

Temperature map computation for X-ray clusters of galaxies

H. Bourdin¹, J.-L. Sauvageot², E. Slezak¹, A. Bijaoui¹, and R. Teyssier^{2,3}

¹ Observatoire de la Côte d'Azur, BP 4229, 06304 Nice Cedex 4, France

² CEA/DSM/DAPNIA, Service d'Astrophysique, CEA/Saclay, 91191 Gif-sur-Yvette Cedex, France

³ NIC Consortium (Numerical Investigations in Cosmology, CEA Saclay)

Received 16 April 2003 / Accepted 19 September 2003

Abstract. Recent numerical simulations have shown that the variations of the gas temperature in clusters of galaxies are indicative of the dynamical state of these clusters. Maps of the temperature variation show complex structures with different shapes at different spatial scales, such as hot compression regions, filaments, cooling flows, or large-scale temperature profiles. A new multiscale spectro-imagery algorithm for restoring the spatial temperature variations within clusters of galaxies is presented here. It has been especially developed to work with the EPIC MOS1, MOS2 and PN spectro-imagers on board the XMM-Newton satellite. The temperature values are fitted to an emission model that includes the source, the cosmic X-ray background and cosmic-ray induced particle background. The spatial temperature variations are coded at different scales in the wavelet space using the Haar wavelet and denoised by thresholding the wavelet coefficients. Our local temperature estimator behaves asymptotically like an optimal minimum variance bound estimator. But it is highly sensitive to the instrumental and astrophysical backgrounds, so that a good knowledge of each component of the emission model is required. Our algorithm has been applied to a simulated 60 ks observation of a merging cluster at $z = 0.1$. The cluster at different stages of merging has been provided by 3-D hydrodynamical simulations of structure formation (AMR). The multiscale approach has enabled us to restore the faint structures within the core of the merging subgroups where the gas emissivity is high, but also the temperature decrease at large scale in their external regions.

Key words. galaxies: clusters: general – galaxies: interactions

1. Introduction

Within the context of the so-called Λ -Cold Dark Matter scenario for hierarchical formation of structures, clusters of galaxies are the largest structures and consequently the last ones to have been formed. They are supposed to have undergone recently the transition between linear and non-linear regimes. Actually, clusters of galaxies are formed by aggregation of groups and subclusters of different masses along filamentary over-densities, as shown by numerical simulations (Katz & White 1993; Colberg et al. 2000; Evrard 1990; Evrard et al. 2002).

Most of the mass of clusters of galaxies is split into a dark matter component – around 70% of the total mass – and a gas halo component. The gas component is an optically thin plasma heated up to 10^8 K and confined within the gravitational potential at thermal equilibrium, which entails a thermal bremsstrahlung emission detectable in X-rays. Since the cluster accretion process is mainly driven by gravitation, observing the X-ray emission of this gas provides a significant piece of information about the dynamical state of the cluster as a whole. Besides, such studies are only limited by photon statistics since

the intra-cluster gas is an extended source, whereas optical studies using galaxy counts are limited by the galaxy distribution. Recent hydrodynamic simulations (Ritchie & Thomas 2001; Ricker & Sarazin 2001; Roettiger et al. 1996; Teyssier 2002) have shown the sensibility of the gas surface brightness and temperature variations to events related to post and on-going mergers such as gas compressions, cooling flows and cold fronts. But our understanding of the details of the merger scenario, especially the physical processes occurring in the dissipative phase, must strongly be improved by benefitting from new observational data on nearby clusters. The new generation of X-ray telescopes (Chandra; XMM-Newton) now provides data with enough spatial and spectral resolution to try to detect such variations. However, the X-ray photon statistics with its typical emission of a few photons per second in a field of view (FOV) of 30 arcmin in diameter leads to a much lower signal-to-noise ratio than obtained with usual optical observations. Consequently any analysis of X-ray spectroscopic data must involve a spatial binning to increase the photon statistics. To do so, while preserving the best spatial resolution, multiscale algorithms can be applied. We introduce here such a multiscale spectro-imagery algorithm dedicated to the optimal mapping of temperature variations in the observed intra-cluster plasma.

Send offprint requests to: H. Bourdin,
e-mail: bourdin@obs-nice.fr

This algorithm has been especially designed to process data from the XMM-EPIC spectro-imagers, MOS1, MOS2 and PN cameras. Thanks to a maximum likelihood algorithm the temperature of the gas inside each element of a grid covering the field of view is computed with its signal-to-noise ratio. Several computations are done for grids with different spatial resolution following a dyadic scheme, in order to obtain the final temperature map from significant temperature variations at different scales. To do so a wavelet formalism is applied: a selection of significant wavelet coefficients at different scales enables us to reconstruct the optimal temperature map. The spectral analysis is introduced in the first section while the multiscale algorithm is described in the second one. Results from simulations of EPIC-XMM data on merging clusters are discussed in the third section. The last section is devoted to our conclusions.

2. Local temperature computation: Spectral analysis

The EPIC spectro-imagers work in photon count mode. So, the data to be processed are a set of photon events, composed of the photon energy and its location on the detector surface. Given the events within an area of the detector we get the observed emission spectrum of the related region of the sky. By modeling this emission into its different astrophysical and instrumental components, the proper cluster emission can be inferred from this spectrum and the temperature and emissivity of the gas can be derived using an optically thin plasma emissivity model.

2.1. The emission model

The observed emission spectrum has been modelled as a combination of three different emission components. The first of these components is the proper cluster emission, the second one is the cosmic X-ray background (CXB) and the last one is a “particle background” which has been introduced to model the EPIC detector response to induced particles from interactions of cosmic particles with the satellite platform.

The first two components correspond to astrophysical sources. In that case the photons have to follow the telescope optical path before reaching the detector and the emission spectra of these sources must be convolved with the spectral instrumental response. The EPIC XMM-Newton photon counts suffer from a vignetting effect which depends both on coordinates on the detector and on the photon energy. Considering this energy dependence, the vignetting function of each EPIC camera is more or less flat at low energy (0.5–5 keV), but it decreases at high energy (>5 keV), so that the net statistics are poorer at high energy. The spatial vignetting function decreases the photons counts in the focal plane as a function of the distance to the detector optical axis. This vignetting effect has to be taken into account when considering the spectral instrumental response. The convolution of the CXB and cluster emissivity component can be done for each camera using the associated vignetting function provided that the mean spatial location of the analyzed events on the detector surface is known.

The cluster X-ray emission model is modelled as an optically thin plasma emission (Mewe 1985, 1986). The plasma

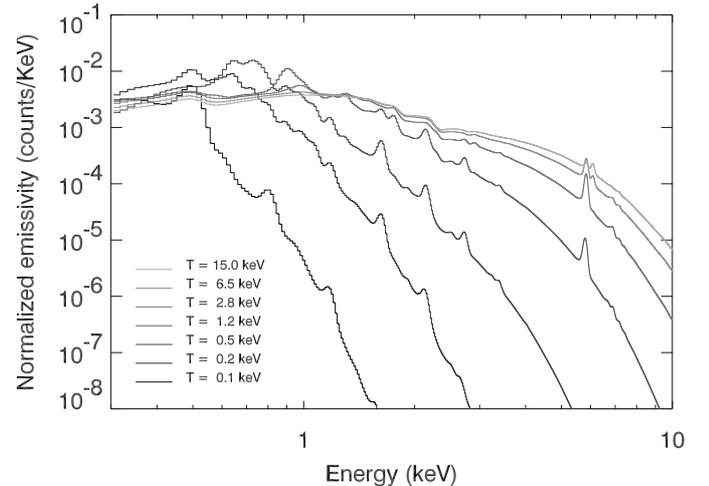


Fig. 1. Normalized emissivity $S(T, e)$ of the optically thin plasma as a function of temperature, ($z = 0.1$, $N_H = 6 \times 10^{20} \text{ cm}^{-2}$). The spectra have been convolved by the EPIC-XMM MOS1 focus instrumental response.

is considered to be a hot ($T > 10^6 \text{ K}$) and low-density one ($n \approx 10^{-3} \text{ cm}^{-3}$) at the coronal limit (Sarazin 1986). Assuming the plasma to be at the ionization equilibrium this emission model is composed of two components. The first component is a continuum of free-free and free-bound radiation emitted during the Coulomb scattering of the electrons by the plasma ions. The second one is a line radiation due to recombination of electrons after the collision process of the different elements of the plasma. Under these assumptions the shape of the emission spectra for the galaxy cluster only depends on the plasma equilibrium temperature T and elements abundances. But the optical path from the source to the detector entails three distortions of this spectral shape when observed: the redshift, the X-ray absorption due to ionization of galactic hydrogen and the EPIC-XMM instrumental response. So, given the element abundances of the plasma, the redshift of the source, the hydrogen column density between the observer and the source and the EPIC-XMM instrumental response, a table of normalized emission spectra $S(T, e)$ can be computed with temperature T and energy e . Such a table is plotted in Fig. 1, where the emission spectra have been convolved with the EPIC-XMM MOS1 focus instrumental response, where no vignetting occurs.

The cosmic X-ray background (CXB) is composed of three components. The first two components correspond to the “soft X-ray background” due to the hot local bubble which can be modelled as an optically thin plasma emission at 0.204 and 0.074 keV respectively. The third component is modelled using a power-law emission with index $\gamma = 1.42$ and corresponds to the hard X-ray background (Lumb et al. 2002). As for the cluster component, the CXB component have to be convolved with the XMM-EPIC instrumental response. Figure 2 shows the three X-ray background spectra convolved with the EPIC-XMM focus instrumental response.

For each camera, the “particle background” can be modelled as a flat energy response plus some fluorescence lines of the detector in the 1.4–2 keV band (Al, Si). Contrary to the CXB the particles background does not correspond to photons

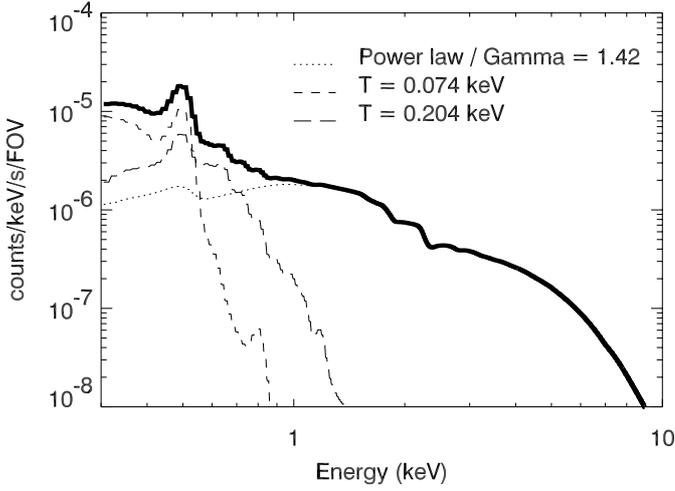


Fig. 2. Three component X-ray background $N_B B(E)$. The three components have been convolved with the EPIC-XMM MOS1 focus instrumental response.

following the telescope optical path but to particles impacting the CCD without interacting with the telescope mirrors. Consequently the particle background is supposed to be uniform on the detector surface and needs not be convolved with the instrumental response.

Let us call $P(e)$ and $B(e)$ the normalized particles and astrophysical background components, respectively. We get an emission spectrum with three components, where the relative contribution of each component, i.e. $S(T, e)$ for the cluster, $B(e)$ for the CXB and $P(e)$ for the particles, can be taken into account by means of three normalization terms, N_S , N_B and N_P , respectively. Let us call $F(T, e)$ the normalized emission model and N_F its normalization term, we get:

$$N_F F(T, e) = N_S S(T, e) + N_B B(e) + N_P P(e), \quad (1)$$

where the normalization terms are linked to each other:

$$N_F = N_S + N_B + N_P. \quad (2)$$

Figure 3 shows the on focus EPIC-XMM MOS1 total emissivity $N_F F(T, e)$ of a simulated isothermal cluster at 5 keV at redshift $z = 0.1$ and for $N_H = 6 \times 10^{20} \text{ cm}^{-2}$, and the related three components, $N_S S(T, e)$, $N_B B(e)$ and $N_P P(e)$.

2.2. The maximum likelihood function

Let us now consider a set of N events of known location on the detector surface corresponding to a simulated or real galaxy cluster observation. Providing that the mean background emissivities N_P and N_B are known, the temperature T and emissivity N_S of the cluster are the unknown parameters to be fitted to the global emission model $F(T, e)$.

The mean background emissivities N_P and N_B are supposed to be known and uniform over the detector surface. Indeed, considering that events in the 10–12 keV energy band are particles events exclusively, the value of N_P can be computed by adding all the events within the 10–12 keV range and extrapolating the bolometric particle emissivity from the particle emission model. The value of N_B can be estimated empirically from

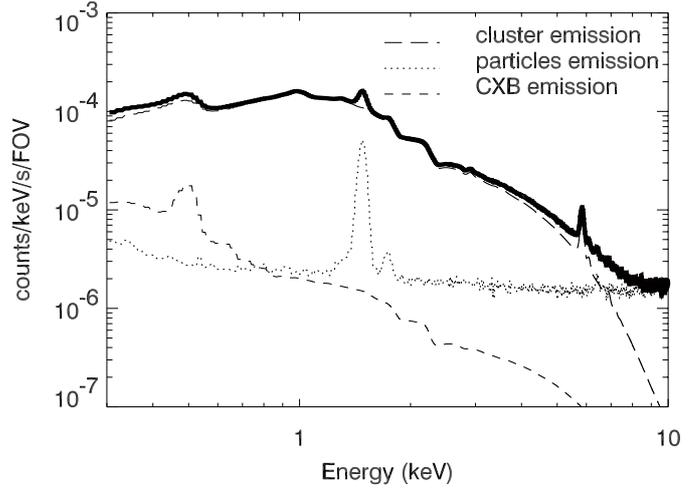


Fig. 3. EPIC-XMM MOS1 total focus emissivity in the line of sight of an isothermal cluster at 5 keV, taking into account the cluster component, but also CXB and particle background.

the emissivity of external regions of the FOV where the CXB emissivity overcomes the cluster emissivity. Introducing the emissivity parameter $r = \frac{N}{N_F}$ to get a bounded parameter space, Eqs. (1) and (2) can be rewritten as:

$$\begin{aligned} F(r, T, e) &= \left[1 - \frac{r}{N} (N_B + N_P) \right] S(T, e) \\ &\quad + \frac{r}{N} [N_B B(e) + N_P P(e)], \\ &= S(T, e) + \frac{r}{N} [N_B (B(e) - S(T, e)) \\ &\quad + N_P (P(e) - S(T, e))]. \end{aligned} \quad (3)$$

$$N_S = N_F - N_B - N_P,$$

$$= N \left[\frac{1}{r} - \frac{1}{N} (N_B + N_P) \right]. \quad (4)$$

So a two parameter model fitting algorithm has to be used to determine simultaneously r and T . A pursuit of the maximum of the likelihood function has been chosen because, contrary to the mean squares estimator, this estimator gives significant results, even in cases of bad statistics. To do so, the emission model is sampled in energy. Each event i corresponds to a photon detection in the energy bin e_i , so that the logarithm of the likelihood function (Kendall & Stuart 1973) can be written as:

$$\ln [L(r, T)] = \sum_i \ln [F(r, T, e_i)]. \quad (5)$$

The maximum of $\ln [L(r, T)]$ gives the values of both parameters T and r . This maximum has to be searched for within a bounded parameter space. Indeed, the emission model $F(r, T)$ only makes sense when both parameters r and T are positive. Furthermore the emission spectra are modelled in a temperature interval corresponding to the intra-cluster plasma usual temperatures, while the emissivity parameter r must be close to 1. Consequently, the parameter space is usually bounded in temperature, from $T_{\min} = 0.1 \text{ keV}$ to $T_{\max} = 30 \text{ keV}$, but also in the emissivity parameter, whose bounds have been set to

$r_{\min} = 0$ and $r_{\max} = 2^1$. In order to pursue the maximum of $\ln[L(r, T)]$ within such a bounded parameter space a modified conjugate gradient algorithm has been chosen for its fast convergence (cf. Appendix A). Eventually, given both values of T and r , the error-bars σ_T and σ_r are obtained from the diagonal elements of the matrix V , deduced from the Fisher matrix F_M :

$$V = -[F_M]^{-1}, \quad \text{with} \quad F_M = \begin{bmatrix} F_1 & \rho \\ \rho & F_2 \end{bmatrix},$$

where:

$$F_1 = \frac{\partial^2 \ln(L)}{\partial r^2}, \quad F_2 = \frac{\partial^2 \ln(L)}{\partial T^2}, \quad \rho = \frac{\partial^2 \ln(L)}{\partial T \partial r}.$$

3. Global detection of temperature variations: A multiscale algorithm

3.1. Introduction

Let us spatially sample the EPIC-XMM FOV along both axes k and l . Given an event-list associated with the pixel k, l , but also an emission model $N_F F(T, e, k, l)$ taking into account the local instrumental response, the maximum likelihood method described in the previous section enables us to define a local estimator $\widehat{T}^{k,l}$ of the parameter $T(k, l)$.

To compute the $\widehat{T}^{k,l}$ value with a high enough signal-to-noise ratio the statistics may have to be enhanced by means of a spatial binning, but this decreases the spatial resolution. Consequently, given the requested signal-to-noise value, an optimal local bin size has to be computed to define $\widehat{T}^{k,l}$, which entails a multiscale exploration of the $\widehat{T}^{k,l}$ signal. A natural multiscale exploration of a given signal $I(k, l)$ is provided by the wavelet transform $W(a, k, l)$ (cf. Appendix B), where the signal $I(k, l)$ is unfolded in the space-scale domain, with a as the scale parameter. Let us define $W_{\widehat{T}^{k,l}}(a)$ the wavelet transform of $\widehat{T}^{k,l}$. The aim of our algorithm is to show the significant $\widehat{T}^{k,l}$ variations at different scales thanks to a thresholding of the wavelet space.

The chosen wavelet transform is based on the Haar wavelet as will be discussed in Sect. 3.2. As a first step, successive binnings of the EPIC-XMM data set at each scale and fitting of the emission model $N_F F(T, e, k, l)$ of Eq. (2) with the maximum likelihood algorithm enable to define successive approximations, $A_{\widehat{T}^{k,l}}(a, k, l)$, of the $\widehat{T}^{k,l}$ distribution. Then the ‘‘first guess’’ set of wavelet coefficients $W_{\widehat{T}^{k,l}}^{(0)}(a, k, l)$ is computed by applying the high-pass analysis filters corresponding to the Haar wavelet to $A_{\widehat{T}^{k,l}}(a, k, l)$. Notice that this set of coefficients does not constitute a real wavelet transform since the coefficients are computed by means of different emission models from scale to

¹ Considering an emission model with a given emissivity N_F , the number of observed counts N is expected to fluctuate like the standard deviation of the Poisson parameter N_F : $\sigma_{N_F} = \sqrt{N_F}$, so that the $r = \frac{N}{N_F}$ values are expected to fluctuate like $\sigma_r = \frac{\sqrt{N_F}}{N_F}$. Consequently the expected r values can be bounded within the $[-\Delta_r, +\Delta_r]$ interval, provided that $\Delta_r > \sigma_r$. The Δ_r value of 1, corresponding to σ_r for one photon per pixel was chosen. Indeed, σ_r decreases with increasing statistics, – for instance $\sigma_r = 0.32$ for $N_F = 10$ –, so that we always get $\sigma_r < \Delta_r$.

scale. Nevertheless, it may be considered as an approximation of the real wavelet transform $W_{\widehat{T}^{k,l}}(a, k, l)$ that we are looking for. Within the wavelet space the respective noise and signal contribution to the signal can be easily separated at each scale. This will be described for the chosen transform in Sect. 3.3. The last step, developed in Sect. 3.4, consists of removing iteratively the noise contribution to $W_{\widehat{T}^{k,l}}^{(0)}$, which leads to a new set of denoised wavelet coefficients $W_{\widehat{T}^{k,l}}^{(n)}(a)$. The final set of coefficients will consist of the expected $W_{\widehat{T}^{k,l}}(a)$ wavelet transform. This process is performed from the lowest to the highest resolution allowing us to construct in the direct space better and better approximations of $\widehat{T}^{k,l}$ and to obtain the optimal temperature map $\widehat{T}(k, l)$.

3.2. Computing the $W_{\widehat{T}^{k,l}}^{(0)}(a, k, l)$ set of coefficients

So as to compute the set of coefficients $W_{\widehat{T}^{k,l}}^{(0)}(a, k, l)$ a wavelet function has to be chosen. The 2-D Haar wavelet (Haar 1910) which matches the square spatial sampling is the most suitable since a square binning has been chosen. It is a compact support square kernel with coefficients $\frac{1}{2}$ and $-\frac{1}{2}$, following the amplitude normalization condition of the wavelet coefficients. The Haar wavelet is anisotropic, so that the wavelet transform results at each scale $a = 2^j$ with $j = 0, \dots, N$ in a decomposition onto three detail spaces with wavelet coefficients $W_p(a, k, l)$ where $p = h, v, d$ indicates the horizontal, vertical and diagonal details, respectively. Their value can be computed from the convolution of the approximation at scale $\frac{a}{2}$, $A(\frac{a}{2}, k, l)$, using the high-pass analysis filter $g_p(\frac{a}{2})$ (cf. Appendix B).

Let us associate with each pixel k, l of the FOV the set of events found within a square bin of size $\frac{a}{2}$. The fitting of this set of events to the model $N_F F(T, e, \frac{a}{2}, k, l)^2$ will enable us to compute the set of coefficients $W_{\widehat{T}^{k,l}}^{(0)}(a, k, l)^3$. To do so, the approximations $A_{\widehat{T}^{k,l}}^{(0)}(\frac{a}{2}, k, l)$ with their rms values $\sigma_{A_{\widehat{T}^{k,l}}^{(0)}}(\frac{a}{2}, k, l)$ are first computed thanks to the maximum likelihood fitting of the model $N_F F(T, e, \frac{a}{2}, k, l)$ to the data corresponding to pixel k, l with scale $\frac{a}{2}$. Then the wavelet coefficients $W_{\widehat{T}^{k,l}}^{(0)}(a, k, l)$ are computed by applying the high-pass analysis filters $g_p(\frac{a}{2})$ (cf.

Fig. B.2). The variances $[\sigma_{W_{\widehat{T}^{k,l}}^{(0)}}(a, k, l)]^2$ of the wavelet coefficients $W_{\widehat{T}^{k,l}}^{(0)}(a, k, l)$ are computed by applying the low-pass analysis filter $h(\frac{a}{2})$ to the variances of the smoothed plane $[\sigma_{A_{\widehat{T}^{k,l}}^{(0)}}(\frac{a}{2}, k, l)]^2$.

The binning of the data has been chosen such that the smoothed planes $A_{\widehat{T}^{k,l}}^{(0)}(a, k, l)$ have the same number of pixels whatever the scale is, yielding a redundant wavelet transform. As it will be explained in the following section the redundancy

² The scale parameter a has now been added because $N_F F(T, e, \frac{a}{2}, k, l)$ is in fact the weighted mean of emission models corresponding to each event found within the square bin of edge a .

³ The p index is hereafter omitted because the following processes do not depend on the orientation of the coefficients.

has been chosen in order to balance the usual poor statistics of the data.

3.3. The noise contribution: The wavelet coefficients thresholding

We get a set of coefficients $W_{\widehat{T}^{k,l}}^{(0)}(a, k, l)$ with zero mean value and variances $\sigma_{W_{\widehat{T}^{k,l}}^{(0)}}(a, k, l)$. Given the signal-to-noise ratio R , the coefficients $W_{\widehat{T}^{k,l}}^{(0),s}(a, k, l)$ with value overcoming R times the local variance will be considered as significant whereas the other ones $W_{\widehat{T}^{k,l}}^{(0),ns}(a, k, l)$ will be considered as non-significant:

$$\begin{aligned} |W_{\widehat{T}^{k,l}}^{(0),s}(a, k, l)| &> R \sigma_{W_{\widehat{T}^{k,l}}^{(0)}}(a, k, l), \\ |W_{\widehat{T}^{k,l}}^{(0),ns}(a, k, l)| &< R \sigma_{W_{\widehat{T}^{k,l}}^{(0)}}(a, k, l). \end{aligned} \quad (6)$$

Let us attribute the Boolean true value to regions of the wavelet space where the $W_{\widehat{T}^{k,l}}^{(0)}(a, k, l)$ coefficients have been considered as statistically meaningful, and conversely the Boolean false value to regions where they have been considered as non-meaningful. Since the set of coefficients $W_{\widehat{T}^{k,l}}^{(0)}(a, k, l)$ is an approximation of the expected wavelet transform $W_{\widehat{T}^{k,l}}(a, k, l)$, the projection of this Boolean space onto the successive wavelet planes at different scales defines an approximation of the so-called significant masks $M(a, k, l)$ of the genuine wavelet transform $W_{\widehat{T}^{k,l}}(a, k, l)$.

Notice that due to the redundancy of the algorithm, the wavelet coefficients $W_{\widehat{T}^{k,l}}^{(0)}(a, k, l)$ are computed from dependent data. A crude estimation of dependency of the coefficients of a wavelet plane is that they are linked to each other, providing that they are included within a region of area a^2 . This information has been used to remove some false detections of significant regions. The connected significant regions of $M(a, k, l)$ are kept providing that their area overcomes a^2 , but are considered non-significant otherwise.

3.4. Constructing the $W_{\widehat{T}^{k,l}}(a, k, l)$ wavelet transform and temperature map $\widehat{T}^{k,l}(k, l)$

This splitting of the wavelet space enables us to construct a true wavelet transform $W_{\widehat{T}^{k,l}}(a, k, l)$ from $W_{\widehat{T}^{k,l}}^{(0)}(a, k, l)$. Since all the available information about the variations of parameter $\widehat{T}^{k,l}$ is considered as being contained within the significant domain of the wavelet space, the wavelet transform $W_{\widehat{T}^{k,l}}(a, k, l)$ can be constructed by keeping the coefficient values of the significant domain of $W_{\widehat{T}^{k,l}}^{(0)}(a, k, l)$. Conversely the coefficient values of the non-significant domain, where information about the variations of $A_{\widehat{T}^{k,l}}^{(0)}(\frac{a}{2}, k, l)$ is considered as lost, will be iteratively modified. Since no variation of $\widehat{T}^{k,l}(k, l)$ has been detected within this domain, a priori information will be added in place of the lack of information: it will be supposed that the $\widehat{T}^{k,l}$ variations are minimal. This condition leads to the Tikhonov smoothness constraint or regularization, where the gradient of the smoothed

planes defined as $\|\frac{\partial A}{\partial x}\|^2 + \|\frac{\partial A}{\partial y}\|^2$ have to be minimized, which implies the following condition (cf. Bobichon & Bijaoui 1997):

$$L_2 A_{\widehat{T}^{k,l}}^{(n)}\left(\frac{a}{2}, k, l\right) = 0, \quad (7)$$

where L_2 is the 2-D Laplacian operator.

Let us call $w_{A_{\widehat{T}^{k,l}}^{(n+1)}}(a, k, l)$ and $w_{L_2 A_{\widehat{T}^{k,l}}^{(n)}}(a, k, l)$ the wavelet coefficients computed by applying the high pass analysis filters $g_p(\frac{a}{2})$ to the respective approximations, $A_{\widehat{T}^{k,l}}^{(n+1)}(\frac{a}{2}, k, l)$ and $A_{\widehat{T}^{k,l}}^{(n)}(\frac{a}{2}, k, l)$, Eq. (7) is solved thanks to an iterative Van-Cittert algorithm on the wavelet coefficients constructed by successive filtering and reconstruction of the smoothed planes $A_{\widehat{T}^{k,l}}^{(n)}(\frac{a}{2}, k, l)$ at scale $\frac{a}{2}$:

$$w_{A_{\widehat{T}^{k,l}}^{(n+1)}}(a, k, l) = w_{A_{\widehat{T}^{k,l}}^{(n)}}(a, k, l) - \beta w_{L_2 A_{\widehat{T}^{k,l}}^{(n)}}(a, k, l). \quad (8)$$

Since $w_{A_{\widehat{T}^{k,l}}^{(n+1)}}(a, k, l)$ and $w_{L_2 A_{\widehat{T}^{k,l}}^{(n)}}(a, k, l)$ correspond to the wavelet coefficients of the temperature map $\widehat{T}^{k,l}(k, l)$ at scale a , Eq. (8) can be rewritten as:

$$W_{\widehat{T}^{k,l}}^{(n+1)}(a, k, l) = W_{\widehat{T}^{k,l}}^{(n)}(a, k, l) - \beta W_{L_2 A_{\widehat{T}^{k,l}}^{(n)}}(a, k, l), \quad (9)$$

where β is a fixed convergence parameter. An additional condition prevents the non-significant coefficient values from overcoming the significance threshold. The convergence criterion $\|W^{(n)} - W^{(n+1)}\| < \varepsilon$, is reached in a few iterations with $\varepsilon = 0.01$. This construction of the wavelet transform $W_{\widehat{T}^{k,l}}(a, k, l)$ is done from the coarsest resolution to the finest one while constructing successive smoothed planes $A_{\widehat{T}^{k,l}}^{(n)}(\frac{a}{2}, k, l)$. The final smoothed plane at the finest resolution defines the final temperature map $\widehat{T}^{k,l}(k, l)$ we are looking for. Notice that the final wavelet transform $W_{\widehat{T}^{k,l}}(a, k, l)$ is the shift invariant wavelet transform introduced by Coifman & Donoho (1995).

3.5. Error map computation

We describe the coarse method we adopted to estimate the final error on the temperature map $\widehat{T}^{k,l}(k, l)$. The temperature map $\widehat{T}^{k,l}(k, l)$ can be considered in the direct space as a linear combination of the different wavelet coefficients $W_{\widehat{T}^{k,l}}^{(n)}(a, k, l)$. For each pixel k, l , the $\widehat{T}^{k,l}(k, l)$ standard deviation is mainly due to the wavelet plane $W_{\widehat{T}^{k,l}}^{(n)}(a, k, l)$ standard deviation, where the scale a corresponds to the highest resolution where $W_{\widehat{T}^{k,l}}^{(n)}(a, k, l)$ is found significant. Consequently, we decided to estimate the error-bar $\sigma_{\widehat{T}^{k,l}}(k, l)$ from the $W_{\widehat{T}^{k,l}}^{(n)}(a, k, l)$ standard deviation. Since $W_{\widehat{T}^{k,l}}^{(n)}(a, k, l)$ is included within the significant domain, the $W_{\widehat{T}^{k,l}}^{(n)}(a, k, l)$ standard deviation is close to the $W_{\widehat{T}^{k,l}}^{(0)}(a, k, l)$ standard deviation value, $\sigma_{W_{\widehat{T}^{k,l}}^{(0)}}(a, k, l)$. We get at pixel k, l ,

$$\sigma_{\widehat{T}^{k,l}}(k, l) = \sqrt{\sum_p \left[\sigma_{W_{\widehat{T}^{k,l}}^{(0)}}(a, k, l) \right]^2}, \quad \text{where } a \text{ is the highest resolution of structure detection.}$$

4. Algorithm performances against simulations

The performance and limits of the algorithm have been tested by means of a set of simulated observations with different

values of source temperatures and emissivities, as well as background emissivities. The spectral fitting performance will be introduced in a first section with uniform source temperatures and emissivities. Then the multiscale algorithm will be tested in a second section on model EPIC-XMM observations with non-uniform temperatures and emissivities but with only one gas emission component $F(k, l, e)$.

4.1. Spectral fitting

4.1.1. Introduction

The spectral fitting performance of our algorithm has been tested against different values of the parameters of the emission model – cf. Eq. (3) –, taking or not taking into account instrumental components. To do so, a cube $F(k, l, e)$ is constructed for different values of $T(k, l)$ and $N_S(k, l)$, the temperature and emissivity of the source, respectively, as well as for N_B and N_P , the background emissivities. The chosen simulated camera is the EPIC-XMM MOS1. An observation can be simulated by drawing randomly within the Poisson distribution of parameter $F(k, l, e)$ a value for the number of photons at position k, l, e . The models used hereafter are 128×128 square images where the source temperature and emissivity are considered uniform, and fixed at 3 keV and 100 photons per pixel, respectively. The fitting is first tested without any background emissivity and then with different particles and CXB background emissivities values. In each case the vignetting influence was also tested by constructing emission models with or without vignetting. Just like for the first step of our multiscale algorithm, the fitting of $F(k, l, a)$ is performed at different resolutions a , leading to different maps $T(k, l, a) \equiv A_{T,k,l}^{(0)}(a, k, l)$ with $a = 2^j$, $j = 0 \dots N$. Actually the fitting has been performed on four successive resolutions with two emission models at 100 and 200 photons per pixel respectively, leading to eight maps with statistics increasing by a factor of two, from 200 to 25 600 cluster photons per pixel. The first spectral fitting test is performed without any background. The mean, standard deviations and a selection of six histograms of the different maps $T(k, l, a)$ corresponding to this spectral fitting are plotted in Fig. 4. An instrumental response with vignetting factor was introduced for the first histogram, but without vignetting for the second one. The histograms of Figs. 5 and 6 correspond to emission models with particles and CXB background, respectively, where each background emissivity values have been set to 0.5 and 2 times the value of the cluster emissivity.

4.1.2. No background vignettted emission model

Let us first consider the fitting of the isothermal emission model $F(k, l, e)$ at 3 keV with a null background emissivity (cf. Fig. 4). The temperature distributions obtained from fittings with more than 1600 cluster photons per pixel are symmetric around a peak value at 3 keV.

The maximum likelihood method used for the model fitting should provide a consistent estimator whose variance is close to the Rao-Cramer limit (Kendall & Stuart 1973) of the Minimum Variance Bound (MVB) estimators. We have checked that the

variances of the temperature histograms are close to the mean of the MVB variances values computed from the Fisher matrix after each spectral fitting. They follow the statistics with N counts, so that $\sigma \propto \frac{1}{\sqrt{N}}$. The variances of distributions obtained from fittings with more than 400 cluster photons per pixel are in agreement with the MVB variance, but the variance of the distribution obtained with 200 cluster photons per pixel is not.

The dispersion obtained with 200 cluster photons is asymmetric and does not peak at the 3 keV value, so that a bias is evidenced. This “low temperature bias” is due to the emissivity of the high-energy part of the source emission spectra, with a variation of several orders of magnitude between the high and low temperature emission models (cf. Fig. 1), and a high relative chance fluctuation of the high-energy photons number. It has been reduced by an histogram equalization of the energy distribution of the photons. It is not worth fully removing this bias since it is only important in the case of very poor statistics when the temperature dispersion is high anyway. Instead of removing it, two conditions have been added to the spectral fitting. The result of the temperature fit of the different resolution elements is only taken into account providing that on the one hand, the 1σ error bars values are lower than 1 keV and on the other hand, more than about 500 photons are available so that the statistics is considered high enough.

4.1.3. No background non-vignettted emission model

Now let us recall that due to the spatial variations of the detector response the different temperatures values are not computed following the same process which is questionable. To get rid of this drawback, but also to test the vignetting influence on the spectral fitting, a model with the same source component as the previous one but without vignetting has been built (cf. Fig. 4). The detector response is now uniform and similar to the focus response all over the detector surface. We get a close to MVB estimator at each resolution with lower temperature dispersions than the dispersions obtained when the vignetting effect occurs. Indeed, the removal of the vignetting effect results in providing much higher energy photons, which enhances the statistics and reduces the “low temperature bias”.

4.1.4. Vignettted emission model with particles background

The particles background influence on the spectral fitting is tested thanks to a model series with vignetting and the same source temperature and emissivities as in Sect. 4.1.1 but with particles background emissivities of 0.5 and 2 times the cluster emissivity (cf. Fig. 5). The temperature variance is larger than in the case of null background emissivity and all the larger since the background emissivity increases. The fitted source temperatures are found to be still distributed around the $T = 3$ keV value in the case of a cluster emissivity higher than 3200 photons per pixels, but biased in the case of lower statistics. The “low-temperature bias” is enhanced here since the source high-energy photons are much less numerous than

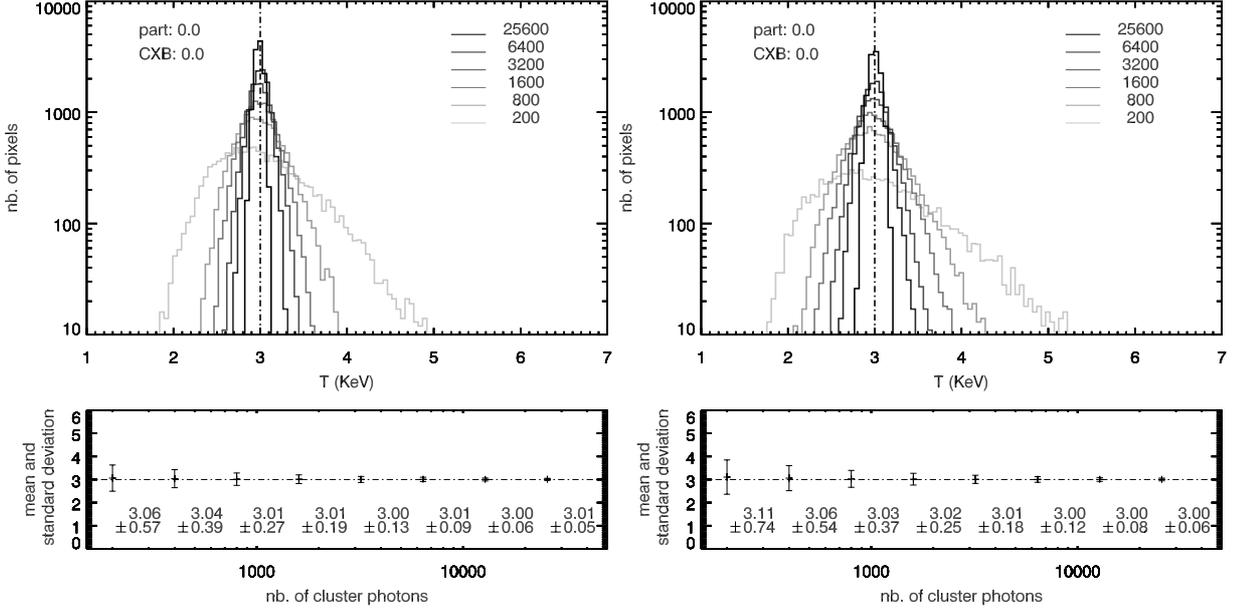


Fig. 4. Distribution of the results of the temperature fits for a simulated EPIC-XMM MOS1 observation of an isothermal extended source. The source emissivity is uniform. No background emission is included. The histograms with increasing darkness correspond to three fixed source emissivity N_S , from $N_S = 200$ to 25 600 photons per pixel. *Left*: fittings performed with uniform instrumental response simulated. *Right*: fittings performed with real instrumental response simulated.

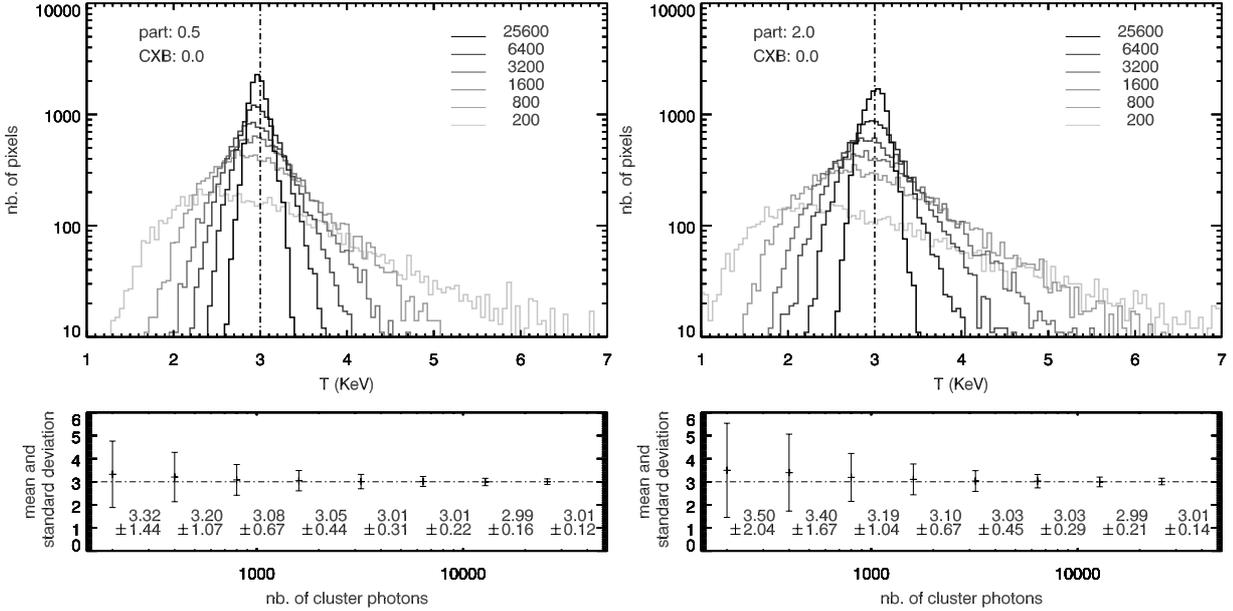


Fig. 5. Same as Fig. 4 with real instrumental response and particles background emission simulated. *Left*: background emissivity value of 0.5 times the cluster emissivity value. *Right*: Background emissivity value of two times the cluster emissivity value.

the high-energy particles. Yet despite the bias, the temperature estimator follows the MVB variance in the case of a cluster emissivity value higher than 400 photons per pixel.

4.1.5. Vignetted emission model with CXB

Let us now discuss the CXB influence on the spectral fitting thanks to a model series with vignetting and CXB emissivities of 0.5 and 2 times the cluster emissivity (cf. Fig. 6). Whatever the CXB emissivity is, the fitted temperatures peak at 3 keV

in the case of a cluster emissivity value higher than 1600 photons per pixel, with lower dispersion and temperature bias than for the toy-model series with particles background. Also, the temperature estimator behaves like an MVB estimator. The vignetted CXB provides many fewer high-energy photons than the particles background. Consequently, on the one hand the global background emission is lowered which reduces the temperature dispersion and on the other hand the lack of background low-energy photons reduces the low-temperature bias.

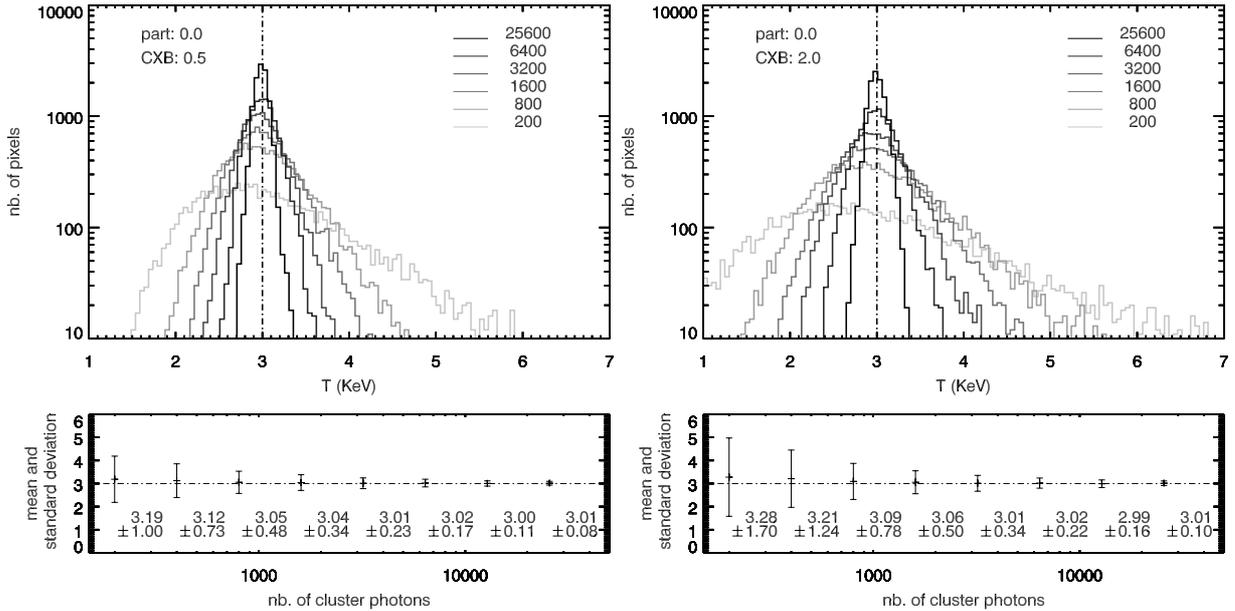


Fig. 6. Same as Fig. 4 with real instrumental response and cosmic X-ray background emission simulated. *Left:* background emissivity of 0.5 times the cluster emissivity value. *Right:* Background emissivity value of two times the cluster emissivity value.

4.1.6. Conclusion

Our temperature estimator is based on the maximum likelihood function computed from the emission model $F(T, e, k, l)$, which provides a close to MVB estimator. Actually the Rao-Cramer limit of MVB estimators cannot be reached, due to the dependency between both parameters r and T in the case of an emission model with background. The variance and bias of the estimator cannot be estimated analytically since the emission model is a numerical model, but tests like those performed enable us to evaluate them for different emission models. The temperature estimator behaves like an MVB estimator in the case of good statistics – more than about 500 photons per pixel – and low background – less than half of the cluster emissivity in our case. Its use in cases of low statistics or high background level is more critical and results must be analyzed carefully.

4.2. Multiscale temperature maps of simulated EPIC-XMM MOS1 observations

Let us now test the second step of our algorithm, i.e. the multiscale analysis, on non-isothermal temperature maps built from a model. The simulated source is an EPIC-XMM MOS1 2D model observation with emission model $F(T, e, k, l)$. Both particles and CXB background emissivities, N_P and N_B , have been set to five photons per pixel. The source is now composed of four isothermal emitting regions: an external ring at 3 keV, two elliptical regions at 4 and 6 keV respectively and the remaining central region at 5 keV. Three emissivity models following a 2-D Gaussian distribution with total flux of 10^5 , 10^6 and 10^7 photons are tested. The corresponding emissivity per pixel range from 10 photons at the edge of the images to 100, 1000 and 10 000 photons, respectively, at the centre of the image.

The images size is 128×128 and the wavelet analysis was performed on six scales in order to prevent large scale edge effects, so that details with typical width from 1 to 32 pixels are analysed. The source temperature distribution and reconstructed temperature maps are shown in Fig. 7. The temperature maps have been cut in the middle of each picture so as to show the temperature variations with error-bars in Fig. 8. As expected, the better the statistics is the better the model is reproduced. Due to the shape of the emissivity map of the source (cf. top-left plot in Fig. 8), the temperature variations can only be seen in the central regions with high statistics whereas the external regions with low statistics have been strongly smoothed. The temperature variations of the central regions have been shown at the first scale for the model with 10^7 photons, but have been smoothed due to the lower statistics for both other models. In the external regions, a large-scale mean temperature is computed whatever the global emissivity is due to the low statistics as well as the isothermality of the model.

The quality of the result can be checked by comparing the temperature profiles along the dashed lines of Fig. 7 with the model temperature values (cf. Fig. 8). Notice that such a comparison is not straightforward due to the spatial rebinning of the restored temperature maps. That is why the temperature profile of each restored map needs to be compared with the profile of a “rebinned” model map. The rebinned model maps are computed by applying the wavelet algorithm to the model temperature map with same thresholding of the coefficients as applied for each restored map. To do so, the temperature values are averaged within grids at different scales instead of fitting the data. Two rebinned model map are compared with each restored map. The first one is obtained by computing the arithmetical mean of the temperatures of the model map within the different bins. The second one is obtained by computing a

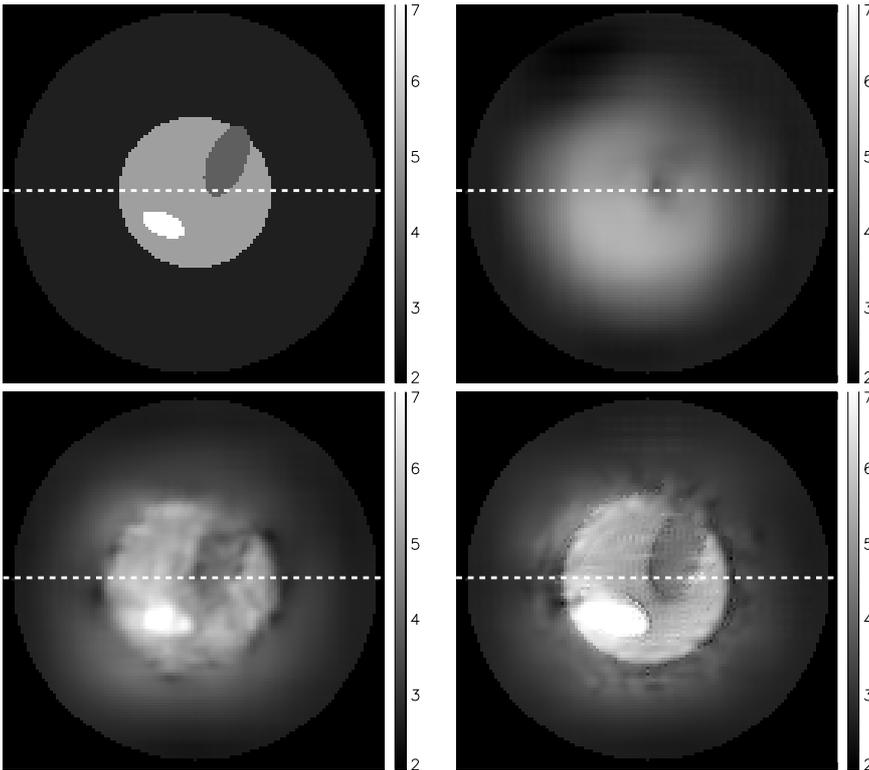


Fig. 7. Temperature maps of simulated EPIC-XMM MOS1 observations, – temperatures coded in keV –. *Top-left:* temperature model. *Top-right:* reconstructed temperature map for a global emissivity of 10^5 photons. *Bottom left and right:* reconstructed temperature map for a global emissivity of 10^6 and 10^7 photons, respectively. The dashed line on the reconstructed maps corresponds to profiles displayed in Fig. 8.

temperature mean which has been weighted by the cluster emissivity. The temperature profile of both rebinned model maps are superimposed to each restored map by means of dotted and dashed lines, respectively. Both cuts are closer to the cuts of the restored temperature maps than the cut of the non-rebinned temperature model. The cut computed from weighted means of the temperature model is in agreement with the restored temperature maps, since it is included within the corresponding error-bars.

To conclude about these simulated observations, the multi-scale algorithm has enabled us to adapt the spatial resolution to the local available statistics. The significance of the result has been checked thanks to a multiscale unfolding of the model temperature map. The agreement between the rebinned model map and restored map shows that the spectral fitting is satisfactory whatever the scale or position where it is performed, despite the spatial variation of the likelihood function.

5. Temperature maps of numerically simulated merging clusters

5.1. Hydrodynamic simulation of a merging cluster

In order to test the ability of our algorithm to detect the temperature variations expected within merging clusters, the observation of a cluster at different stages of merging have been simulated. This cluster has been obtained from adaptive mesh refinement 3-D hydrodynamical simulations of structures formation (for a complete description of the numerical simulation code, see Teyssier 2002). In order to simulate a merging cluster in a realistic cosmological environment, a low resolution ($N_{\text{part}} = 128^3$) large scale ($L_{\text{box}} = 400 h^{-1}$ Mpc)

Λ -CDM simulation was first performed. More than 100 large clusters of galaxies formed in this run. One cluster undergoing a violent merger was selected. Then, the exact same universe was re-simulated, adding many more particles and mesh points in the region surrounding the cluster. The same original box ($400 h^{-1}$ Mpc aside) is now described with 400 000 particles and 4 times more mesh points within the cluster virial radius. The mesh spacing in the cluster core has gone down to $6 h^{-1}$ kpc, corresponding to a formal resolution of 65536^3 . This “zoom” simulation lasted more than 100 CPU h on the Fujitsu VPP 5000 at CEA Grenoble. For this work, the cluster is always at the same comoving distance of $z = 0.1$. Just the time has increased from $z \approx 0.13$, $z \approx 0.09$, up to $z \approx 0.0$. This allows us to keep the same scale for all the images which is a square of 26.5 arcmin aside, i.e. a full EPIC-XMM Newton field of view. The projected gas emissivity and temperature of the merging cluster at each time are shown in Fig 9.

The map of averaged gas temperatures corresponding to the early stage of merging ($z \approx 0.13$) shows hot filaments and substructures ($T \approx 4$ keV) superimposed on a temperature background at 0.5 keV. The processes of gas heating and shocking are balanced by gas relaxation within each subgroup at $z \approx 0.09$, leading to two isothermal subgroups at respective temperature of 2 and 4 keV, where the hottest subgroup is the most massive. Eventually both subgroups have merged leading to a hot single one at $z = 0$. This merging cluster begins to reach its equilibrium temperature of 4 keV, with hot shocked regions superimposed from 5 to 6 keV. The temperature discontinuity between the central regions of the cluster and the background has increased during the merging process, since the central regions of the cluster have been heated whereas the background temperature remained constant.

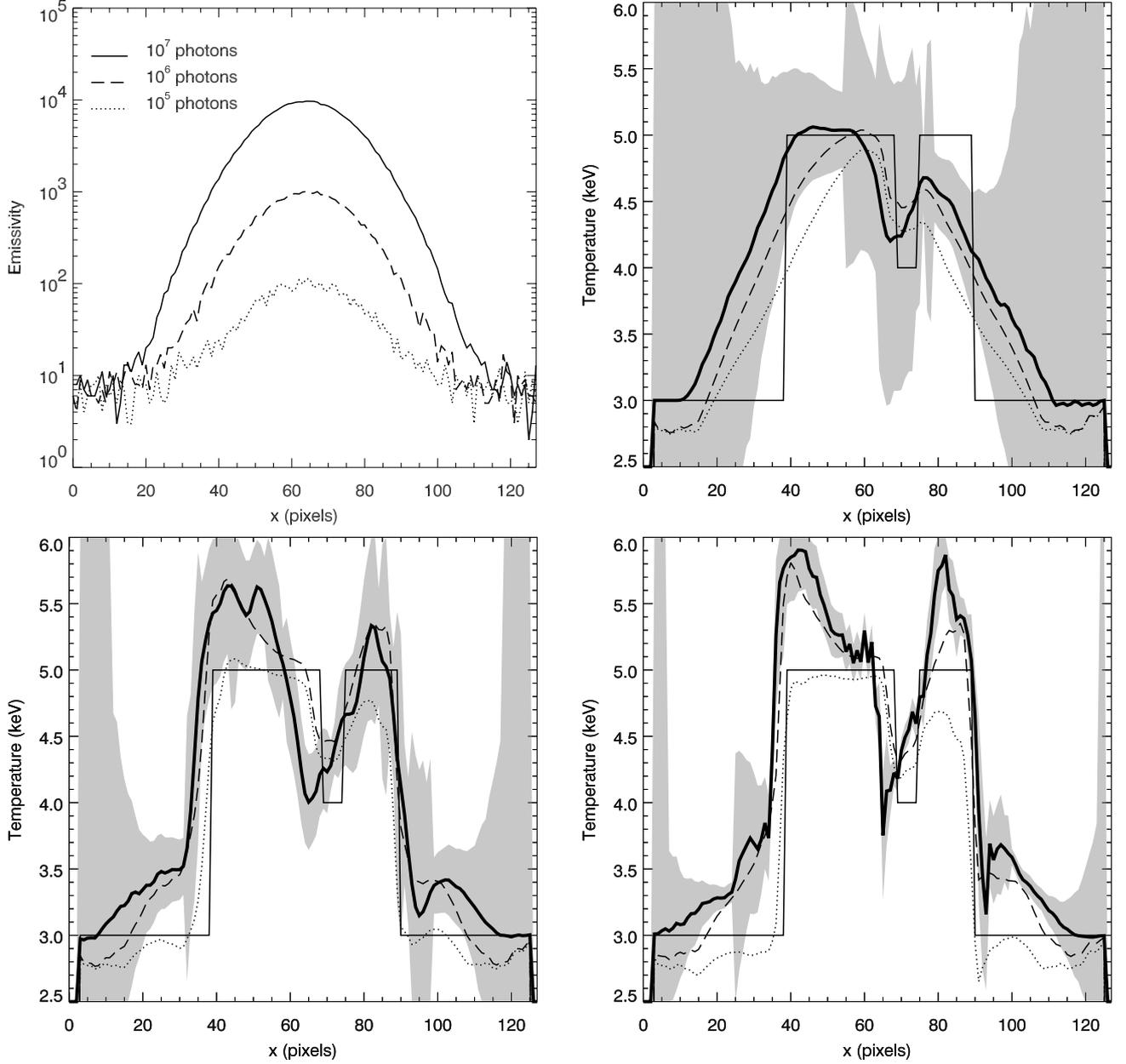


Fig. 8. Temperature profiles across the central region of simulated EPIC-XMM MOS1 observations, – temperatures coded in keV –. *Top-left:* emissivity profiles of the three emission models corresponding to global emissivities of 10^5 , 10^6 and 10^7 photons. *Top-right:* temperature profile with error-bars for a global emissivity value of 10^5 photons. *Bottom left and right:* temperature profiles with error-bars for respective global emissivity values of 10^6 and 10^7 photons. Bold line: temperature cuts, with grey shaded 3σ significance region superimposed. Thin line: temperature model. Dotted line: temperature model with the same wavelet thresholding as the temperature maps. Dashed line: temperature model with the same wavelet thresholding as the temperature maps and emissivity weighted mean temperature per bin.

5.2. Simulation of an EPIC-XMM MOS1 observation

The 3-D temperature $T(k, l, m)$ at pixel k, l, m , and emissivity $N_C(k, l, m)$ distributions of the intra-cluster plasma being known, an EPIC-XMM MOS1 observation can be simulated. To do so, the emitted photons are randomly drawn among different temperatures from the emission model of Poisson parameter $T(k, l, m)$ and have to follow the optical path from the source to the EPIC-XMM MOS1 focal plane. The different

distortions of the signal that would occur for a real cluster at $z = 0.1$ are taken into account: redshift, galactic absorption and instrumental response including the PSF, vignetting and energy response of the CCD detectors. The cluster and particle background emissivities correspond to standard emissivities for a real 5 keV cluster at $z = 0.1$ observed during 60 ks, yielding a background emissivity of $N_P = 0.67$. The CXB has not been taken into account. The simulated EPIC-XMM count maps with restored temperature maps are shown in Fig. 10.

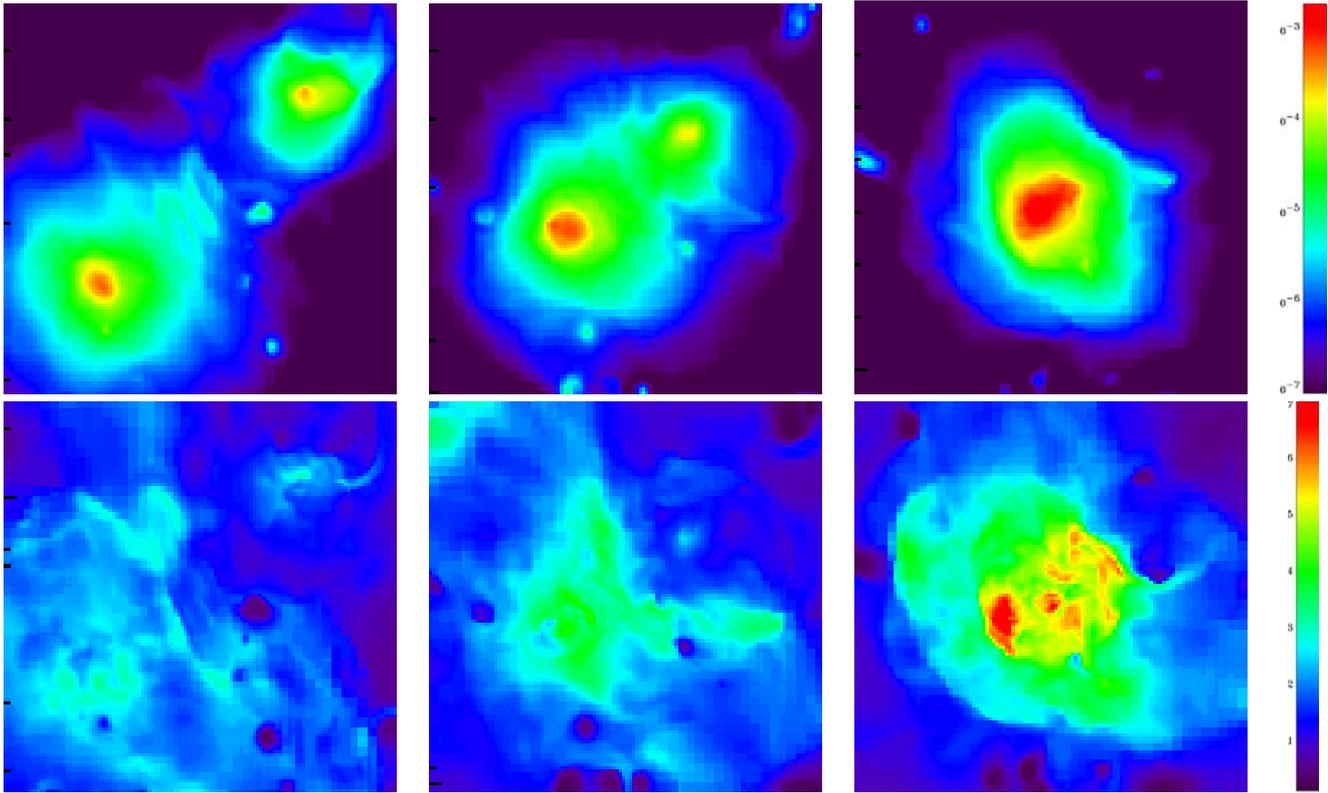


Fig. 9. Simulated merging cluster at $z \approx 0.13$, $z \approx 0.09$ and $z \approx 0.0$ from left to right. *Top*: emissivity maps, with emissivity color code in $\text{ergs s}^{-1} \text{cm}^{-2}$. *Bottom*: maps of the mean gas temperature averaged along the line of sight, with temperature colors coded in keV. The width of the field of view is 26.5 arcmin.

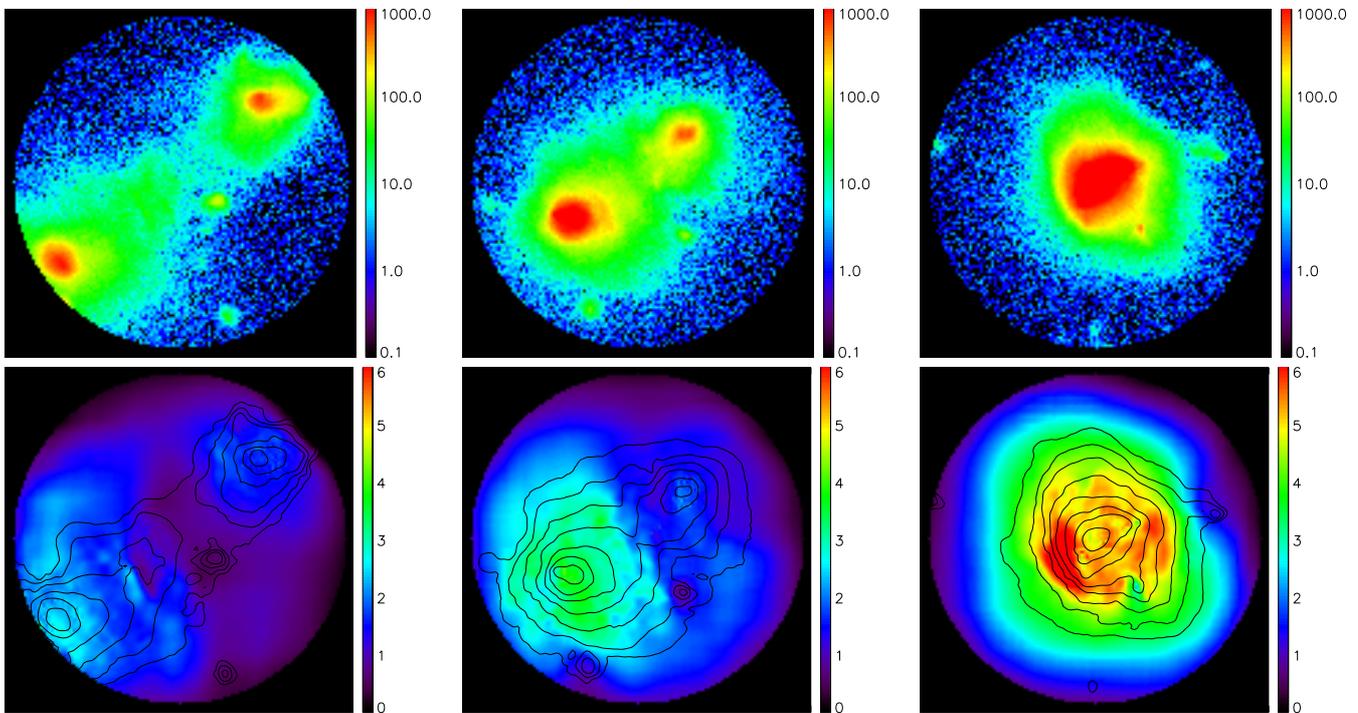


Fig. 10. Simulated merging cluster at $z \approx 0.13$, $z \approx 0.09$ and $z \approx 0.0$ respectively. *Top*: simulated EPIC-XMM MOS1 photons counts maps, with emissivity coded in counts. *Bottom*: temperature maps with logarithmic emissivity contours superimposed, the temperatures are coded in keV.

5.3. Temperature restoration

The temperature maps of Fig. 10 have been restored with a 1σ significant threshold. This simulated observation is closer to the observation of a real cluster than the models of the previous section. Indeed the instrumental components are supposed to be similar, but the source emission is no longer a single component model due to projection effects. We consequently must bear in mind that the spectral fitting of each resolution element is not as accurate as described in Sect. 4.1. Let us nevertheless compare the restored temperature maps of Fig. 10 with the maps of averaged gas temperatures of Fig. 9.

The counts vary by several orders of magnitude between high and low density regions (cf. count maps of Fig. 10). As expected from a multiscale algorithm, the resolution of the temperature maps is better in regions with high emissivity and statistics, than in regions with poor statistics. For instance, the closer to the core of the subgroups, the better resolved are the details of the hot substructures of the early stage of merger ($z \approx 0.13$). Likewise the hot shocked regions in the central region of the cluster at $z = 0$ are well restored despite the light smoothing applied, whereas both subgroups at $z \approx 0.09$ and the single cluster at $z = 0$ have been detected at large scale as isothermal structures because of low statistics in the external regions.

These results on a simulated merging cluster at $z = 0.1$ allow us to conclude that our multiscale algorithm applied to real EPIC-XMM Newton observation of merging clusters should enable us to detect the expected temperature variations if they exist. Nevertheless the main building blocks of our algorithm, i.e. the spectral fitting and the multiscale temperature restoration, may still be improved. The spectral fitting uses a single component temperature model and a spectral modelling of projection effects with a multicomponent emission model might improve the detection of temperature variations. Also, the multiscale algorithm relies on a Haar wavelet which matches the square sampling of the data but leads to square patterns on the restored image, despite the regularization. Future improvements of the algorithm will include the use of other wavelets in order to better match the shape of the sources.

6. Conclusion

To take advantage of the spatial resolution of the new X-ray spectro-imagers, a new spectro-imagery algorithm has been introduced. This algorithm has been especially designed for the EPIC-XMM Newton cameras in order to restore the gas temperature variations within clusters of galaxies. The algorithm consists of two parts. First, an emission model of the source is fitted around each pixel of the detector surface, using regions of different areas, leading to a set of temperatures with error-bars at different scales. Then, the temperature variations are coded and thresholded in the wavelet space according to a significance criterion, and the wavelet transform of the optimal temperature map is obtained by regularizing the non-significant wavelet coefficients.

For the spectral fitting step, an emission model is fitted using a maximum likelihood method. The source emission

spectrum, the astrophysical and instrumental backgrounds, but also the instrumental response and especially the vignetting factor are taken into account within the emission model. As far as the vignetting is concerned, this approach is different from the usual solution, where the vignetting is taken into account by modifying the data instead of the emission model, e.g. the “weight method” described in Arnaud et al. (2001). The spectral fitting with full instrumental response modelling is more accurate in regions with very poor statistics and enables us therefore to reduce significantly the error bars in external regions of the FOV where the vignetting factor is important. However a temperature bias has been revealed in cases of poor statistics. This bias is enhanced in the case of high emissivity of the background with regard to the source emissivity, due to degeneracy effects in the shape of the emission spectra of each component. This result has to be kept in mind for sources like compact groups with low emissivities with regard to the background emissivity in the central region of the focal plane.

Our multiscale algorithm enables us to adapt the spatial resolution to the local statistics. In our test on a numerical simulation of a merging cluster at $z = 0.1$, the multiscale algorithm has enabled us to reveal not only faint temperature variations in the core of the merging subgroups, but also a large-scale temperature decrease in their external regions. The multiscale approach provides much more information than other methods using square or ring kernels. Nevertheless the results have still to be interpreted carefully. The final map has indeed been computed from a linear combination of spatially rebinned informations, and the knowledge of the area used for spectral fitting is not straightforward. Also, the chosen Haar wavelet transform is well suited to match the square spatial sampling but is not the best one since it implies block effects. Other wavelets that better match the temperature structures shape should be investigated.

Due to the spatial variation of the instrumental response, the different temperatures within different spatial bins are first computed from spectral fittings using different emission models. Thus, the wavelet transform of the expected temperature map is not computed linearly from the data. But a real wavelet transform is a linear process. This wavelet transform of the searched temperature map is obtained iteratively by regularizing the non-significant wavelet coefficients, which is questionable due to the non-linearity of the process. Yet this process has been tested on simulations and seems to perform well. The good behavior of the algorithm is due to the low spatial variations of the emission model, and especially to the low relative spatial variations of temperatures within the intra-cluster plasma. Consequently such an algorithm is well suited for restoring spatial variations of temperature within extended sources like clusters of galaxies but would not work for sources with high local variations of the emission model such as supernovae remnants.

The current spectro-imagery algorithm will be applied to a sample of merging clusters at different phases. Future developments are considered. The spectral fitting process could include spatial variations of the emission model in order to take into account projection effects. Moreover variations of other parameters such as the gas metallicity could be measured.

Appendix A: The maximum likelihood pursuit: A modified conjugate gradient method

The model fitting algorithm is based on the pursuit of the maximum of the likelihood function $L(r, T)$ – cf. Eq. (5) – over the 2-D space with parameters r and T . To do so, a 2-D minimization algorithm is required. The conjugate gradient method was chosen for its fast convergence, but adapted to a bounded parameter space with possibility of multiple maxima of the likelihood function.

Let us recall the general method of minimizationalgorithms. Considering a n -D function $S(X)$. The function $S(X)$ and its gradient can be developed over the initial position X_0 :

$$S(X) = S(X_0) + [X - X_0]' \nabla_0 S(X_0) + \frac{1}{2} [X - X_0]' H_0 [X - X_0], \quad (\text{A.1})$$

$$\nabla S(X) = \nabla_0 S(X_0) + H_0 [X - X_0], \quad (\text{A.2})$$

where H and ∇ are the respective Hessian and gradient operators and H_0 and ∇_0 their value at position X_0 .

The minimization is obtained for $\nabla S(X) = 0$. Since $S(X)$ is usually non-linear S has to be developed iteratively over successive positions X_i . Minimization are usually performed over 1-D cuts of $S(X)$. The successive directions of the cuts depending on the algorithms. In case of the conjugate gradient algorithm, a given position X_i corresponds to a minima in direction P_i . Then the next direction P_{i+1} is chosen to be a *conjugate* direction from P_i . Contrary to the simple gradient algorithm where P_{i+1} is normal to P_i , here P_{i+1} is chosen so that ∇S stays as far as possible normal to direction P_i , i.e. $P_i' \delta[\nabla S(X)] = 0$. This condition prevents successive explorations in the same direction. By differentiation of Eq. (A.2) and by taking into account that $P_{i+1}' \delta X = 0$ at this iteration, we get:

$$\delta[\nabla S(X)] = H_0 \delta X, \quad (\text{A.3})$$

$$P_i' H_0 P_{i+1} = 0. \quad (\text{A.4})$$

Let us now describe the adaptation of this standard algorithm to our case with a 2-D bounded parameter space (cf. Fig. A.1) with possibility of multiple maxima. At iteration i , starting from X_{i-1} , a maximum of $L(r, T)$, X_i , is fetched along the conjugate direction from P_{i-1} : P_i . We are looking for the maximum of the 1-D cut of $L(r, T)$: $L_1(x)$ with origin $x = 0$ at point X_{i-1} . To do so we define a bounded exploration support $[x_{\min}, x_{\max}]$ of $L_1(x)$.

In any case, both bounds x_{\min} , x_{\max} must be included within the bounded parameter space, so that the exploration support $[x_{\min}, x_{\max}]$ must be included within the $[\tilde{x}_{\min}, \tilde{x}_{\max}]$ interval where \tilde{x}_{\min} and \tilde{x}_{\max} are the parameter space bounds along the direction P_{i-1} (cf. Fig. A.1). In case of good statistics the likelihood function $L(r, T)$ has only one maximum within the parameter space and the conditions $x_{\min} = \tilde{x}_{\min}$, $x_{\max} = \tilde{x}_{\max}$ are enough. But in the case of bad statistics, some multiple maxima events can appear, so that a choice has to be made among the possible solutions. We observed in simulations of observations that the good one is generally the closest to the center of the parameter space. Indeed two kinds of false maxima

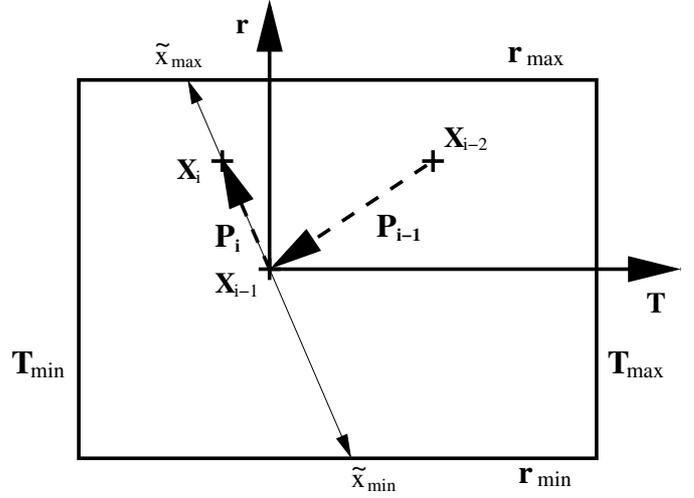


Fig. A.1. The bounded parameter space. At iteration i the minimum is obtained from X_{i-1} over direction P_i along the bounded segment $[\tilde{x}_{\min}, \tilde{x}_{\max}]$.

can appear on the T -axis. The first ones concern the low temperatures ($T < 1$ keV) in case of degenerescence due to the similar shape of the emission-lines complex and CXB emission in the soft band. The second ones concern the high temperatures ($T > 10$ keV) in case of degenerescence between the bremsstrahlung part of the cluster emission and CXB emission in the hard band. Consequently, the closest solution to the center of the parameter space, – where the algorithm starts its pursuit –, is chosen in the case of multiple maxima. To do so, a new constraint is added to the exploration support $[x_{\min}, x_{\max}]$ at iteration i : the right solution must be the closest to the starting point $x = 0$, so that x_{\min} and x_{\max} are randomly drawn in the respective $[\tilde{x}_{\min}, 0]$, and $[0, \tilde{x}_{\max}]$ intervals up to get values where $\frac{dL_1(x_{\min})}{dx} > 0$ and $\frac{dL_1(x_{\max})}{dx} < 0$. In case of bad statistics where no support $[x_{\min}, x_{\max}]$ can match the previous condition, the undefined value is returned for both parameters.

This local pursuit of function $L_1(x)$ is performed at each iteration of the standard conjugate gradient algorithm. In case of good statistics where a maximum is found, the convergence criterion $\|\nabla L(r, T)\| < \epsilon$ is obtained in a few iterations with $\epsilon = 1 \times e^{-6}$.

Appendix B: The 2-D Haar wavelet transform

The wavelet transform of a 2-D signal $I(k, l)$ is based on a wavelet function which can be dilated and translated. The 2-D wavelet leads to functions with compact support depending on three parameters: the “scale” or dilation factor a and both space coordinates k, l . In our case, for geometrical and computational reasons, the chosen wavelet is the Haar wavelet (cf. Fig. B.1).

Since the $I(k, l)$ distribution is sampled in the direct space a discrete wavelet transform $W(a, k, l)$ can be associated to $I(k, l)$ while preserving all the signal information. To do so the scale axis can be sampled within a dyadic scheme so that $a = 2^j$ with $j = 0, \dots, N$. The wavelet transform can be redundant within the space axes, such as in the so called “a trous”

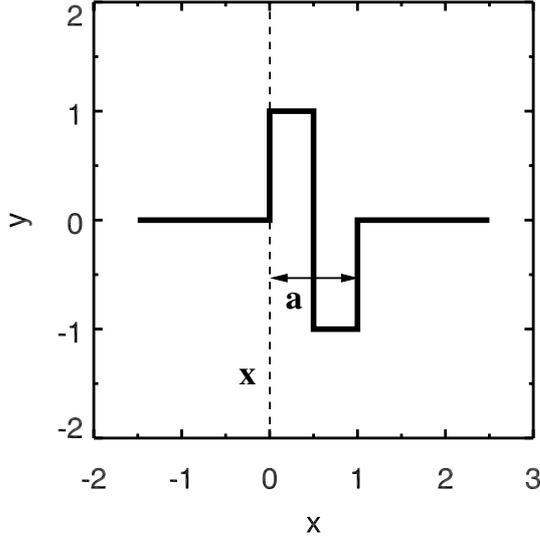


Fig. B.1. The 1-D Haar wavelet and its characteristic parameters scale (a) and location (x).

algorithm (Holschneider 1989), or non-redundant, such as in the Mallat multi-resolution algorithm (Mallat 1989).

On the one hand, the computation of the wavelet transform $W(a, k, l)$ can be performed from the 2-D signal $I(k, l)$. Given a scale a , “smoothed planes” $A(a, k, l)$ and wavelet coefficients $W(a, k, l)$ can be defined, both being computed from the smoothed plane $A\left(\frac{a}{2}, k, l\right)$ by applying low-pass $h\left(\frac{a}{2}\right)$ and high-pass $g\left(\frac{a}{2}\right)$ analysis filters, respectively. The smoothed plane $A(a, k, l)$ corresponds to the smoothing of $I(k, l)$ with a kernel of width a , whereas the wavelet coefficients $W(a, k, l)$ corresponds to the information lost between $I(k, l)$ and $A(a, k, l)$. Starting at scale $a = 1$, the first smoothed plane $A(1, k, l) \equiv I(k, l)$ can be defined, and the coefficients $W(a, k, l)$ and planes $A(a, k, l)$ can be computed iteratively from scale to scale.

The wavelet transform using the anisotropic 2-D Haar wavelet is composed of three kinds of coefficients, the horizontal, vertical and diagonal coefficients, $W_h(a, k, l)$, $W_v(a, k, l)$, $W_d(a, k, l)$, respectively. According to the redundant shift invariant wavelet transform (Coifman & Donoho 1995), they are computed by applying the corresponding, horizontal, vertical and diagonal analysis filters, of Fig. B.2, following the analysis equations:

$$A(a, k, l) = H\left(\frac{a}{2}\right)A\left(\frac{a}{2}, k, l\right), \quad (\text{B.1})$$

$$W_h(a, k, l) = G_h\left(\frac{a}{2}\right)A\left(\frac{a}{2}, k, l\right), \quad (\text{B.2})$$

$$W_v(a, k, l) = G_v\left(\frac{a}{2}\right)A\left(\frac{a}{2}, k, l\right), \quad (\text{B.3})$$

$$W_d(a, k, l) = G_d\left(\frac{a}{2}\right)A\left(\frac{a}{2}, k, l\right). \quad (\text{B.4})$$

On the other hand the computation of the 2-D signal $I(k, l)$ can be performed from its wavelet transform $W(a, k, l)$. According to the biorthogonal wavelet formalism, the so-called wavelet reconstruction is performed by applying at higher and higher resolution the respective low-pass $\tilde{h}\left(\frac{a}{2}\right)$ and high-pass $\tilde{g}\left(\frac{a}{2}\right)$

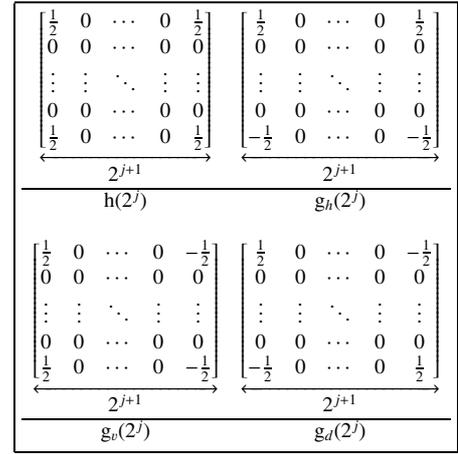


Fig. B.2. The two dimensional analysis filters at scale 2^j . The low-pass $h(2^j)$ and the three high-pass $g_{h,v,d}(2^j)$ filters associated to the horizontal, vertical and diagonal details respectively.

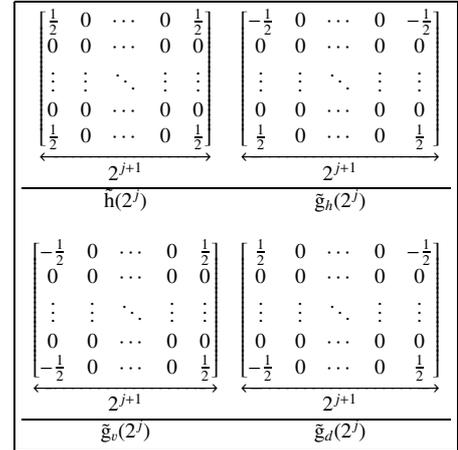


Fig. B.3. The two dimensional synthesis filters at scale 2^j . The low-pass $\tilde{h}(2^j)$ and the three high-pass $\tilde{g}_{h,v,d}(2^j)$ filters associated to the horizontal, vertical and diagonal details respectively.

synthesis filters to the wavelet coefficients $W(a, k, l)$ and smoothed planes $A(a, k, l)$. The filtered wavelet and smoothed planes are added, and the right renormalization leads to a new smoothed plane $A\left(\frac{a}{2}, k, l\right)$, and so on at higher and higher resolution. The last smoothed plane at the highest resolution constitutes the 2-D signal $I(k, l)$.

In the case of the 2-D Haar wavelet transform, the smoothed plane $A\left(\frac{a}{2}, k, l\right)$ at scale $\frac{a}{2}$, is computed from a linear combination of the smoothed plane $A(a, k, l)$ and the wavelet coefficients $W_{h,v,d}(a, k, l)$, following the synthesis equation:

$$A\left(\frac{a}{2}, k, l\right) = \left[\tilde{H}A(a, k, l) + \tilde{G}_h W_h(a, k, l) + \tilde{G}_v W_v(a, k, l) + \tilde{G}_d W_d(a, k, l) \right]. \quad (\text{B.5})$$

References

- Arnaud, M., Neumann, D. M., Aghanim, N., et al. 2001, *A&A*, 365, L80
- Bobichon, Y., Bijaoui, A. 1997, *Exp. Astron.*, 7, 239
- Coifman, R. R., & Donoho, D. 1995, Translation invariant denoising, Technical report 475, Dept. of Statistics, Stanford University
- Colberg, J. M., White, S. D. M., Yoshida, N., et al. 2000, *MNRAS*, 319, 209
- Evrard, A. E. 1990, *ApJ*, 363, 349
- Evrard, A. E., MacFarland, T. J., Couchman, H. M. P., et al. 2002, *ApJ*, 573, 7
- Sanders, J. S., & Fabian, A. C. 2001, *MNRAS*, 325, 178
- Holschneider, M., Kronland-Martinet, R., Morlet, J., & Tchamitchian, P. 1989, A real-time algorithm for signal analysis with help of the wavelet transform, *Wavelets: Time frequency methods and phase-space*, 286
- Haar, A. 1910, Zur theorie der orthogonalen Funktionensysteme, *Mathematische Annalen*, 69, 331
- Jammal, G., & Bijaoui, A. 1999, A multiresolution image restoration method for photon imaging systems. *Proc. IEEE Intern. Conf. Image Processing*, 4, 3401
- Katz, N., & White, S. D. M. 1993, *ApJ*, 412, 455
- Kendall, M. G., Stuart, A. 1973, *The advanced theory of statistics*, T.2 (Londres: Griffin)
- Lumb, D. H., Warwick, R. S., Page, M., & De Luca, A. 2002, *A&A*, 389, 93
- Mallat S. 1989, A theory for multiresolution signal decomposition: the wavelet representation, *IEEE Trans. on pattern analysis and machine intelligence*, 11, 674
- Mewe, R. 1972, *SpH*, 22, 459
- Mewe, R. 1975, *SpH*, 44, 383
- Mewe, R., Gronenschild, E. H. B. M., & van den Oord, G. H. J. 1985, *A&AS*, 62, 197
- Mewe, R., Lemen, J. R., & van den Oord, G. H. J. 1986, *A&AS*, 65, 511
- Ricker, P. M., & Sarazin, C. 2001, *ApJ*, 561, 621
- Ritchie, W. R., & Thomas, P. A. 2002, *MNRAS*, 329, 675
- Roettiger, K., Burns, J. O., & Loken, C. 1996, *ApJ*, 473, 651
- Sarazin, C. L. 1986, *Rev. Mod. Phys.*, 58, 1
- Teyssier, R. 2002, *A&A*, 385, 337
- Tikhonov, A. N. 1963, Regularization of incorrectly posed problems, *Sov. Math. Dokl.*, 4, 1624
- Van Cittert P. H. 1931, *Z. Phys.*, 69, 298