THE ASTROPHYSICAL JOURNAL, 471:673–682, 1996 November 10


# MAPPING THE GAS TEMPERATURE DISTRIBUTION IN EXTENDED X-RAY SOURCES AND SPECTRAL ANALYSIS IN THE CASE OF LOW STATISTICS: APPLICATION TO *ASCA* OBSERVATIONS OF CLUSTERS OF GALAXIES

E. CHURAZOV AND M. GILFANOV
Space Research Institute (IKI), Profsouznaya 84/32, Moscow 117810, Russia
AND
W. FORMAN AND C. JONES
Smithsonian Astrophysical Observatory, 60 Garden Street, Cambridge, MA 02138
*Received 1995 October 20; accepted 1996 May 29*

## ABSTRACT

A simple method for mapping the temperature distribution in extended sources is developed for application to *ASCA* observations of galaxy clusters. Unlike the conventional approach to spatially resolved spectral analysis, this method does not require nonlinear minimization and is computationally fast and stable. Therefore, it can be implemented for a large number of regions or on a fine spatial grid. Although based on a Taylor expansion over the nonlinear parameter, the method is found to be accurate in many practical situations, the relative error for the temperature estimate being less than 2%–4% when the plasma temperature exceeds $\sim 2$ keV. This method is not intended to replace conventional spectral analysis but to supplement it, providing relatively fast and easy construction of temperature maps, which may be used as a guide to further detailed analysis of particularly interesting regions using conventional spectral fitting.

Conventional spectral analysis in the case of moderate and low numbers of counts is discussed. A practical recipe for unbiased parameter estimation is suggested and verified in Monte Carlo simulations for commonly used spectral models. A simple modification of the $\chi^2$ statistic (calculation of weights based on the smoothed observed spectrum) yields nearly unbiased parameter estimates and correct confidence interval determination with no need for regrouping (binning) the energy channels even in the case of low statistics ($\sim 50$–$100$ counts in the observed spectrum with several hundred channels).

*Subject headings:* galaxies: clusters: general — methods: data analysis — methods: statistical — X-rays: galaxies

## 1. INTRODUCTION

The instruments on the *ASCA* satellite (Tanaka et al. 1994) provide sufficiently powerful imaging and spectroscopic capabilities so that, for the first time, it is possible to map the temperature and iron abundance distributions in the intracluster medium in clusters of galaxies. However, three serious problems are associated with analysis of *ASCA* cluster observations:

1. In practice, spatially resolved spectral analysis can be performed only on a rather coarse grid with cells of large angular size because iterative minimization procedures, required in conventional spectral analysis, are usually time consuming and not always stable, especially in the case of low statistics.

2. Very often observed spectra contain only few (or even zero) counts per bin in high-energy channels, and special care must be taken to avoid biased parameter estimates while fitting the data with the model.

3. The point-spread function (PSF) of the *ASCA* mirrors has broad, energy-dependent wings, which lead to an energy-dependent redistribution of counts over the image and consequently can lead to a distortion of the derived spectral parameter map.

Problems (1) and (2) are common in the analysis of X-ray data, while (3) is specific to X-ray telescopes with nested conical mirrors like those used for *ASCA*. Because of (1), various quantities like hardness ratios or X-ray colors are often used to characterize the properties of observed spatially resolved spectra. The disadvantage of these methods is that usually they are not "optimally sensitive" and do not provide a direct measure of the physical parameter of interest (e.g., gas temperature). To overcome problem (2), a conventional approach is to use maximum likelihood fitting or to group the data into broad energy channels. However, sometimes neither of these solutions is acceptable. In this paper we describe two practical methods for addressing (1) and (2) and describe their application to *ASCA* observations of galaxy clusters. In § 2 we describe a method of mapping the temperature distribution in extended sources, which is as fast as hardness ratio calculations but has the advantage of providing a direct measure of the gas temperature with nearly the minimum possible uncertainty on the value of the fitted parameter. In § 3 we discuss a simple modification to the conventional $\chi^2$ fitting technique leading to unbiased parameter estimates. Although developed and tested in application to *ASCA* observations of galaxy clusters, these methods can be easily applied to the analysis of observations from other imaging missions (e.g., *ROSAT*). Possible ways to account for the energy-dependent *ASCA* PSF (problem 3) will be discussed in a subsequent paper.

## 2. MAPPING THE TEMPERATURE DISTRIBUTION IN EXTENDED SOURCES

One of the common tasks in X-ray studies of extended sources is constructing a temperature map. An obvious and direct approach is to perform spatially resolved spectral fits to the data. In practice, however, conventional spatially

resolved spectral analysis can be implemented only for a limited number of regions or on a rather coarse spatial grid because of the very time consuming nature of the minimization procedure and because the minimization is not always stable, especially in the case of moderate and low statistical precision data. For this reason various quantities characterizing the spectral shape, such as hardness ratios, X-ray colors, and average photon energies, are used instead. These alternatives are usually easy to calculate and are applicable for almost any type of spectrum, e.g., the hardness ratios or average photon energies are higher for regions having harder spectra. However, they often are not the "most sensitive" (i.e., the dispersion of the parameter estimate is larger than can be obtained using conventional spectral fitting) and they do not directly give the value of the physical quantity of interest (e.g., hardness ratio must be converted to temperature).

On the other hand, for many classes of extended sources we almost always have in mind a particular spectral model (e.g., emission from an optically thin plasma with some admixture of heavy elements), and, in fact, for each particular source, we even frequently know a priori the reasonable dynamic range for the interesting physical parameter, such as temperature (e.g., 2–10 keV for a typical, rich galaxy cluster). Using this a priori knowledge one can replace conventional spectral analysis with a method which is (1) computationally fast and stable (similar to hardness ratio calculations) and therefore can be implemented for any number of regions or on a fine spatial grid, (2) directly gives the value of the physical quantity, and (3) the parameter estimate dispersion of which is nearly optimal and approaches that of direct spectral fitting.

### 2.1. *Method of Linearized Parameter Estimation*

Consider for simplicity a spectral model with one nonlinear parameter $p$: $M(A, p, E) = Af(p, E)$, where $E$ is the energy and $A$ is the normalization. We choose reference values of $p_1$ and $p_2$ such that $[p_1, p_2]$ bounds the expected range of the interesting physical parameter $p$. Any spectrum characterized by the physical parameter $p$ can then be approximated by a linear combination of the two reference spectra $f(p_1)$ and $f(p_2)$:

$$M(A, p, E) \approx L(K_1, K_2, E) = K_1 f(p_1, E) + K_2 f(p_2, E)\,, \quad (1)$$

where $K_1$ and $K_2$ are the relative normalizations of the two reference spectra. Examples for an optically thin thermal bremsstrahlung and optically thin thermal plasma emission models, convolved with the *ASCA* SIS0 response, are shown in Figure 1 and demonstrate that a spectrum with temperature $kT = 6$ keV can be reasonably well approximated as a weighted sum of spectra with $kT_1 = 3$ keV and $kT_2 = 8$ keV.

Obviously, an unambiguous relation exists between values of $K_1$, $K_2$, $p_1$, $p_2$ and the physically interesting parameter $p$. In the above example for the thermal bremsstrahlung model, the coefficients $K_1$ and $K_2$ correspond to the emission measures of two components at temperatures $T_1$ and $T_2$ which approximate the spectrum for emission appropriate to a temperature $T$. A reasonable expression for the temperature $T$ would be: $T \approx (EM_1 T_1 + EM_2 T_2)/(EM_1 + EM_2)$. Similarly, in general, the simplest form of such a relation giving the exact values of parameter $p$ at the endpoints of the interval $[p_1, p_2]$ [i.e., for $M(A, p, E) = Af(p_i, E)$, $i = 1, 2$] would be

$$p \approx \frac{K_1 p_1 + \Psi K_2 p_2}{K_1 + \Psi K_2}\,, \quad (2)$$

with a free coefficient $\Psi$. A possible (but not unique) way to choose the value of $\Psi$ is to require that equation (2) give an exact value for $p$ at some arbitrary third point $p = p_m$ [i.e., for $M(A, p, E) = Af(p_m, E)$] with a natural choice of $p_m$

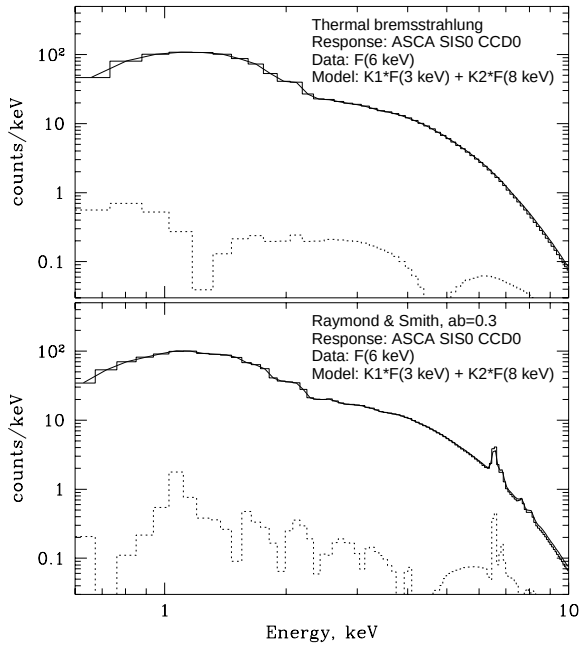


FIG. 1.—Approximation of spectra for thermal bremsstrahlung (*top*) and thermal emission from an optically thin plasma (*bottom*) with $kT = 6$ keV as a linear combination of two reference spectra ($kT_1 = 3$ keV and $kT_2 = 8$ keV). The spectra were convolved with the response of *ASCA* SIS0 CCD0 (s0c0g0234p40e1_512v0_8i.rmf in standard *ASCA* GOF notation). The dotted lines show the absolute value of the residuals. The fraction of the photons incorrectly characterized by the model [i.e., $\sum$ abs(Data–Model)] is 0.58% for bremsstrahlung and 0.74% for optically thin thermal plasma emission.

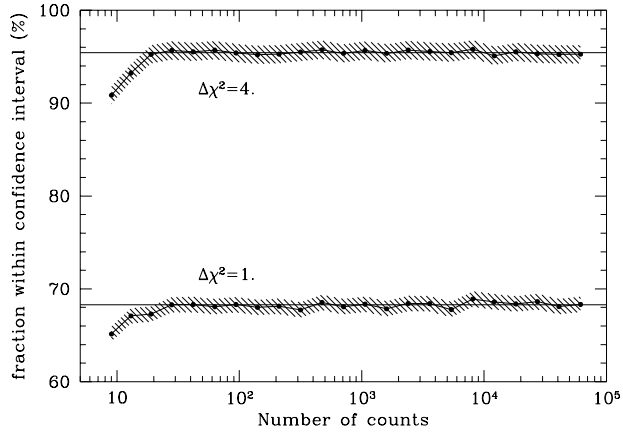


FIG. 2.—Monte Carlo simulations of the error estimates according to the algorithm described in the text. The chosen reference spectra were $T_1 = 2$ keV, $T_2 = 12$ keV, $T_m = 7$ keV. A total of 10,000 spectra with $T = 5$ keV have been simulated for each value of the total number of counts in the spectrum. The shaded area shows the 1 $\sigma$ uncertainty of the Monte Carlo simulations. The two solid lines are 68.3% and 95.45% levels.

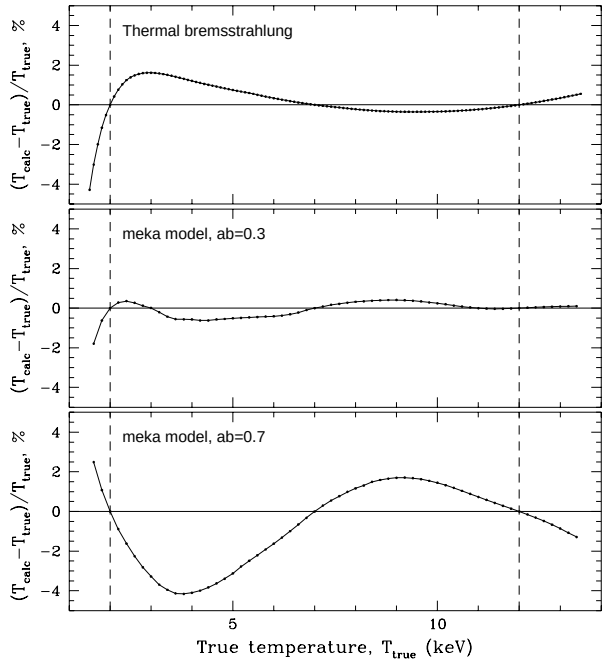


FIG. 3.—Application of the method of temperature estimation described in § 2 to various emission models: optically thin thermal bremsstrahlung (*top*) and thermal emission from an optically thin plasma with heavy element abundances $a = 0.3$ (*middle*) and $a = 0.7$ (*bottom*). The energy response of the *ASCA* SIS0 (s0c0g0234p40e1_512v0_8i.rmf in standard *ASCA* GOF notation) was used. The choice of reference spectra were $T_1 = 2$ keV, $T_2 = 12$ keV, $T_m = 7$ keV. The vertical axis represents the expected relative error of the temperature estimate, in percent. The temperature, $T_{calc}$, calculated from the linear approach described in the text, provides a remarkably accurate estimate of the true temperature $T_{true}$ over a wide temperature range with errors of only a few percent.

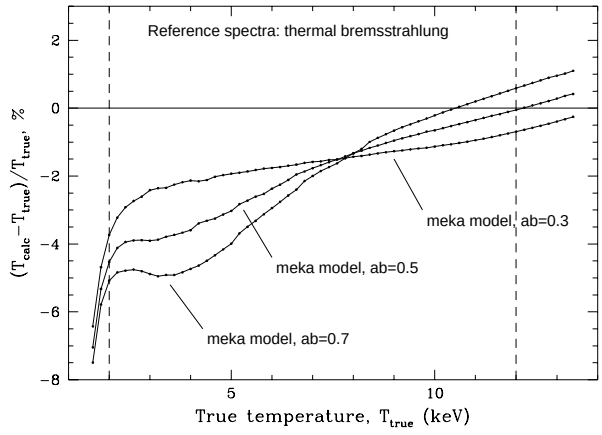


FIG. 4.—Application of the method of temperature estimation described in § 2 to the case of a "wrong model": the reference spectrum (optically thin thermal bremsstrahlung, the "measured" spectrum for which the temperature is to be determined) thermal emission from an optically thin plasma with heavy element abundances of $a = 0.3$, 0.5, and 0.7. The choice of reference spectra were $T_1 = 2$ keV, $T_2 = 12$ keV, $T_m = 7$ keV.

lying midway between $p_1$ and $p_2$.[1] If we then fit the measured spectrum $C(E)$ with the *linear* model $L(K_1, K_2, E)$ defined by equation (1), where $K_1$, $K_2$ are parameters of the model and $p_1$, $p_2$ are fixed, then equation (2), with a proper choice of $\Psi$, provides an estimate of the parameter $p$. This transforms the nonlinear iterative fitting of the parameter $p$ into the straight forward determination of the coefficients $K_1$ and $K_2$ at each desired spatial location.

There are several important points worth noting about the parameter estimation method described above:

1. This method of parameter estimation is equivalent to a first-order Taylor expansion over the nonlinear parameter to be fitted and therefore, in the general case, does not give an exact value of the parameter. We found, however, that in many practical cases it is sufficiently accurate (see examples in the next section).

2. The particular choice of values $p_1$ and $p_2$ is determined by the expected range of the parameter $p$ in the data. In general, the smaller the range, the more accurate the parameter estimate.

[1] In fact, the method of nonlinear parameter estimation defined by eqs. (1) and (2) is based on the first order Taylor expansion of the model over the nonlinear parameter $p$:

$$M(A, p, E) = Af(p, E) = A[f(p_1, E) + (df/dp)(p_1, E)(p - p_1)]$$

and further approximation of the first derivative by $(df/dp)(p_1, E) = [f(p_2, E) - f(p_1, E)]/(p_2 - p_1)$. The choice of $\Psi$ described in the text corresponds to the minimization of the contribution of the second derivative $(d^2f)/(dp^2)$, with the second derivative being estimated using the function values at points $p_1$, $p_m$, $p_2$.

3. Since the coefficients $K_1$ and $K_2$ are determined through the data approximation by the *linear* model, these coefficients are *linear* combinations of the data, and an analytical expression for their calculation can be easily derived.

4. Fitting the measured spectrum with the linear model $L(K_1, K_2, E)$ (eq. [1]) must be done in "count space" to account for the energy response of the instrument. The $\chi^2$ statistic $[\![\chi^2 = \sum \{[C(E_i) - K_1 f(p_1, E_i) - K_2 f(p_2, E_i)]\}^2/w_i]\!]$ can be used to determine the best-fit values of $K_1$ and $K_2$. A modification of the $\chi^2$ technique, described in the next section, is fully applicable to moderate and low statistic observations and yields unbiased results.

5. A practical way to implement this method for mapping the parameter $p$ (e.g., temperature) distribution for an extended source is to calculate and store maps of $K_1$ and $K_2$. All further calculations can be performed by manipulating these $K_1$ and $K_2$ maps. The reconstruction of the parameter map on any coarser grid (or smoothed parameter map) reduces to regrouping (or smoothing) the $K_1$ and $K_2$ maps and the relatively simple application of equation (2).

6. The algorithm described above was developed and verified in the context of *ASCA* observations of relatively hot clusters with $kT \geq 2$ keV, and it should be used with caution in the case of lower temperatures, when the contribution from emission lines may be dominant.

7. This method is not intended to be a substitute for conventional spectral analysis. It is intended rather for relatively fast and easy construction of temperature maps and as a guide to further detailed analysis of regions of particular interest using conventional spectral fitting.

### 2.2. *Error Calculation*

One important aspect of any method of parameter estimation is the determination of confidence intervals. The expression for the parameter estimate given by equation (2) is, in fact, a ratio of two linear combinations of the observed counts (summation is over detector channels):

$$p = \frac{\sum \alpha_i C_i}{\sum \beta_i C_i}, \qquad (3)$$

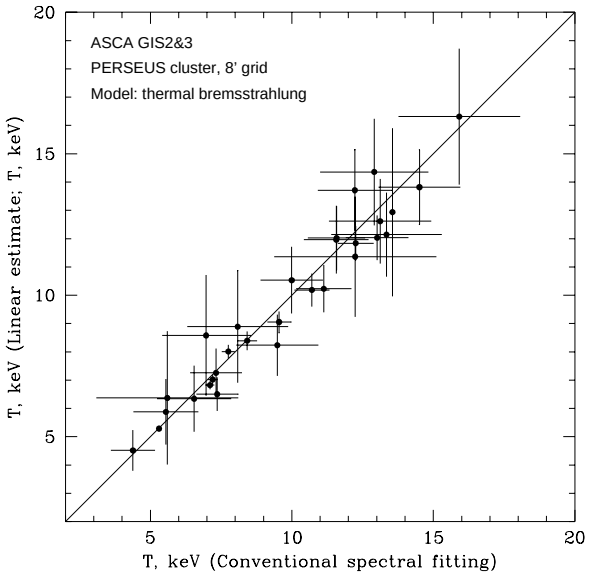


FIG. 5.—Application to *ASCA* GIS observations of the Perseus cluster (sequence Nos. 80007000, 80008000, and 80009000). Comparison with results of direct spectral fitting shows the excellent agreement between the two approaches. The spatial grid was centered on NGC 1275 and had a pixel size of 6′. For each cell the temperature has been determined using conventional spectral fitting in XSPEC and using the method described in § 2. Both for direct spectral fitting and linear estimation, the optically thin thermal bremsstrahlung emission model was used. The choice of reference spectra were $T_1 = 2$ keV, $T_2 = 12$ keV, $T_m = 7$ keV. Note that the data shown are not corrected for the effects produced by the telescope PSF and are shown to demonstrate that standard spectral fitting and the method described here yield equivalent results.

the numerator and denominator being, obviously, correlated. The coefficients $\alpha_i$, $\beta_i$ of these linear combinations can be explicitly expressed in terms of the known values and functions $f_1(p_1, E_i), f_2(p_2, E_i), p_1, p_2, \Psi$.

A simple and also sufficiently accurate way to calculate the confidence interval for $p$ is based on the fact that the confidence region for the pair of *linear* parameters ($q_1 = \sum \beta_i C_i$, $q_2 = \sum \alpha_i C_i$) can be easily determined as the interior of the ellipse defined in the $(q_1, q_2)$ plane by the equation

$$a(q_1 - q_1^m)^2 + b(q_2 - q_2^m)^2 + c(q_1 - q_1^m)(q_2 - q_2^m) = \frac{\Delta\chi^2}{2}, \quad (4)$$

where $q_1^m$ and $q_2^m$ are the measured values of $q_1$ and $q_2$ [i.e., $(q_2^m/q_1^m) = p^m$ is the estimated value of $p$] and the coefficients $a$, $b$, $c$ can be derived from $\langle q_1^2\rangle$, $\langle q_2^2\rangle$, $\langle q_1 q_2\rangle$. The confidence interval for $p = q_2/q_1$, corresponding to the given confidence region in the $(q_1, q_2)$ plane, is obviously restricted by the two tangents to the ellipse (eq. [4]) and can be found by solving the quadratic equation

$$\left[1 - \frac{\langle x_1^2\rangle}{(q_1^m)^2}\Delta\chi^2\right]\left(\frac{p}{p^m}\right)^2 - 2\left(1 - \frac{\langle x_1 x_2\rangle}{q_1^m q_2^m}\Delta\chi^2\right) \times \left(\frac{p}{p^m}\right) + \left[1 - \frac{\langle x_2^2\rangle}{(q_1^m)^2}\Delta\chi^2\right] = 0, \quad (5)$$

where $x_1 = q_1 - \langle q_1\rangle$, $x_2 = q_2 - \langle q_2\rangle$. Since we wish to determine the confidence interval for *one* parameter $p$, $\Delta\chi^2$ in equation (5) should be set to 1, 4, or 9 for 1, 2, or 3 $\sigma$ confidence regions.

Although $p$ is a nonlinear function of the linear parameters $q_1$ and $q_2$ and the above consideration is not exactly correct, as our simulations have demonstrated, the approach described above does provide sufficiently accurate estimates of the confidence intervals. The results of Monte Carlo simulations of the confidence interval determination are shown in Figure 2. In these simulations, we generated 10,000 spectra for each value of the total number of counts with $kT = 5$ keV (thermal bremsstrahlung model) in the 0.5–9.0 keV energy range and with channel width 0.01 keV. For each simulated spectrum, the temperature was estimated according to the algorithm described in the previous subsection with $T_1 = 2$ keV, $T_2 = 12$ keV, $T_m = 7$ keV (eq. [2]), and the 1 $\sigma$ and 2 $\sigma$ confidence intervals were estimated by solving equation (5) with $\langle x_1^2\rangle$, $\langle x_2^2\rangle$, $\langle x_1 x_2\rangle$ estimated from the *simulated* spectrum (i.e., $\langle x_1^2\rangle = \sum \alpha_i^2 C_i$, $\langle x_2^2\rangle = \sum \beta_i^2 C_i$, $\langle x_1 x_2\rangle = \sum \alpha_i \beta_i C_i$). It is clear from Figure 2 that this method of confidence limit determination works quite well even if the total number of counts in the spectrum is quite small, ~20–30 counts.

### 2.3. *Application to ASCA Data*

To verify the method described above we have chosen the energy response of the *ASCA* SIS0 CCD0 (s0c0g0234p40e1_512v0_8i.rmf in standard *ASCA* GOF notation). In Figure 3, the results of the application of the method to two emission models (optically thin thermal bremsstrahlung and thermal emission from an optically thin plasma with various heavy element abundances) are shown. For each emission model and each value of the temperature in the specified range we generated the expected count spectrum and determined the temperature using the method described above. The weights $w_i$ in the expression for $\chi^2$ (see previous paragraph) used for determination of the best-fit values of $K_1$ and $K_2$ were set to unity. (More realistic weights together with true simulations which account for Poisson statistics are considered in the next section.) As is seen from Figure 3, despite the quite large temperature range (2–12 keV), the accuracy of the temperature estimate is satisfactory for most practical cases, the relative error being less than 2%–4%. Figure 4 illustrates a more complicated situation where we are fitting the "wrong model"; we determine the temperature of a thermal optically thin plasma model by fitting a thermal bremsstrahlung spectrum. As this figure shows, the method still gives quite accurate parameter estimates with the relative error being below 3%–5% for most of the temperature range and for a wide range of heavy element abundances.

Finally, the scatter plot in Figure 5 compares our method with the results of conventional spectral fitting using the *ASCA* GIS observations of the Perseus cluster (sequence Nos. 80007000, 80008000, and 80009000). Each point corresponds to an individual cell (6′ on a side) located on the spatial grid centered on NGC 1275. The abscissa of each point is the temperature determined by direct spectral fitting using XSPEC, while the ordinate is the temperature determined using the method described above. The weights $w_i$ in the expression for $\chi^2$ for both methods were determined according to the algorithm described in the next section. Note that the data plotted in Figure 5 were not corrected for the effects induced by the energy-dependent *ASCA* PSF and are shown here only to demonstrate that standard spectral fitting and the method described here yield equivalent results.

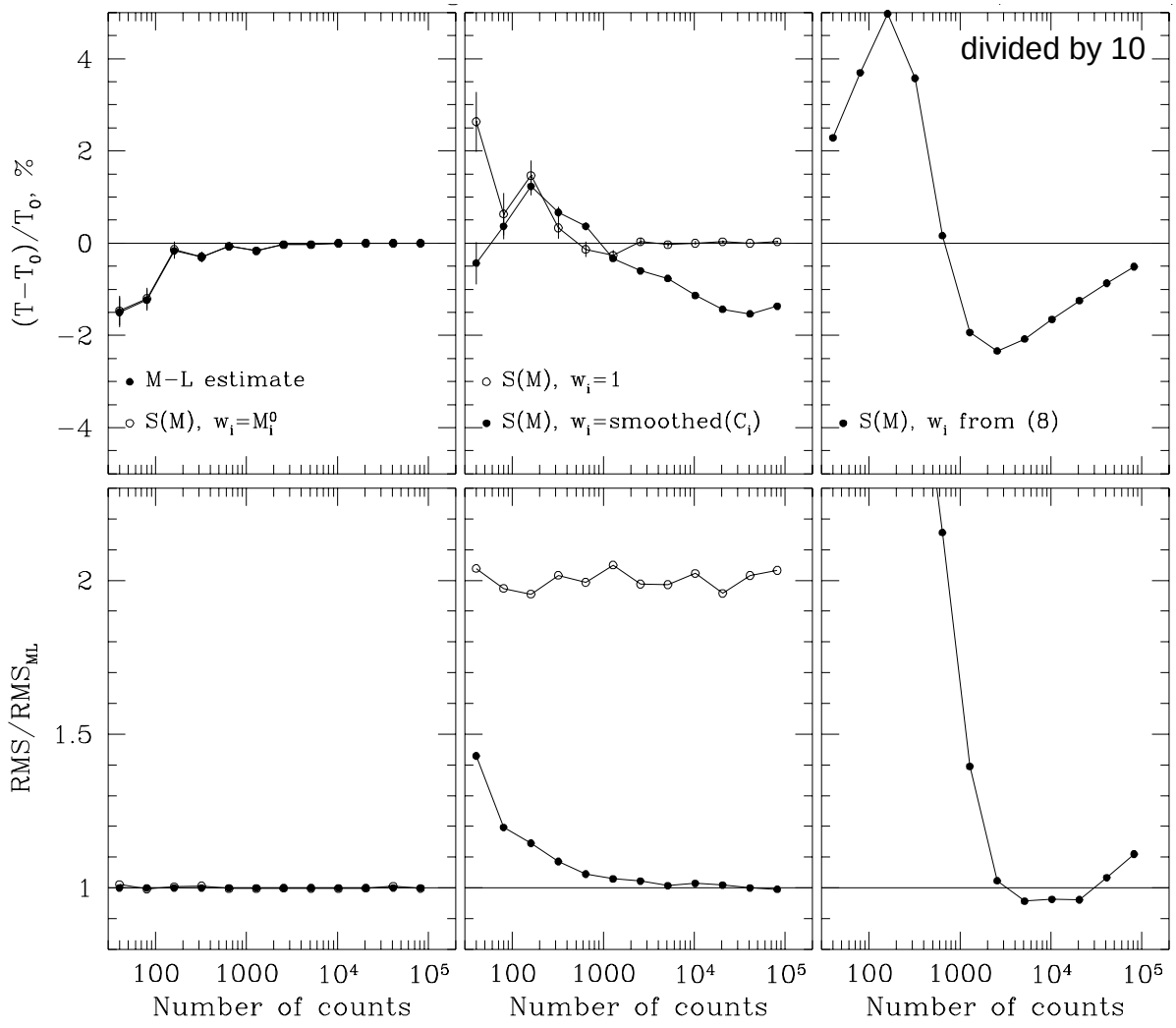


FIG. 6.—Comparison of various fit statistics (see text for detailed description of the statistics used). Monte Carlo simulations (9000 runs for each value of the total number of counts) for an optically thin thermal bremsstrahlung spectrum with $kT = 3$ keV, assuming an ideal instrument with a diagonal energy response, a 0.5–9.0 keV energy range, and a channel width of 0.01 keV. *Top panels*: relative bias (in percent) of temperature estimates for various statistics as a function of total number of counts in the spectrum. Note that the curve in the rightmost top panel is scaled down by a factor of 10. *Bottom panels*: ratio of dispersion of the parameter estimate with respect to that obtained in the maximum likelihood method for the same parameter estimation methods as in the corresponding top panels.

### 2.4. *Distinguishing Multi-Temperature from Single-Temperature Emission*

The rather good single-temperature fit to the two-temperature model (Fig. 1) raises another question: how long an exposure do we need to reliably distinguish between single-temperature and multi-temperature plasma emission? We have found that over the *ASCA* energy range, the difference between one- and two-temperature models starts to be significant at $\approx 10^5$ source counts. For example, the approximation of a SIS spectrum with $10^5$ counts (sum of 3 and 8 keV bremsstrahlung components with the same flux at 1 keV) by a single-temperature model in the 0.5–9 keV energy range (290 channels) gives an increase of $\chi^2$ due to an inadequate model of $\Delta\chi^2 \approx 13$. Adding emission lines from heavy elements (abundance 0.3 of the cosmic value) somewhat improves the sensitivity: the increase of $\chi^2$ (when fitting 2T meka simulation with 1T meka model) is $\Delta\chi^2 \approx 27$. Note that the value of $\Delta\chi^2$ scales linearly with the number of counts in the spectrum (e.g., for $3 \times 10^4$ counts $\Delta\chi^2$ is $\sim 4$ and 9, respectively). Note also that for as many as $\approx 10^5$ counts the calibration uncertainties might become important. Of course, if the temperatures of the two components differ very strongly (e.g., 1 and 10 keV) the situation becomes much better for the case considered above.

## 3. MODIFICATION OF THE $\chi^2$ FITTING TECHNIQUE

The aim of this section is to derive and verify by Monte Carlo simulations a simple and practical recipe for unbiased parameter estimation in X-ray spectral analysis when statistics are poor and only moderate or low numbers of counts are available. It is all too common that all or part of the observed spectrum has few or even zero counts per energy bin (especially in the high-energy bins) even when the total number of counts in the spectrum is quite large. In this case special attention must be paid to avoid biased parameter estimates. A similar problem has been recently considered by Primini & Kearns (1995) and Wheaton et al. (1995). In the discussions below, we present the arguments which are particularly relevant to *ASCA* data analysis. Although the justifications are not all presented in detail, they can be found in standard textbooks or are transparent.

In considerations throughout this section the variables

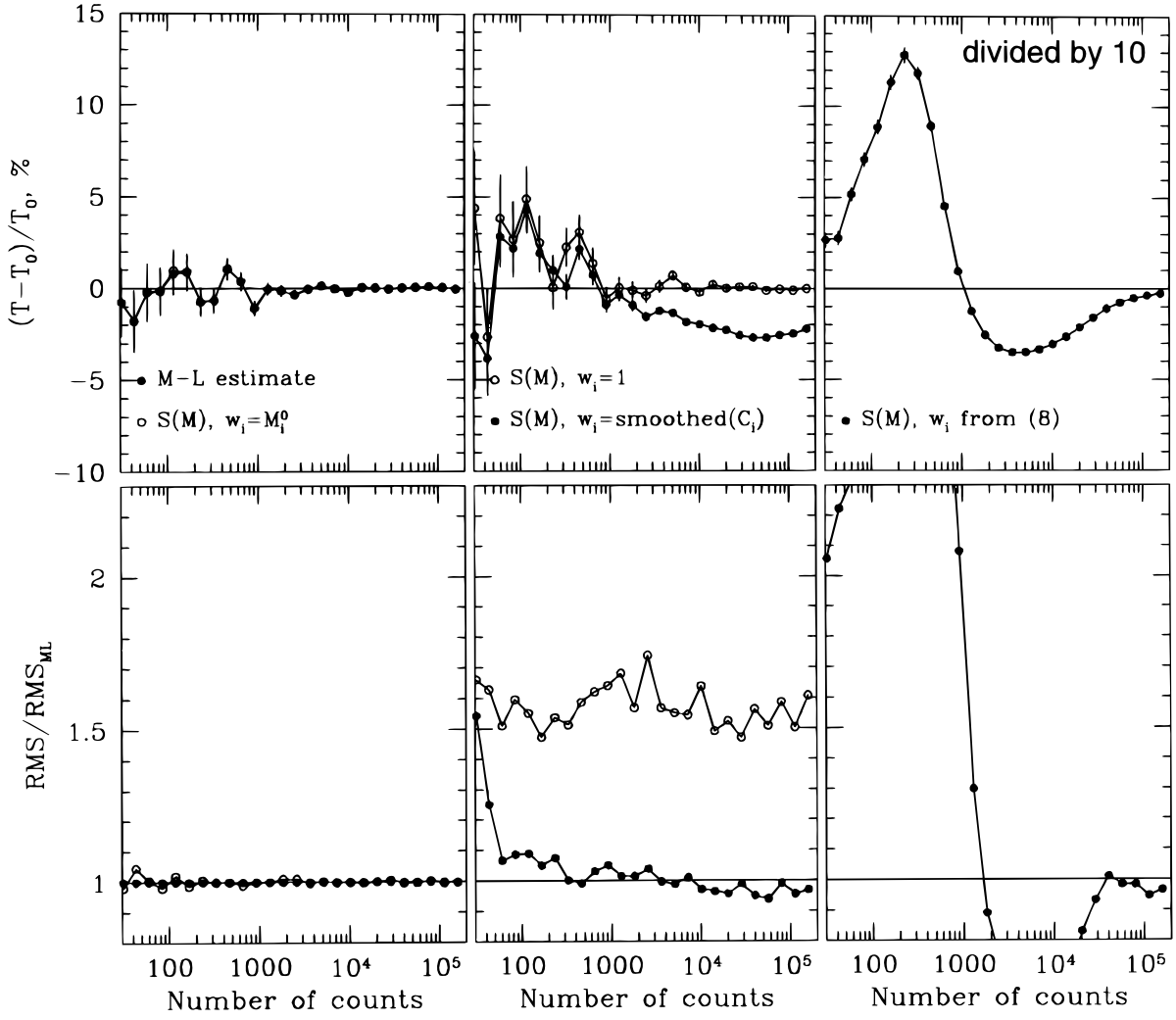


FIG. 7.—Same as Fig. 6 for a temperature of $T = 8$ keV (1000 Monte Carlo runs for each value of the total number of counts)

$M_i$, $C_i$, $M_i^0$ refer to the model, data, and true model (i.e., expectation value) counts in the $i$th spectral bin, $i = 1, N$; $p_k$ is $k$th parameter of the model, $k = 1, L$. Unless stated otherwise, we consider linear models $M_i(p) = \sum_k p_k F_i^k$. As we demonstrate through extensive Monte Carlo simulations (see below), the main conclusions formulated in this section for linear models hold for typical nonlinear models, commonly used in X-ray data analysis.

A reasonable set of requirements for any spectral fitting technique used in X-ray data analysis includes (in order of decreasing priority, as viewed by the authors):

*a*) The parameter estimates should be bias-free (at least for linear models);

*b*) The parameter estimates should be the most accurate (ideally, the method should provide a minimal dispersion of the best-fit parameters with respect to their true values);

*c*) The method should allow straight forward construction of confidence intervals for the parameters;

*d*) The method should provide a goodness-of-fit criterion.

Typically, spectral fitting packages search for a minimum of the function:

$$S(M) = \sum_i^N \frac{(M_i - C_i)^2}{w_i}, \tag{6}$$

where $w_i$ are the inverse weights assigned to the $i$th energy bin. For the linear model, the minimum of equation (6) can be explicitly found:

$$p_k = \sum_i^N \alpha_i^k C_i, \tag{7}$$

where $\alpha_i^k$ are functions of $F_j^l$, $l = 1, L$, and $w_j, j = 1, N$.

If all $C_i$ are Gaussian distributed with mean $M_i^0$ and standard deviation $\sigma_i$, then with $w_i = \sigma_i^2$ the value of $S(M^0)$ will follow a $\chi^2$ distribution with $N$ degrees of freedom. In practice, $C_i$ has either a Gaussian or Poisson distribution or is a linear combination of them (e.g., when background subtraction is performed). For Poisson-distributed $C_i$, one of the commonly used recipes for the determination of the weights $w_i$ in equation (6) is

$$w_i = \begin{cases} C_i & \text{if } C_i > 0, \\ 1 & \text{if } C_i = 0 \end{cases}. \tag{8}$$

As is well known, parameter estimation based on the above formula will be strongly biased (e.g., Cash 1979) if $M_i^0$ is small in even only a portion of the spectrum since higher weights (lower $w_i$) are assigned to the spectral bins where the number of counts is smaller due to statistical fluctuations.

Two commonly implemented solutions are the following:

1. Use of the maximum likelihood (ML) statistic (see, e.g.,

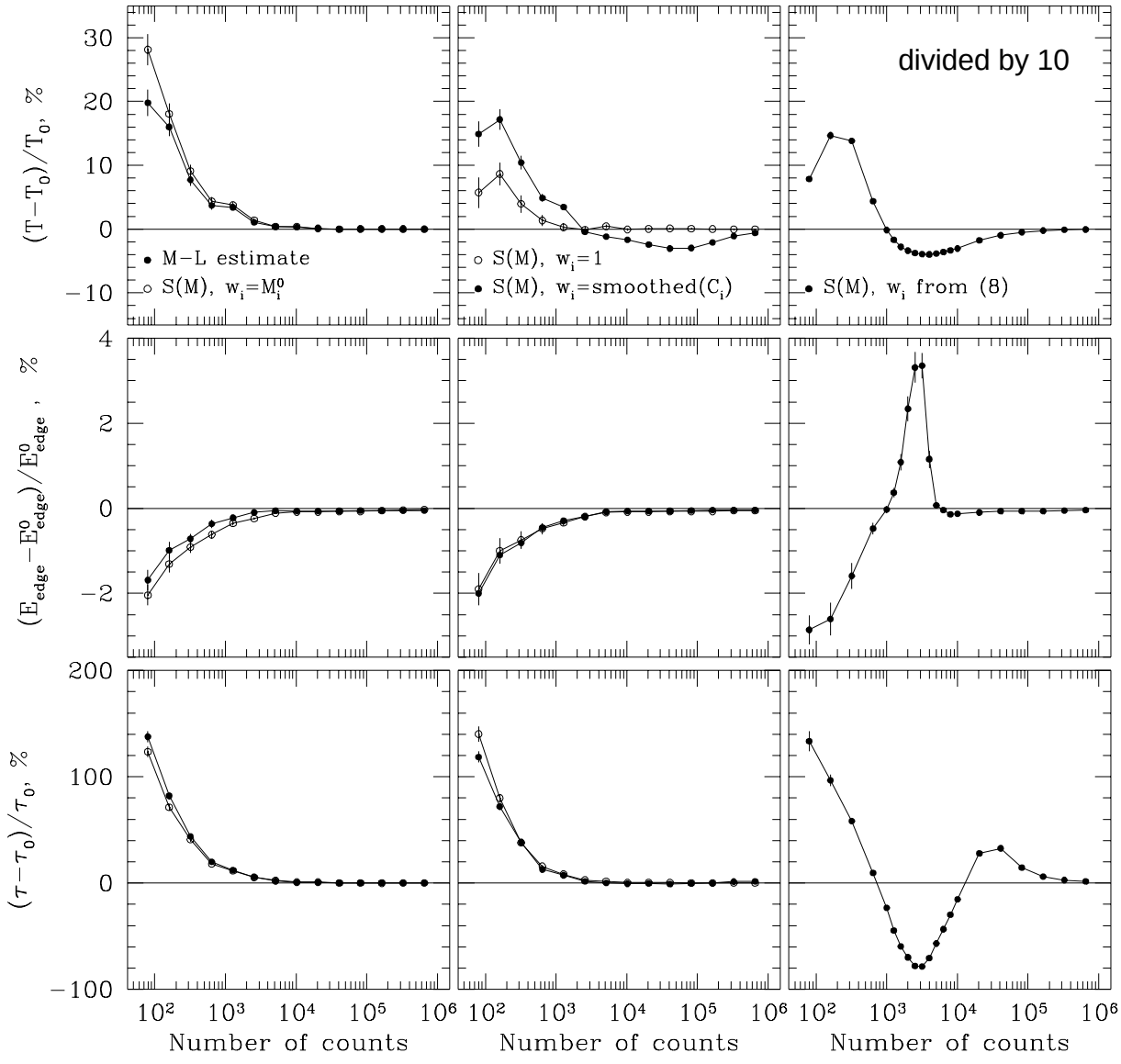


FIG. 8.—Comparison of various fit statistics (see text for detailed description of the statistics used). Monte Carlo simulations (1000 runs for each value of the total number of counts) for an optically thin thermal bremsstrahlung spectrum ($kT = 8$ keV) with a $K$-edge ($E_{edge} = 7.1$ keV, $\tau = 1$) absorption, assuming an ideal instrument with a diagonal energy response, 0.5–9.0 keV energy range, and channel width of 0.01 keV. Relative bias (in percent) of parameter determination for various statistics as a function of total number of counts in the spectrum is shown. *Top*, results for temperature; *middle*, results for edge energy; *bottom*, results for optical depth of the edge. Note that the curve in the rightmost top panel is scaled down by a factor of 10.

Cash 1979) instead of $S(M)$ defined by equation (6). The ML statistic satisfies requirements (*a*), (*b*), and (*c*) above, but fails with (*d*); it provides no goodness-of-fit test. The major difficulty is connected with its practical implementation with more complicated (than Poisson) statistical distributions of $C_i$ (e.g., especially for background subtracted spectra).[2]

2. Grouping the detector channels into broader energy bins until each bin contains more than $\sim 20$ counts followed by minimization of $S(M)$ with weights defined by equation (8). Although all requirements (*a*)–(*d*) above are satisfied, this recipe is not always acceptable, with the primary objection being an obvious loss of spectral information when significant regrouping is necessary.

[2] Note, however, that if background is known with "sufficient accuracy," one can fit measured counts (not background subtracted) with a model consisting of *the source model spectrum + the background spectrum.* In this case, the ML statistic can be easily applied, provided the measured counts follow a Poisson distribution.

Below we describe an alternative approach which, in addition to its simplicity, meets the requirements (*a*)–(*c*) listed above for many practical cases.

### 3.1. *Unbiased Estimates with Minimal Dispersion*

As can be easily demonstrated, there must be an absence of statistical correlation between the weights $w_i$ and the measured counts $C_i$ if a spectral fitting technique based on minimization of S(M) is to yield unbiased parameter estimates. Indeed, recalling that the $\alpha_i^k$ in equation (7) are functions of $F_j^l$ and $w_j$ only, if $w_i$ and $C_i$ are uncorrelated, we can write

$$\langle p_k \rangle = \left\langle \sum_i^N \alpha_i^k C_i \right\rangle = \sum_i^N \alpha_i^k \langle C_i \rangle = \sum_i^N \alpha_i^k M_i^0 = p_k^0 \quad (9)$$

since $\langle C_i \rangle \equiv M_i^0$ by definition. Therefore, an *arbitrary choice of $w_i$ which are uncorrelated with the fluctuations of the measured counts* (e.g., $w_i \equiv 1$ or any other function of $i$)

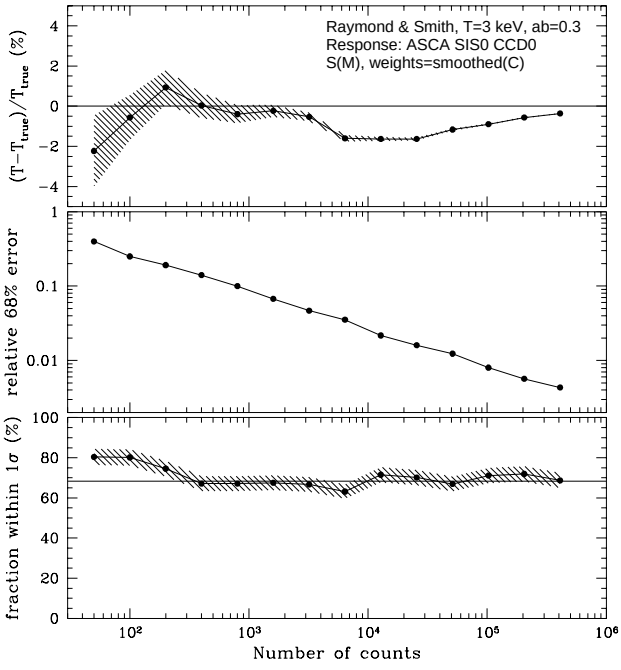


FIG. 9.—Verification of the method for determination of weights $w_i$ from the smoothed measured spectrum (as described in the text) for the *ASCA* SIS energy response. Monte Carlo simulations (500 runs for each value of the total number of counts) for emission from an optically thin plasma with $T = 3$ keV and a heavy element abundance of 0.3 (of the cosmic value). Both temperature and abundance were allowed to vary during minimization (see Fig. 10 for abundance results). *Top*: the relative bias in the temperature estimate. *Middle*: the fractional error (68% confidence) on the temperature. *Bottom*: the fraction of those runs for which the 68% confidence interval determined by XSPEC using the conventional algorithm ($\Delta\chi^2 = 1$) includes the mean value (shown in the top panel).

provides *unbiased estimates* of $p_k$. The definitions of weights according to equation (8) obviously does not satisfy this requirement; the $w_i$ and therefore $\alpha_i^k$ in equations (7) and (9) are strongly correlated with the fluctuations in $C_i$ for small values of $C_i$.

The *accuracy of the parameter estimates*, i.e., the dispersion of the $p_k$ estimates with respect to $p_k^0$, depends critically, however, upon the choice of $w_i$. It can be easily demonstrated that to achieve a minimal dispersion in the estimates for $p_k$ for any distribution of $C_i$ having variance $v_i$, the weights should be chosen as $w_i = v_i$.[3]

In the simplest case where $C_i$ follows a Poisson distribution, the variance of $C_i$ is equal to $M_i^0$, and the best choice for weights in equation (6) is $w_i = M_i^0$. Unfortunately, the $M_i^0$ are unknown, but a reasonable approximation to them can be made from the observed spectrum.[4] One possible solution based on the iterative determination of $w_i$ has been considered recently (Primini & Kearns 1995).

We propose another practical recipe for determining the weights $w_i$, which is to smooth the observed spectrum with a sliding window having a variable size which is adjusted to have more than $C_{\min}$ counts within the window. The weights $w_i$ can then be set equal to the values of the smoothed spectrum. Note that the smoothing of the spectrum is performed only to calculate the weights $w_i$, while the fitting procedure is applied to the original unsmoothed spectrum. As the Monte Carlo simulations described below confirm, a choice of $C_{\min} = 30$ counts gives satisfactory results by sufficiently suppressing the correlation between $w_i$ and $C_i$ and by simultaneously providing a good approximation to $M_i^0$.

Extending this algorithm to spectral fitting where background has been subtracted is straightforward. If the spectrum to be fitted is $C_i = C_i^{\rm tot} - C_i^{\rm bkg}$, then the weights for equation (6) should be chosen as $w_i = w_i^{\rm tot} + w_i^{\rm bkg}$, where both $w_i^{\rm tot}$ and $w_i^{\rm bkg}$ are calculated according to the algorithm described above.

For linear models, the above algorithm guarantees an almost unbiased estimation of parameters (the remaining bias being of the same order as for regrouping the original spectrum into broader channels with more than $C_{\min}$ counts in each channel). In addition, as long as the number of counts in the observed spectrum is large enough, so that the smoothed spectrum roughly reproduces the shape of the true spectrum, the above algorithm also provides minimal error estimates of the parameters and correct determination of the confidence intervals. In practice and as we show below, these same properties are found when the approach described above is applied to those typical nonlinear models commonly encountered in X-ray spectral analysis.

### 3.2. *Error Estimation and Goodness of Fit*

Let us now investigate the properties of confidence intervals determined from conventional parameter estimation as applied to the new method described above. We refer to conventional confidence interval determination as that given by $S_{\min} + \Delta$, where $\Delta$ is defined such that for $n$ parameters of interest $\int_0^\Delta \chi_n^2(p)dp$ yields the desired confidence level (e.g., 0.67, 0.95). We investigate the distribution of $\Delta S = S(M^0) - S(M^{\min})$, where $S(M^{\min})$ is the minimum value of the function. Without reducing the generality, we can rewrite the model as a function of another set of parameters $M_i(p_k, k = 1, L) = \sum_k^L \tilde{p}_k \tilde{F}_i^k w_i$, where $\tilde{F}_i^k$ are an orthonormal set of vectors. The value of $\Delta S$, of course, remains the same. Now it is easy to find an explicit expression for $\Delta S$:

$$\Delta S = S(M^0) - S(M) = \sum_k^L (\Delta\tilde{p}_k)^2 , \qquad (10)$$

where $\Delta\tilde{p}_k = \tilde{p}_k^{\min} - \tilde{p}_k^0$ are the deviations from the true parameter values. If $C_i$ are Gaussian distributed and $w_i \equiv M_i^0$, then all $\Delta\tilde{p}_k$ have normal distributions, and therefore $\Delta S$ is distributed as $\chi^2$ with $L \equiv$ *number of fitted parameters* degrees of freedom. The expression of $\Delta\tilde{p}_k$ as a function of $\Delta C_i$ remains the same for either a Gaussian or Poisson distribution. Therefore, $\Delta S$ will be distributed as $\chi^2$ as long as $\Delta\tilde{p}_k$ is normally distributed. But $\Delta\tilde{p}_k$ are linear functions of $\Delta C_i$, and, even if each individual $\Delta C_i$ has a Poisson distribution, the sum of many $\Delta C_i$ with almost the same weight may have a Gaussian distribution. Thus, we can intuitively state that $\Delta S$ *is distributed as* $\chi_L^2$ *as long as the total number of photons in those channels which contribute significantly to* $\Delta\tilde{p}_k$ *is much larger than unity*. Our simulations (see below) have shown that for many cases about 100–200 source photons in the *ASCA* energy band are suffi-

[3] Indeed, requiring that the weights $w_i$ maximize the change in $S(M) - S(M^0)$ for any deviation of the parameter $p_k$ from its true value $p_k^0$, compared to the root mean square fluctuations of $S(M) - S(M^0)$, one immediately finds $w_i = v_i$.

[4] Note that if there is a priori knowledge of the spectral shape one can generate $w_i$ without recourse to the observed counts. For example, for the Crab Nebula observations with *ASCA*, choosing $w_i$ proportional to a power-law distribution with photon index $\Gamma \approx 2$ and convolved with the *ASCA* energy response, would provide an acceptable estimate of $M_i^0$. As one can see from eq. (6), to satisfy requirements (*a*) and (*b*) above, the absolute normalization of $w_i$ is not important.

cient to yield correct error estimates when derived with standard procedures.

It is easy to show that the mean value of $S(M^{\min})$ will be $N - L$, i.e., the same either for Gaussian or Poisson distributions of $C_i$ if $w_i \approx M_i^0$. In practice, for spectra sharply decreasing at high energies and with very low background, if $w_i$ is derived from the smoothed observed spectrum, then $M_i^0$ may be overestimated significantly at high energies and the average value of $S(M^{\min})$ can be artificially reduced. The width of the distribution of $S(M^{\min})$ also can be narrower than that of $\chi^2_{N-L}$. Therefore, *it is not possible to use the absolute value of* $S(M^{\min})$ *as a reliable goodness of fit estimator*. Note, however, that the distribution of $S(M^{\min})$ will remain much closer to $\chi^2_{N-L}$ and will at least retain the correct mean value compared with that calculated using the conventional choice of $w_i$ according to equation (8).

### 3.3. *Monte Carlo Simulations*

We performed a series of Monte Carlo simulations to compare various choices of weights $w_i$ in equation (6) and to verify the validity of the algorithm for the determination of $w_i$ proposed above. The following statistics were tested for total numbers of counts varying from 10 to $10^6$:

*a*) Maximum likelihood, i.e., $\sum_i^N M_i - \sum_i^N C_i \ln M_i$;

*b*) $S(M)$ given by equation (6), with $w_i \equiv M_i^0$;

*c*) $S(M)$ given by equation (6), with $w_i$ equal to the observed spectrum smoothed with a sliding window the size of which is adjusted to contain more than $C_{\min} = 30$ counts;

*d*) $S(M)$ given by equation (6), with $w_i \equiv 1$;

*e*) $S(M)$ given by equation (6), with $w_i$ defined by equation (8).

The results for a thermal bremsstrahlung spectrum with $T = 3$ and 8 keV (assuming an ideal instrument with a diagonal energy response) are shown in Figures 6 and 7. One can see that methods (*a*)–(*d*) give nearly unbiased parameter estimates with counts as few as $\sim$50–100, while method (*e*) produces a strong bias even when the total number of counts is large ($\sim 10^4$) and only converges to the true value when the number of counts approach $\sim 10^5$ counts. As expected, methods (*a*) and (*b*) give the most accurate parameter estimates, method (*c*) gives only slightly larger errors, and method (*d*), although unbiased, results in significantly larger dispersion of the best-fit parameter.

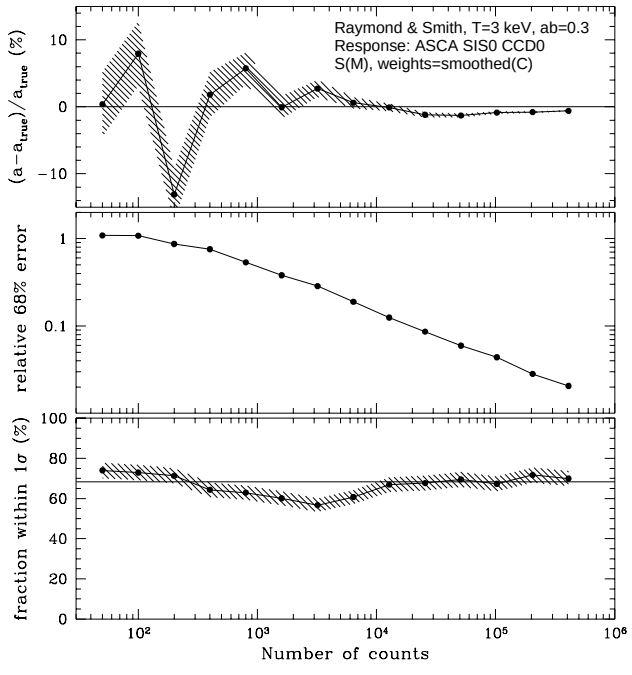


FIG. 10.—Same as Fig. 9 for the heavy element abundance

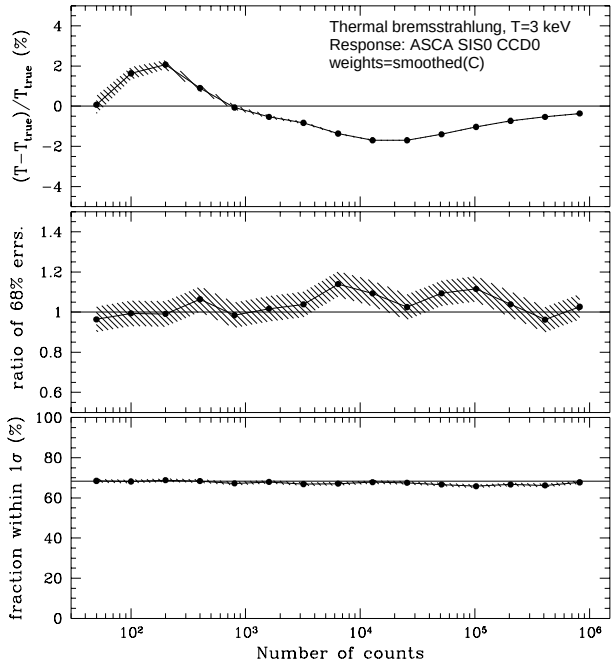


FIG. 11.—Verification of the method for temperature determination, described in § 2 for the *ASCA* SIS energy response. Monte Carlo simulations (9000 runs for each value of the total number of counts) for optically thin thermal bremsstrahlung spectrum with $kT = 3$ keV. *Top*: the relative bias in the temperature estimate. *Middle*: the ratio of the 68% error for the temperature determination in this method to that for conventional spectral analysis. *Bottom*: the fraction of runs for which the 68% confidence interval determined according to eq. (5) includes the mean value (shown in the top panel).

Figure 8 shows the results of simulations for a bremsstrahlung spectrum ($T = 8$ keV) including $K$-edge absorption ($E_{\rm edge} = 7.1$ keV, $\tau_{\rm edge} = 1$). All four parameters of the model were allowed to vary during minimization. The extremely nonlinear nature of the model (due to the absorption component) leads to a strong bias in parameter estimation by all five methods even with moderate numbers of counts. Nevertheless, method (*c*) is not significantly more biased than (*a*) or (*b*) and is apparently less biased than (*e*).

Figures 9 and 10 demonstrate the validity of method (*c*) for determining weights for spectral fitting as applied to the *ASCA* SIS (s0c0g0234p40e1_512v0_8i.rmf in standard *ASCA* GOF notation) for emission from an optically thin plasma. In the Monte Carlo simulations the best-fit parameters and conventional confidence interval determination for each simulated spectrum were made using XSPEC with $w_i$ substituted for the errors. Displayed in the upper panel of each graph is the relative bias of the parameter estimate versus total number of counts. The middle panel shows the fractional 68% error determined as a fractional half-width $\Delta p$ of the interval $[\langle p\rangle - \Delta p, \langle p\rangle + \Delta p]$, symmetric with respect to the mean $\langle p\rangle$ and including 68% of the best-fit values obtained in individual Monte Carlo runs. The bottom panel illustrates the applicability of the conventional method for confidence interval determination; the frac-

tion of the runs in which the 68% confidence interval determined by XSPEC using the conventional method (with $\Delta\chi^2 = 1$) includes the mean parameter value.

Finally, Figure 11 shows the results of Monte Carlo simulations for the method of temperature determination described in § 2 with weights $w_i$ (used for determination of coefficients $K_1$ and $K_2$) calculated according to method (*c*). The energy response and energy channels were those of the *ASCA* SIS CCD0. The top panel is the same as in Figure 9. The middle panel shows the ratio of 68% error of the temperature determination in this method to that of the conventional spectral analysis (i.e., the ratio of *actual* widths of 68% intervals, determined as described in the previous paragraph). The bottom panel demonstrates the validity of the method for determination of confidence intervals from equation (5), the fraction of the runs in which the 68% confidence interval includes the mean value of the parameter. Clearly, systematic biases are no larger than those for conventional spectral analysis, and the accuracy of the temperature determinations are also similar to conventional techniques.

## 4. CONCLUSIONS

The advent of powerful X-ray observatories capable of obtaining high-quality spatially resolved spectral data have necessitated the development of new approaches to data analysis. We have presented a novel approach to the mapping of the spectral properties of extended sources. This approach employs two features; the use of a linear combination of two spectra to characterize the spectrum at any desired spatial position and the approximation of the inverse weight, in the standard $\chi^2$ minimization by the smoothed observational data.

The first feature of our approach, the determination of a spectral parameter from the coefficients of the linear combination of two spectra which bound the expected range of the parameter, has two important characteristics. First, because it is a linear process, it obviates the need for time-consuming, iterative fitting typical of spectral analysis in X-ray astronomy. Therefore, it is rapid and allows the production of spectral parameter maps on fine spatial scales. Second, because the process is linear, it permits, through simple calculations, the generation of confidence levels.

The second feature of our approach, the use of the smoothed data to approximate the inverse weights in the standard $\chi^2$ expression, is important for the optimal use of the data. This use of the smoothed data reduces the bias inherent in low count situations where the weights would otherwise be correlated with the data. Furthermore, as the simulations show, even for very sparse data the parameter estimates have very small bias and the parameter estimates are nearly optimal, i.e., the uncertainties on the fitted parameters are as small as the statistics permit.

This technique was developed for application to *ASCA* observations of clusters of galaxies. In particular, we are interested in studying the temperature distributions in clusters. For regular clusters, a few judiciously selected spatial regions can be analyzed in detail. For irregular systems, which are very common and in which we are particularly interested, the approach described above provides an overall view of the cluster and is a guide to the selection of spatial regions for more detailed analysis. We report on the results of our analysis of *ASCA* observations of extended X-ray emission from clusters of galaxies in subsequent papers.

E. C. and M. G. acknowledge support from RBRF grant 93-02-17167. W. F. and C. J. acknowledge support from the Smithsonian Institution and NASA contract NAS 8-39073. E. C. and M. G. also acknowledge hospitality of the Harvard-Smithsonian Center for Astrophysics where this work was started.

## REFERENCES

Cash, W. 1979, ApJ, 228, 939
Primini, F., & Kearns, K. 1995, ApJ, submitted
Tanaka, Y., et al. 1994, PASJ, 46, L37
Wheaton, Wm. A., et al. 1982, ApJ, 438, 322