instrument to be compared to the outputs of the end-to-end simulations. Ideally, this projection should be deterministic and give an unequivocal results. However, since the parameter distribution along the line-of-sight cannot be perfectly integrated (we only measure discrete number of counts, affected by statistics and background sources), multiple schemes exist to compare inputs and outputs depending on the physical quantity we wish to compare. Among those, the most widely used include:

– Emission-weighted projection, using the product $\rho^2\sqrt{T}$ of each element along the line-of-sight.
– Mass-weighted projection, using the mass of each element
– Emission-measure-weighted projection, using the emission-measure of a given line derived from Eq. (2)
– Spectroscopic schemes, as defined in (Mazzotta et al. 2004)

The accuracy of the simulated measurements was first tested by taking as estimator the relative error distribution of the output map, using emission-weighted-input maps as proxy for the input parameters. Being each region much larger than the telescope PSF they are considered independent on a strict statistical term. The relative error is assumed to follow a Gaussian

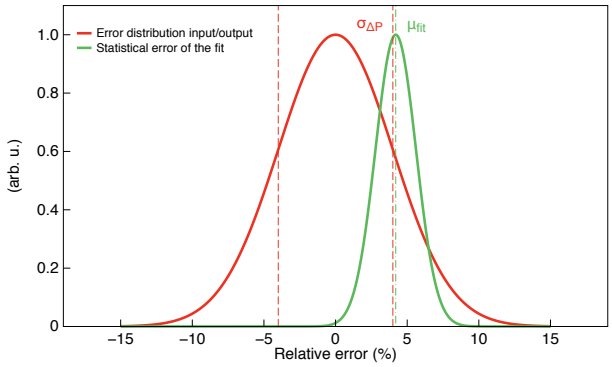

Fig. B.1. Comparison between the relative error distribution of the parameters (red) and the fitting error distribution (green – symmetric but not shown here for clarity). Ideally, if no biases are present relative error distribution should be centred and its standard deviation $\sigma_{\Delta P}$ should be very close to the mean value of the fitting error distribution $\mu_{\rm fit}$.

distribution given the sufficiently high number of regions considered (≥80), with a mean $\mu_{\Delta P} = 0$ (if no biases are present) and a standard deviation of $\sigma_{\Delta P}$, which indicates the total error on the parameter. The fitting error returned by XSPEC should also be Gaussian, and centred on given a value $\mu_{\rm fit}$ which depends on the exposure time and the emission model parameters. For an accurate measurement, the value of $\mu_{\rm fit}$ should be comparable to $\sigma_{\Delta P}$ do all parameter (Fig. B.1). A second test can also be performed using as estimator the ratio between the relative difference and the XSPEC error $\sigma_{\rm fit}$ for each region, i.e. $\chi_j = (P_{{\rm fit},j} - P_{{\rm in},j})/\sigma_{{\rm fit},j}$ and the corresponding reduced chi-square. Using the emission-measure-weighted input and the output distributions, let us take as null hypothesis (H_0): "The measurements obtained with the pipeline are consistent with the statistical errors for a given exposure time" and (H_1): "The measurements are biased" with a threshold $p_\alpha = 5\%$ (i.e. 97.5% of the Gaussian distribution, or $\sim 2.5\sigma$).

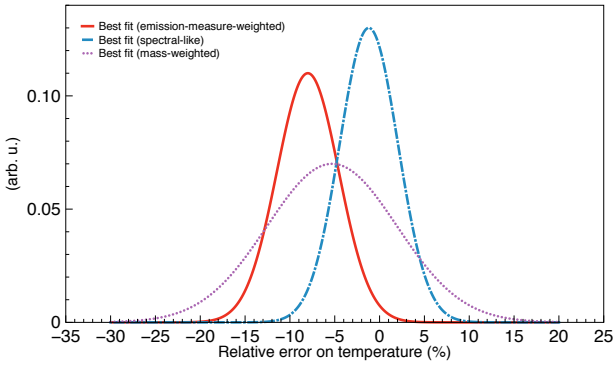

Fig. B.2. Gaussian best fits of the normalised relative error distribution on the measured temperature for different input map weighting scheme for cluster C4. The red solid curve indicates the emission-weighted best fit of the distribution ($\mu_{\Delta T} = -8.0\%$, $\sigma_{\Delta T} = 3.4\%$). The blue dash-dotted line indicates the spectroscopic temperature best fit ($\mu_{\Delta T} = -1.2\%$, $\sigma_{\Delta T} = 2.2\%$), while the dotted violet line indicates the best fit for a mass-weighted input ($\mu_{\Delta T} = -5.2\%$, $\sigma_{\Delta T} = 7.2\%$).

H_0: (H_0) can be rejected if the value of $\sigma_{\Delta P}$ is outside $\mu_{\rm fit} \pm 2.5\sigma_{\rm fit}$ for all the parameters. This can also be seen on the χ_j maps, if the number of regions with $|\chi_j| \geq 2.5$ is high. For all runs, the value of the error dispersion is always within this threshold and the number of outliers regions is small (see also Appendix C and Table 3), indicating that the (H_0) is valid and errors are consistent with the corresponding statistics.

H_1: (H_1) can be rejected if $\mu_{\Delta P} \sim 0$, within ~2.5 times the mean standard error of the distribution (due to the finite size of the sample). The samples are generally composed of $N_{\rm reg} \sim 80$ regions and the standard error on $\mu_{\Delta P}$ is given by $\sigma_{\Delta P}/\sqrt{N_{\rm reg}}$. Under these assumptions, clear biases are visible in the reconstructed emission-weighted and mass-weighted temperature maps, which presented a systematic underestimation of ~5–10% (Fig. B.2). This bias is explained by mixing effects along the line-of-sight and complexity to disentangle multi-temperature plasma with a single plasma model (Mazzotta et al. 2004). It can be reduced using spectroscopic temperature maps (Fig. B.2). The use of the broad-band fit induced in fact many other visible biases, notably between abundances (O, Si, and Fe) and temperature (Fig. B.3, upper panel). The use of multi-band fitting (detailed in Sect. 4.6) significantly reduces these biases (Fig. B.3, lower panel) within the statistical variations of the parameters. Despite this improvement, small correlations are visible between abundances and temperature (~1%) and between the normalisation and all other parameters (~5%). Efforts to reduce this bias were conducted by fixing or releasing various fitting parameters, without significant success. A clear rejection of the null hypothesis (H_1) cannot be performed, although results suggest that small residual biases and correlations remain in the current fitting procedure, mainly on normalisation.

Influence of the projection scheme

Depending on the weighting schemes (mass-weighted or emission-measure-weighted) different error distributions of the same quantities can be obtained. These discrepancies add a further complexity to evaluate any potential bias in the pipeline (Fig. B.2).

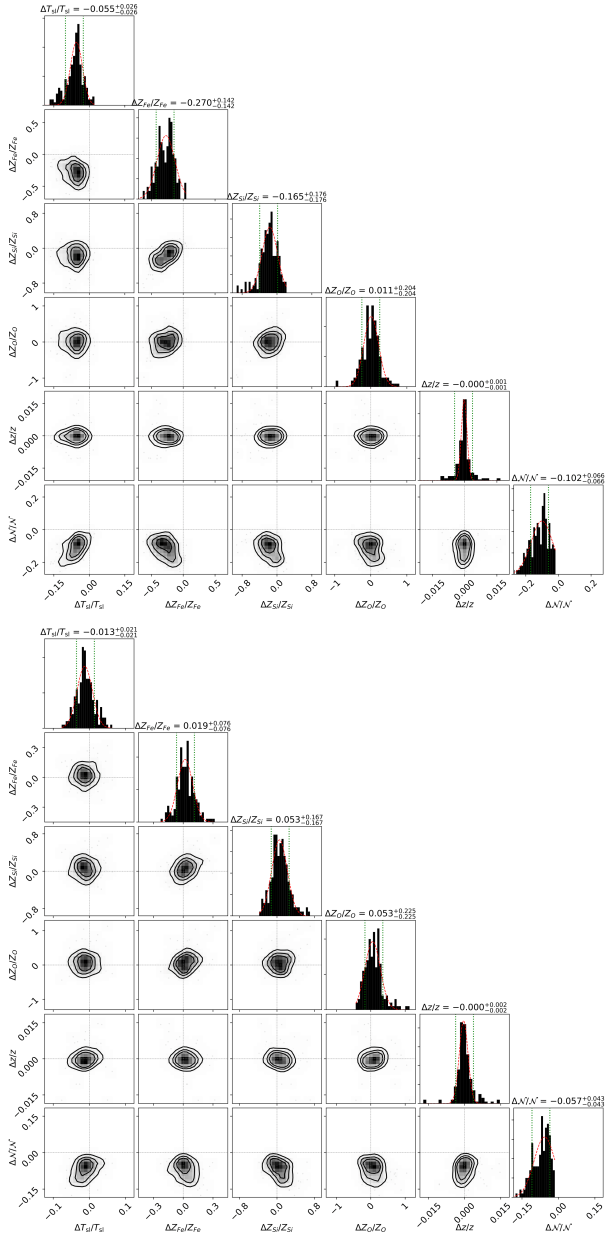

Fig. B.3. Corner plots of the relative error on parameter $(X_{\rm fit} - X_{\rm inp})/X_{\rm inp} = \Delta X/X_{\rm inp}$ as function of parameter $(Y_{\rm fit} - Y_{\rm inp})/Y_{\rm inp} = \Delta Y/Y_{\rm inp}$, for the spectroscopic temperature, $T_{\rm sl}$, oxygen, silicon, and iron abundance, redshift, z, and the normalisation $\mathcal{N}$, for cluster C4. The diagonal panels are the corresponding relative error distribution, where the red solid line indicates the Gaussian best fit of the distribution (parameters $\mu_{\Delta P}$, $\sigma_{\Delta P}$ given above) and the dotted line is the value of $\mu_{\Delta P} \pm \mu_{\rm fit}$. *Top* Broad-band fit. *Bottom*: multi-band fit (Sect. 4.6) considering a spectroscopic temperature.

Accuracy in terms of probability distributions

The previous test is only valid whenever the error distributions are assumed to be Gaussian, which is unfortunately not always the case (slight deviations from Gaussian behaviour are observed). If so, the accuracy of our method needs to be tested in the sense of the statistical distributions performing for instance a Kolmogorov–Smirnov (KS) test over the output and input distributions. The KS test compares the probability $p_{\rm KS}$ for two random variables to be drawn from the same data set (i.e., same

probability density). Using both the input and output distribution, this test showed high $p_{\rm KS}$ values (between 0.6 and 1 for the different parameters), which gives strong hints that the output distribution indeed matches the input. Statistically speaking, no real conclusion can be achieved with a single realization of the observation. To fully validate the pipeline, a large number of observations of the same cluster (either along the same line-of-sight or by taking multiple lines-of-sight) would be needed to perform a meaningful comparison using a KS method. Unfortunately, the duration of one full simulation of a cluster is of the order of a day, making it computationally cumbersome to carry out this test. For simplicity, a very high exposure time simulation of these extended sources were carried out instead. Although beyond the scope of this paper, such observations ($\geq$1 Ms) with the same binning regions decrease most of the biases below 2% and create distributions which much more alike ($p_{\rm KS} \approx 0.8/1$), suggesting that the residual errors are in part related to statistics and to the fitting scheme.

Appendix C: Further results of the simulations

We show here the reconstructed maps of the main physical parameters (spectroscopic temperature, oxygen, silicon and iron) for clusters C1, C3, and C4.

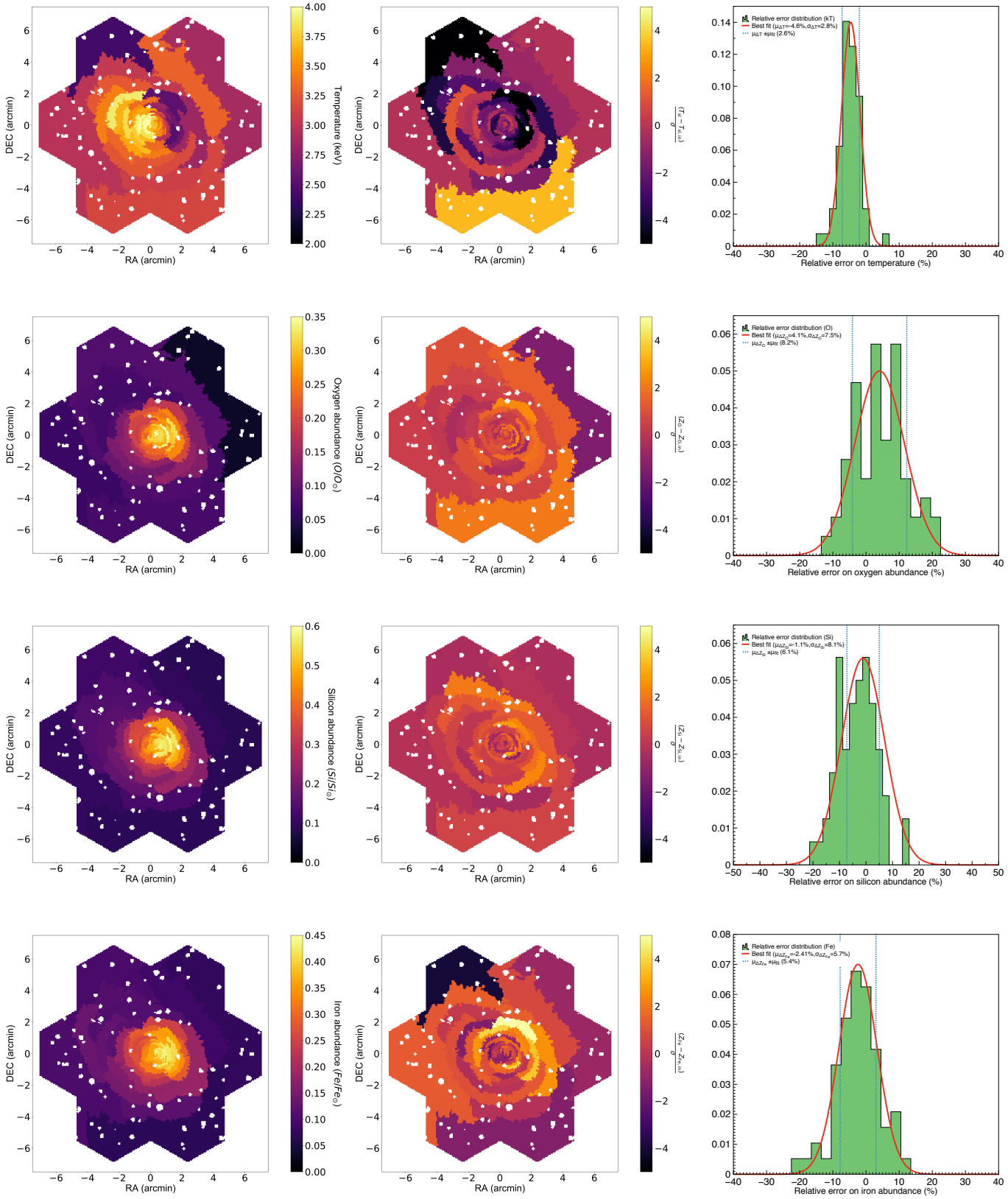

Fig. C.1. As Fig. 3 for cluster C1.

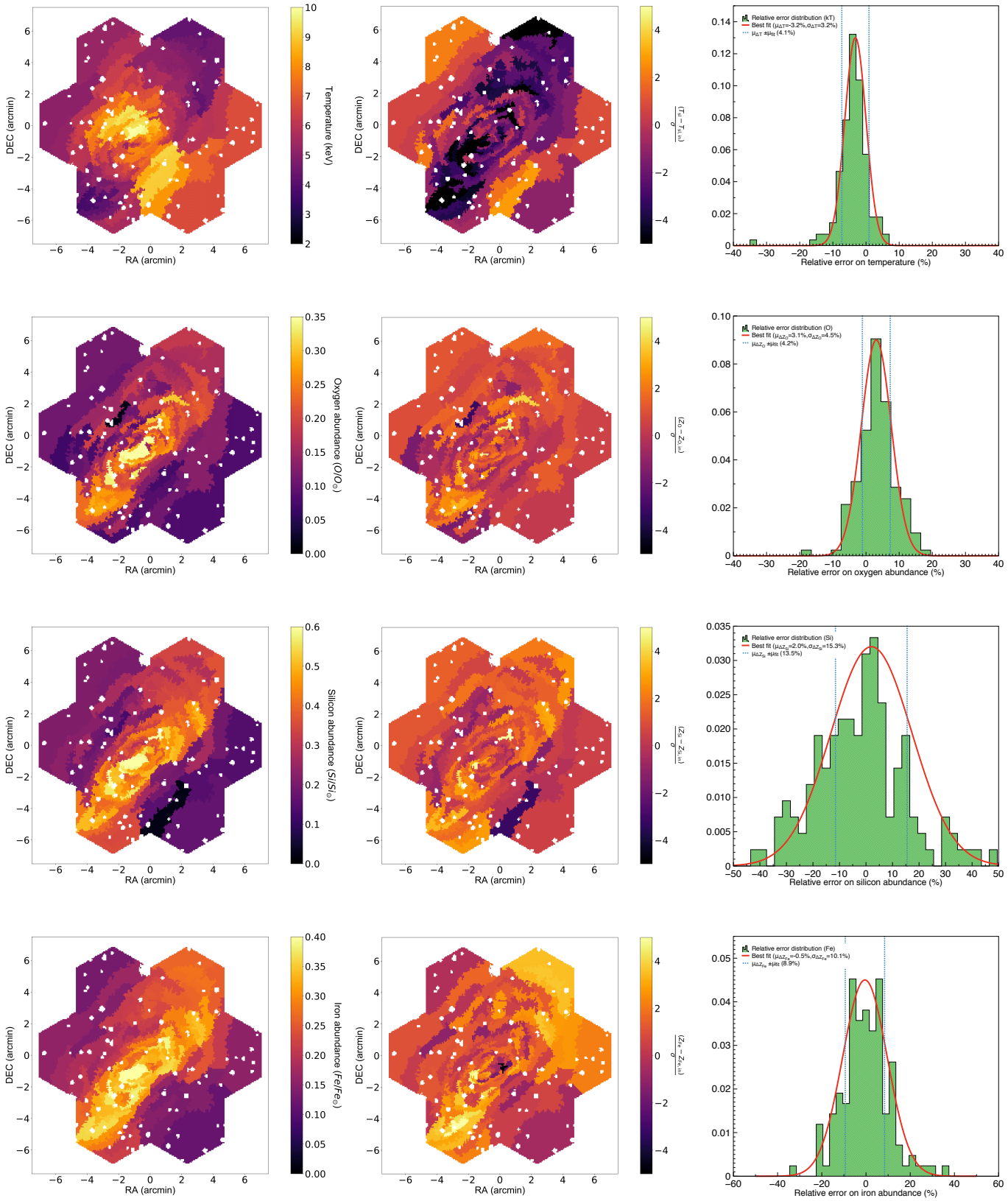

Fig. C.2. As Fig. 3 for cluster C3.