The photon clean method: an event-based approach to analyzing X-ray spectra[1]

Matthew H. Carpenter, J. Garrett Jernigan, Peter Beiersdorfer, and Gregory V. Brown

Abstract: The Photon Clean Method (PCM) is an inverse Monte-Carlo method of spectral fitting that differs from traditional fitting routines found in spectral modeling packages by fitting event lists as opposed to binned spectra. The model spectrum is represented in event form as well. Thus, using this method it is possible to fit data of higher dimensionality than can be fit using binned spectra and standard routines based on Chi-Square statistics, such as event-mode data from electron beam ion traps or satellite observations that are tagged, for example, as a function of time, position, or energy. To demonstrate some of the power of the PCM and aid in its development, we have implemented a simplified one-dimensional version of the PCM algorithm (PCM1D). Using our implementation, which is a command-line program intended for public release, we have performed tests on simulated and observed Chandra ACIS CCD data, and present two examples, one on Cassiopeia A and another on a simulated multitemperature plasma in collisional ionization equilibrium.

PACS No.: 52.65.Pp

Résumé : La méthode Photon Clean (PCM) est une méthode de Monte-Carlo inverse d'ajustement numérique spectral qui diffère des programmes traditionnels de modélisation en ajustant des listes d'événements plutôt que des spectres regroupés en classes (binned). Le spectre modèle est également représenté sous la forme événement. Ainsi, cette méthode permet un ajustement à des données de plus grande dimension que ce qui peut être ajusté en utilisant des spectres classés ou des programmes standard basées sur la statistique des moindres carrés, telles les données en mode événement provenant de pièges ioniques ou d'observations par satellite qui sont identifiées, par exemple, par le temps, la position ou l'énergie. Afin d'illustrer la puissance de la PCM et aider à son développement, nous utilisons une version à une dimension de l'algorithme PCM (PCM1D). Il s'agit d'un programme ouvert avec commandes que nous avons utilisé sur des données réelles et simulées ACIS CCD de Chandra et nous présentons deux exemples, un de Cassiopée et l'autre une similation d'un plasma à plusieurs température dans un équilibre d'ionisation par collisions.

[Traduit par la Rédaction]

1. Introduction

There are many tools freely available to perform spectral analysis of laboratory or astrophysical data. The standard practice in plasma modeling involves building synthetic spectra from one or more models, each with one or more parameters. Goodness of fit is usually determined by the Chi-Square test, in which a minimum number of parameters as practicable are desired. The Photon Clean Method (PCM), first described by Jernigan and Vezie [1], approaches this problem from a fundamentally different direction. Instead of normalized spectral bins, it compares discrete lists of photon events. This quantized treatment opens up several possibilities, including the use of Monte-Carlo (MC) methods, complex detector geometries, and models with numbers of degrees of freedom proportional to the number of photons. While it suffers the same problems of energy and spatial resolution imposed on standard methods, its reliance on random MC trials with feedback coupled with the use of bootstrap methods to determine sample variability allows complex solution spaces to be evaluated.

There are other analysis methods related to PCM that also derive models from event data using MC approaches. These related methods are all inverse MC approaches that find solutions in one or more dimensions via iteration with feedback as first described in Jernigan and Vezie 1996. Peterson developed a variation of this type of algorithm that fits models of a few free parameters based on a binned Chi-Square metric for feedback, in contrast to the PCM, which is based on the Kolmogorov–Smirnov (KS) metric. In this version, all the model photons and simulated detected events are produced by a detailed MC simulation of both the model and the detector system [2, 3]. Subsequently, Peterson extended the algorithm to allow many parameters in the form of multidimensional "blobs", each an aggregate of photons characterized by a set of a few parameters. The feedback method in this approach is based on a Monte-Carlo Markov Chain (MCMC; see ref. 4). Gu has also implemented a similar one-dimensional version of this MCMC approach [5]. Each of these inverse MC methods also includes a particular

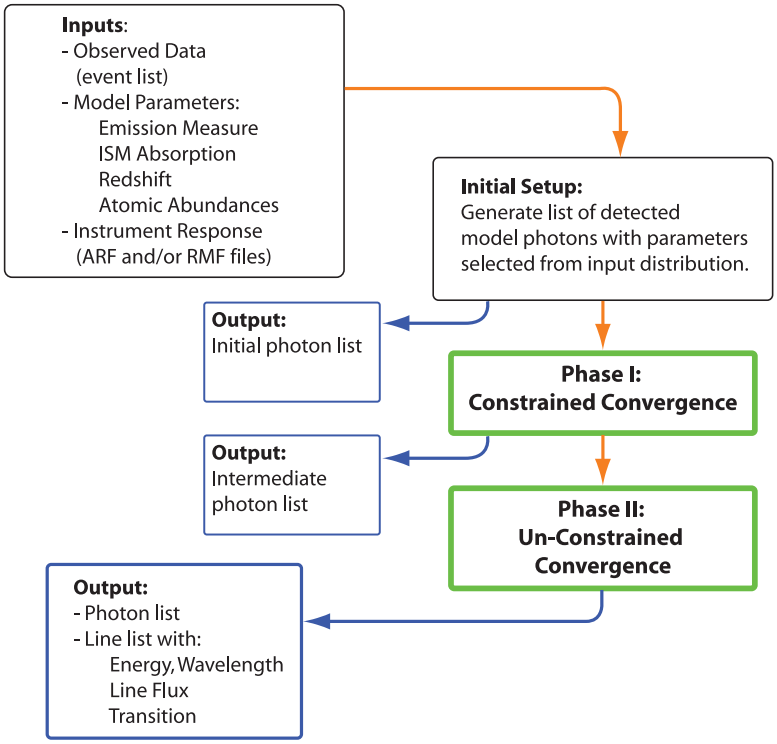
Received 7 March 2007. Accepted 16 May 2007. Published on the NRC Research Press Web site at http://cjp.nrc.ca/ on 9 February 2008.

M.H. Carpenter[2] and J.G. Jernigan. Space Sciences Laboratory, University of California, Berkeley, CA 94720-07450, USA.
P. Beiersdorfer and G.V. Brown. High Temperature and Astrophysics Division, Lawrence Livermore National Laboratory, Livermore, CA 94550-9234, USA.

[1]Paper given at the Workshop on Twenty Years of Spectroscopy with EBIT held in Berkeley, California, 13–15 November 2006.
[2]Corresponding author (e-mail: carpenter@ssl.berkeley.edu).

 doi: 10.1139/P07-153

Fig. 1. Flow-chart for one-dimensional Photon Clean Method with the Atomic Plasma Emission Database model.



methodology for the estimation of errors in the derived models. This entire class of methods has significant advantages over traditional methods in situations in which complex models and complex detectors require accurate MC techniques typical of the current generation of imaging spectrometers.

We have implemented a trial version of the PCM algorithm as an ANSI C command line program named PCM1D. This test version will be released publicly under open-source license in 2007 [6]. The PCM algorithm may be generalized to high dimensionality; however, this current manifestation operates in one dimension in both data and parameter space. In many ways this is sufficient to address analysis issues associated with, for example, event-mode data generated at the University of California Lawrence Livermore National Laboratory electron beam ion traps. This includes spectral data of K-shell and L-shell iron emission recorded with the XRS microcalorimeter on the Electron Beam Ion Trap I (EBIT-I) device using a simulated Maxwell–Boltzmann electron distribution [7, 8]. Since this version is intended for eventual general release with a focus on the astrophysics community, special effort has been made to make the program inter-operate with NASA HEASARC[3] tools via use of standard-format input and output products. In the examples presented here, the model is limited to collisional ionization equilibrium (CIE) plasma models constructed from the publicly available Atomic Plasma Emission Database (APED) developed at the Chandra X-Ray Center [9].

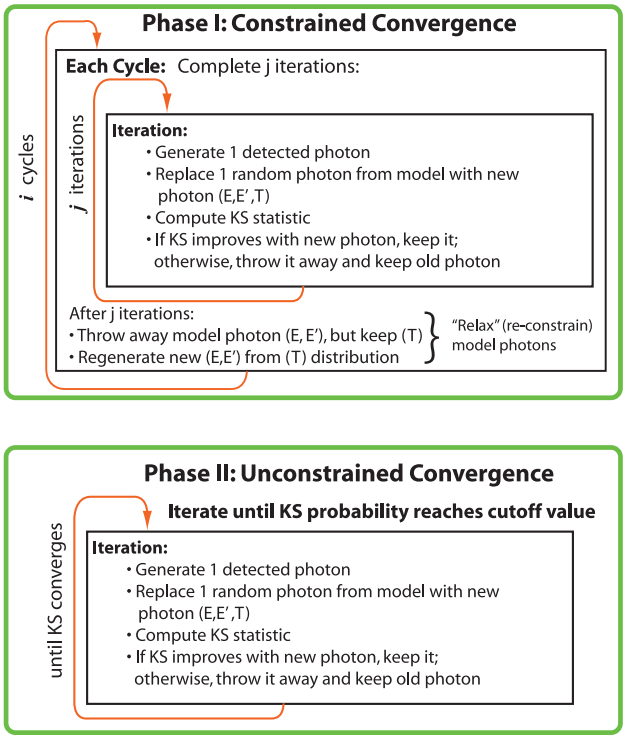
[3] High Energy Astrophysics Science Archive Research Center. See http://heasarc.gsfc.nasa.gov/

2. Method

The main departure from traditional analysis methods is that PCM operates directly on lists of photons without creating any bins or histograms. Observed events are loaded directly into an internal array, and model data are generated as a list of photons with associated model parameters. As each simulated photon is generated with a discrete set of model parameters, it is regarded as a parameter itself, or one member of the span of the parameter space. Goodness-of-fit statistics are provided by the two-sample Kolmogorov–Smirnov (KS) test [10] (see also Jernigan and Vezie [1]). The KS test is both well-suited to comparing event-mode data and computationally inexpensive. As with any Monte-Carlo-based method, the speed of individual components of the algorithm, as well as their scaling with size of the data set, is crucial to the practicality of the method.

The Photon Clean Method makes an important distinction between "observed events" and "real photons". To compute discrete photons from a variety of models, the program implements a Monte-Carlo "photon generator". The flow chart displaying the various steps, including inputs and outputs, is shown in Fig. 1. Given a fixed set of model parameters as inputs, the generator outputs a detected photon; to do this, the generator may need to produce and discard many undetected photons. The probability that a photon is detected as a discrete event is determined by the instrument model provided by instrumental response files (i.e., standard detection efficiency and point-spread function matrices, given by the so-called Ancillary Response File (ARF) and Redistribution Matrix File (RMF) files in the HEASARC tool sets). Due to the nature of the APED data base, it is known, exactly, which atomic transition produced a given photon, or whether it was emitted at part of the thermal

Fig. 2. Detail of Phase I and II steps. The photon parameters, energy (E), detected energy (E'), and temperature (T), are specific to the analysis presented here using an Atomic Plasma Emission Database collisional ionization equilibrium plasma; in general, any set of parameters may be used to describe the photon model source.

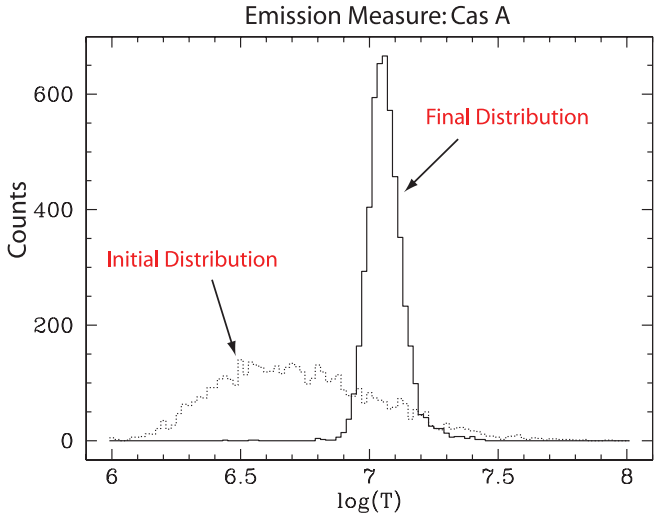


continuum or background. This information is retained by the program for the life of the photon, so that each photon is directly linked to a set of model parameters. This coupling allows the photon to be treated as an independent concept with a specific energy selected from a model distribution determined by a CIE plasma of a selected temperature. The variable parameter in this analysis is plasma temperature. In general, there could be more than one variable parameter, which is distinct for each emitted X-ray.

The functional process may be divided into two main cycles. After generating an initial model photon list from input parameters, the program enters the first iterative cycle, labeled in Fig. 1 as "Phase I". In this phase, the program works primarily on improving the parameter space of the model. Its goal is to produce the closest match to the observed data possible, given the limited set of model parameters to vary. "Phase II" in Fig. 1 breaks strict adherence to the model to produce the most accurate spectral distribution possible. While this final solution may not be reproducible from its parameters, it contains information that can be of great help to the analysis process, such as quantitative measurements of line ratios and fluxes.

Both phases have the same Monte-Carlo iteration and feedback loop at their core. For each iteration, an arbitrary photon is selected from the model list and set aside, then replaced with a new photon from the photon generator. The parameters used to generate the new photon may vary from completely random parameters to a linear combination of parameters taken from the existing model photon list. The program computes the KS parameter, and keeps the new photon if the KS value is the same

Fig. 3. Histogram of photon temperatures from the Photon Clean Method fit to knot in Cassiopeia A (see Fig. 6). The continuous line depicts the final distribution of temperatures, and the dotted line depicts the initial distribution of temperatures. The distribution of temperatures may be taken as the emission measure distribution of the fit.



or better than it was before the change. If the KS value has degraded, it throws away the new photon and restores the original photon set aside earlier.

During phase I (see Fig. 2), after a set number of iterations defining one cycle, the program throws away the entire list of model photon energies, but leaves the photon parameters intact. A new list of photons is generated from the old photons' parameters, and the cycle starts again. This "relaxation" step is needed because the algorithm is most likely to keep a photon with very different parameters than the one it replaces when the rest of the population is well-constrained by the model parameters. By "re-constraining" the population in this manner every so often, the algorithm is able to work most effectively to correctly change the distribution in parameter space. If the set of models cannot produce a realizable solution, phase I will not be able to arrive at a solution of sensible KS probability. Phase II takes over where phase I leaves off, and iterates until it produces a solution of highest KS probability. As noted earlier, without re-constraining the model population, phase II diverges from a physically realizable model, at the benefit of fine-tuning the discrepancies.

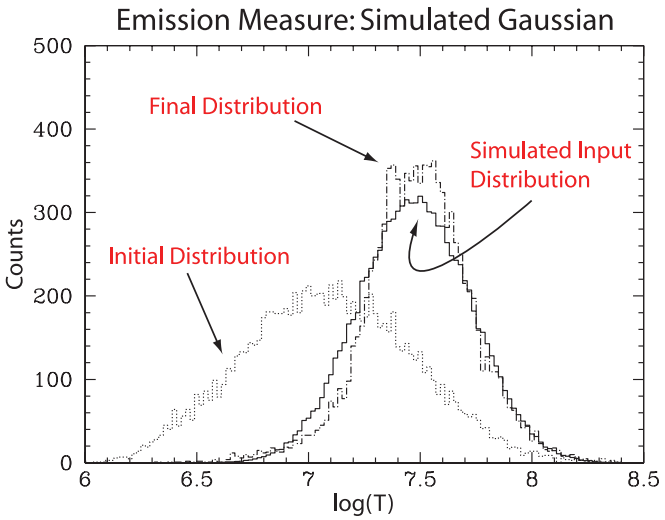
At the end of the fitting process, PCM outputs the model photon list with accompanying photon parameters. Emission measure diagrams in Figs. 3 and 4 are histograms of the temperature parameters of the photons in this list. Note that the histograms shown here are for visualization of results only and are not computed as part of the PCM algorithm. PCM also aggregates photons by element and transition, normalizing the counts in a given group by the ARF file to generate a line list with fluxes. Table 1 is copied directly from the line list output by the program. The program may also produce outputs at regular intervals during the process, so that plots such as the three panels in Fig. 5 are possible.

The natural way for PCM to estimate errors and variance in fit is using "bootstrap" re-sampling methods. This approach, introduced by Efron in 1979 [11], is a fully-developed branch

Table 1. Ten brightest emission lines, Cassiopeia A analysis. This list corresponds to Phase II solution (see Fig. 2). The complete model derived from the Atomic Plasma Emission Database (APED) includes thousands of X-ray lines. For illustration, only the ten brightest lines are listed here. The element, ionization, and atomic transition are derived directly from APED. Both the energy and wavelength of the lines are listed, along with computed flux in photon and energy units.

Element	Ion	Transition	E (keV)	λ(Å)	Flux (photons (s/cm^2)$^{-1}$)	Flux (erg (s/cm^2)$^{-1}$)
Si	XIII	$1s2p\ ^1P_1 \rightarrow 1s^2\ ^1S_0$	1.865	6.648	6.955×10^{-7}	2.045×10^{-15}
Mg	XII	$2p\ ^2P_{3/2} \rightarrow 1s\ ^2S_{1/2}$	1.473	8.419	2.982×10^{-7}	7.039×10^{-16}
Si	XIII	$1s2s\ ^3S_1 \rightarrow 1s^2\ ^1S_0$	1.839	6.740	2.913×10^{-7}	8.584×10^{-16}
Si	XIV	$2p\ ^2P_{3/2} \rightarrow 1s\ ^2S_{1/2}$	2.006	6.180	2.822×10^{-7}	9.070×10^{-16}
S	XV	$1s2p\ ^1P_1 \rightarrow 1s^2\ ^1S_0$	2.461	5.039	2.719×10^{-7}	1.072×10^{-15}
Fe	XXI	$1s^22s^22p3d\ ^3D_1 \rightarrow 1s^22s^22p^2\ ^3P_0$	1.009	12.284	2.695×10^{-7}	4.356×10^{-16}
Fe	XXII	$1s^22s^23d\ ^2D_{3/2} \rightarrow 1s^22s^22p\ ^2P_{1/2}$	1.053	11.770	1.947×10^{-7}	3.286×10^{-16}
Fe	XXI	$1s^22s^22p4d\ ^3P_1 \rightarrow 1s^22s^22p^2\ ^3P_0$	1.308	9.480	1.844×10^{-7}	3.863×10^{-16}
Si	XIV	$2p\ ^2P_{1/2} \rightarrow 1s\ ^2S_{1/2}$	2.004	6.186	1.835×10^{-7}	5.893×10^{-16}
Mg	XII	$2p\ ^2P_{1/2} \rightarrow 1s\ ^2S_{1/2}$	1.472	8.425	1.730×10^{-7}	4.080×10^{-16}

Fig. 4. Emission measure of the PCM fit to simulated CIE plasma. The continuous line depicts input data, and the dotted lines are initial and final distributions of the PCM fit. The model includes both a theoretical spectrum and a model of detector performance, including quantum efficiency and detector resolution.



in modern statistics. Bootstrap methods are broadly useful for many analysis algorithms, both binned and unbinned. Since we have implemented PCM using Monte-Carlo methods for both the model and simulated data to be compared to actual observed-event data, the estimation of the errors in the model derived by PCM are quantified by re-running PCM using bootstrapped versions of the observed data. This re-sampling of the observed data is analogous to a Monte-Carlo realization of the data. Details of the bootstrap method are beyond the scope of this paper.

3. Example analysis

Results from this implementation are preliminary and serve primarily as illustrative examples, but nonetheless may demonstrate the versatility of the method. As mentioned, all models were constructed from APED CIE plasma models. However, other models could readily be substituted in these examples.

3.1. Cassiopeia A

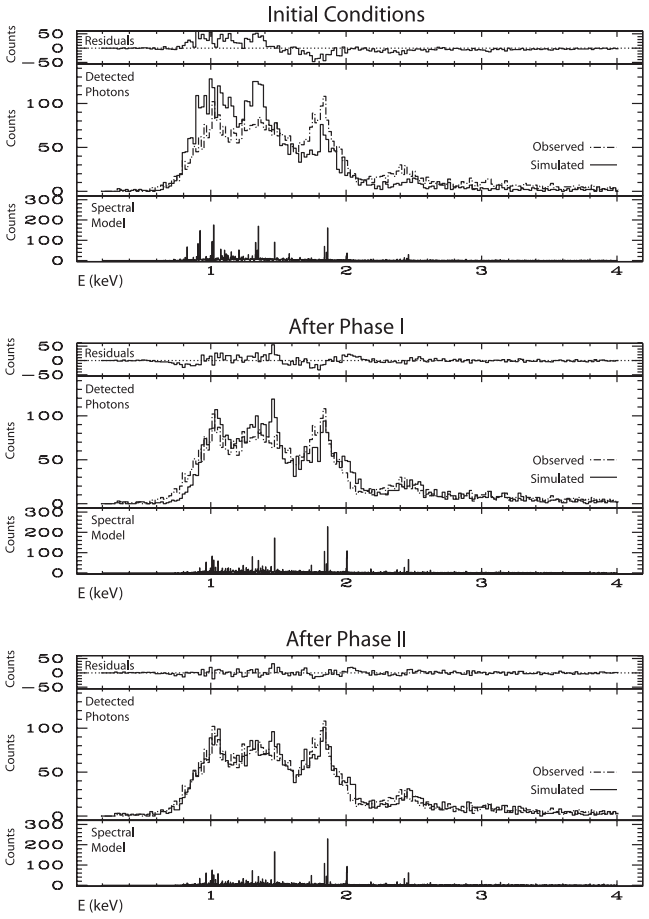
The first example is a Chandra ACIS-S3 observation of Cassiopeia A (Cas A). The observation used, Chandra ObsID 4638, comprises 167 ks of exposure and totals over 24×10^6 events. An extraction region defined by a circle of radius 19.8″ centered about RA 23° 23′ 32.0″ Dec +58° 49′ 29.6″ was used to generate a subset of almost 30 000 events (see Fig. 6). This region was chosen to capture X-rays from the diffuse central portion of the remnant while keeping the total number of events to a practical amount. The events were extracted from Chandra pipeline data products using the CIAO[4] tool `dmcopy`, and RMF and ARF files were generated with `specextract`. The design goal is not to implement any processing code as part of PCM1D that is already available by using a standard processing tool.

An initial arbitrary guess is generated from a CIE plasma of solar abundance with a broad Gaussian emission measure distribution in temperature, as shown in Fig. 3. Interstellar medium attenuation is modeled as an effective hydrogen absorption column with a density of $N_H = 1.3 \times 10^{22}$ cm^{-2} [12]. The choice of the centroid temperature and deviation of the initial model's emission measure is arbitrary, and is seen to provide a poor fit in the top panel of Fig. 5. As the abundance ratios and ISM absorption are fixed parameters in this fit, phase I is not expected to converge to a high confidence level. Subsequent tuning during the second phase fixes discrepancies around the Fe L-complex and the Mg and Si lines at 1.47 and 1.87 keV, respectively. See Table 1 for the location of these emission features.

The resulting emission measure is in agreement with published temperature ranges for knots in Cas A, which vary near 1 keV for the soft X-ray component of the spectrum below 6 keV [13]. The highly peaked emission measure distribution suggests this region is approximately fit with a single-temperature model. The line list in Table 1 is generated directly from counts in the final model photon list, divided by detector effective area contained in a standardized ARF. These numbers are useful in determining the correct elemental abundance ratios, as the fractional change in total flux per element from the end of phase I to the final solution is directly related to the abundance ratio in the model.

[4] Chandra Interactive Analysis of Observations. See http://cxc.harvard.edu/ciao/

Fig. 5. Analysis progression, Cassiopeia A. The top panel of the figure shows the initial conditions input to the Photon Clean Method (PCM). The middle portion of the panel shows a spectrum in the form of a histogram of the Chandra event data for a portion of Cassiopeia A (broken-dotted line). The continuous line shows the simulated data derived from an initial model of a Collisional Ionization Equilibrium (CIE) plasma with solar elemental abundances and a broad Gaussian distribution in emission measure (see the initial emission measure distribution in Fig. 3). The top portion of the panel shows the discrepancy between the simulated and observed histograms. The large difference residuals indicate that the initial model is a poor fit to the observations. The middle panel shows the same information, in the same form as for the top panel to aid comparison. The displayed data corresponds the end of Phase I of the PCM (see Fig. 2). Phase I finds an optimal solution based on a blend of CIE plasmas with a more highly constrained distribution in emission measure (see final distribution in Fig. 3). Similarly, the lower panel shows the end of Phase II. Phase II finds an optimal solution by modifying individual photons which are not constrained by strict adherence to a CIE model.



3.2. Simulated CIE blend

A simulated Chandra ACIS observation of a CIE plasma source with a broad Gaussian emission measure distribution centered about 2.6 keV was produced using ARF and RMF files from the previous Cas A analysis (see Fig. 7). The emission measure distribution of this simulated data set is plotted with a continuous line in Fig. 4, and is seen to have broad high-

Fig. 6. Unsmoothed Chandra broadband image of Cassiopeia A, with logarithmic intensity scaling. Circular extraction region for Cassiopeia A analysis is marked in white.

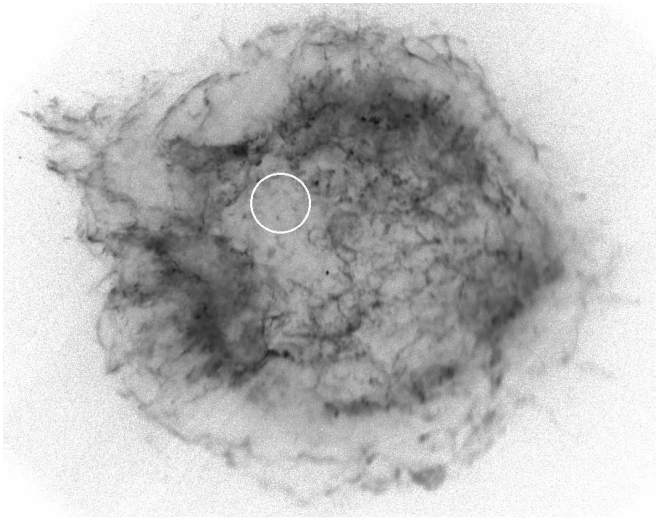
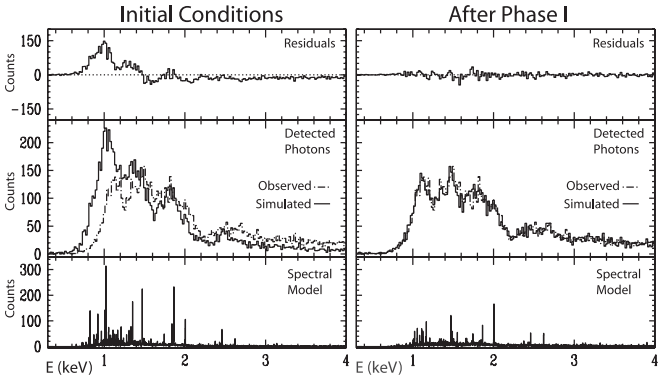


Fig. 7. The Photon Clean Method fit to simulated Collisional Ionization Equilibrium plasma with Gaussian distribution in emission measure. The left and right panels computed similarly to the top and middle panels of Fig. 5. Since this example is based on simulated data, the Phase I solution residuals are consistent with Poisson noise. Therefore, there is no need to compute a Phase II solution. Note the initial and final solutions correspond to the emission measure distributions shown in Fig. 4.



and low-temperature tails. The initial guess input to PCM is a much broader arbitrary Gaussian distribution, with a lower centroid of 1 keV. The lower overall temperature in the initial guess is evident in the Fe L-shell forest in the left panel of Fig. 7.

The purpose of testing in this manner is that we may use identical calibration files for the simulated observed data and the model data, which removes the effects of calibration errors. This provides a control environment that establishes that PCM works under ideal conditions. In addition, PCM is shown to fit data with complex models of high dimensionality with nontrivial parameter distributions. Because of the precise control over free parameters ($N_{\rm H}$, element abundances) in the simulated input, PCM is able to produce a solution with residuals consistent with Poisson noise (right panel, Fig. 7) in phase I alone, and does not need to proceed to phase II as expected for ideal simulated data that matches a selected model perfectly.

4. Summary and future work

The method as described in this paper is limited in function but retains several advantages over traditional fitting methods, most importantly in its ability to handle complex models with many parameters. It is well suited to resolving error between blends of lines and other problems related to detector resolution and calibration. Future versions of the code will benefit from being able to vary multiple parameters independently, multiple simultaneous models, and fitting in two or more dimensions at once, for example, fitting to energy and spatial coordinates simultaneously.

Early testing with the current version of the software has produced promising results on a number of celestial sources. Further tests to quantify its performance and to extend the number of models available to the program are underway, in anticipation of the release of a beta version to be made available to the general research community for evaluation and feedback in 2007.

It is also planned to apply PCM to the thermal emission or iron recorded LLNL's electron beam ion trap using a microcalorimeter under simulated Maxwell–Boltzmann conditions. The microcalorimeter records the spectral emission from about 300 eV to the above ionization potential of Fe^{25+} near 10 keV. The spectrum thus includes both the L-shell as well as the K-shell emission of iron. The temperature distribution derived by PCM can thus be compared to the temperature setting of the Maxwell–Boltzmann simulator. This will provide a stringent test of the accuracy of the method.

Acknowledgment

Work by the University of California Lawrence Livermore National Laboratory was performed under the auspices of the Department of Energy under contract No. W-7405-Eng-48. We thank H. Tananbaum and J. McDowell of the Chandra Science Center (NASA contract NAS8-39073), S. Kahn, The RGS/XMM-Newton US team leader (NASA contract NNG04GL76G) and W. Craig and S. Labov of the I Division of LLNL for their support of the earlier development of the PCM concept.

References

1. J.G. Jernigan and D. Vezie. Ast. ASP Conf. Series, **101**, 167 (1996).
2. J.R. Peterson, F.B.S. Paerels, J.S. Kaastra, M. Arnaud, T.H. Reiprich, A.C. Fabian, R.F. Mushotzky, J.G. Jernigan, and I. Sakelliou. Astron. Astrophys. **365**, L104 (2001).
3. J.R. Peterson, J.G. Jernigan, and S.M. Kahn. Astrophys. J. **615**, 545 (2004).
4. J.R. Peterson, P.J. Marshall, and K. Anderson. Astrophys. J. **655**, 109 (2007).
5. M.F. Gu, R. Gupta, J.R. Peterson, M. Sako, and S.M. Kahn. Astrophys. J. **649**, 2 (2006).
6. M.H. Carpenter and J.G. Jernigan. Public release of a one-dimensional version of the Photon Clean Method (PCM1D), AAS, HEAD, 9, 13.63C (2006) (see also OSTI ID 894786 (2006), http://www.osti.gov/bridge/).
7. D.W. Savin, P. Beiersdorfer, S.M. Kahn, B.R. Beck, G.V. Brown, M.F. Gu, D.A. Liedahl, and J.H. Scofield. Rev. Sci. Instrum. **71**, 3362 (2000).
8. F.S. Porter, G.V. Brown, K.R. Boyce, R.L. Kelley, C.A. Kilbourne, P. Beiersdorfer, H. Chen, S. Terracol, S.M. Kahn, and A.E Szymkowiak. Rev. Sci. Instrum. **75**, 3772 (2004).
9. R.K. Smith, N.S. Brickhouse, D.A. Liedahl, and J.C. Raymond. ASP Conf. Series, **247**, 161 (2001).
10. A. Kolmogorov. Ann. Math. Stat. **12**, 461 (1941).
11. B. Efron. Ann. Stat. **7**, 1 (1979).
12. J. Vink, J.S. Kaastra, J.A.M. Bleeker, and H. Bloemen. Adv. Space Res. **25**, 689 (2000).
13. J.S. Lazendic, D. Dewey, N.S. Schulz, and C.R. Canizares. Astrophys. J. **651**, 250 (2006); **408**, 1087 (2003).