X-RAY SPECTROSCOPY

Jean Ballet [1]

Abstract. Spectral analysis is the main tool of astrophysical research. Current instrumental technology offers very sensitive CCD detectors with enough spectral resolution for many purposes. This introductory text is aimed at graduate students starting a thesis in X-ray astronomy, and at non-specialists who wish to analyse X-ray data themselves. It gives the basic tools and explains the basic issues often implicit in specialised papers.

X-ray spectroscopy differs from longer wavelengths because it is photon based with relatively small numbers of photons. As a result the Poisson process is at the core of the statistical description of X-ray spectra. The classical model-fitting procedure is described and illustrated on a detailed example. It is applicable to any parametric model and any instrumental data.

High resolution X-ray spectroscopy, now applicable to many sources thanks to sensitive grating instruments, allows to measure individual lines accurately. The diagnostics that it offers are very similar to those offered by mid-resolution optical spectroscopy: line ratios, line broadening, absorption lines.

1 Introduction

This chapter relates the instrumental characteristics of X-ray detectors (as described by Barret, this volume) and the radiation processes described by Marcowith and Tatischeff (this volume).

The X-ray telescopes and detectors collect photons, which are detected one by one and sent to the ground. Instrument-specific procedures then transform this raw data into a calibrated list of photons, characterised by a time, a position and an energy channel. Any kind of selection may be applied to that list, and any kind of histogram may be built, most classically images, spectra and light curves. Most of this text deals with spectral analysis, which is the most direct way of diagnosing

---

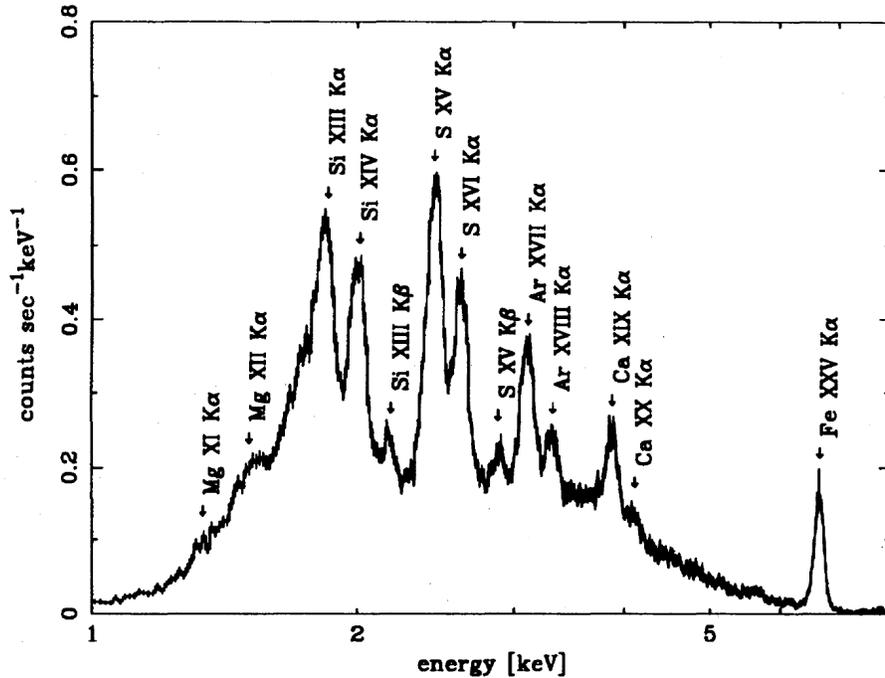[1] DSM/DAPNIA/SAp, CEA Saclay, 91191 Gif sur Yvette Cedex, France

Fig. 1. X-ray spectrum of the heavily absorbed ($N_H = 4\ 10^{22}$ cm^{-2}) supernova remnant W 49 B observed with ASCA (Fujimoto et al., 1995). K lines of elements Si, S, Ar, Ca and Fe are clearly visible.

physical conditions in the source. Using several examples, I show how to compare observed spectra with theoretical models, using a statistical framework.

Section 2 describes the main spectral and instrumental features. Spectral analysis may be divided into two cases, using quite different methods:

- When spectral lines are clearly resolved (high resolution spectroscopy, for example with gratings), a local analysis (around each line) is usually preferred. Section 3 deals with that situation.

- When spectral lines are not resolved (low resolution spectroscopy, for example with CCDs), a global modelling of the spectrum as a whole is necessary. Section 4 deals with that situation.

The text is openly biased towards line-dominated spectra, but the methods of Section 4 apply to featureless spectra as well. Section 5 opens a few perspectives on more advanced methods.
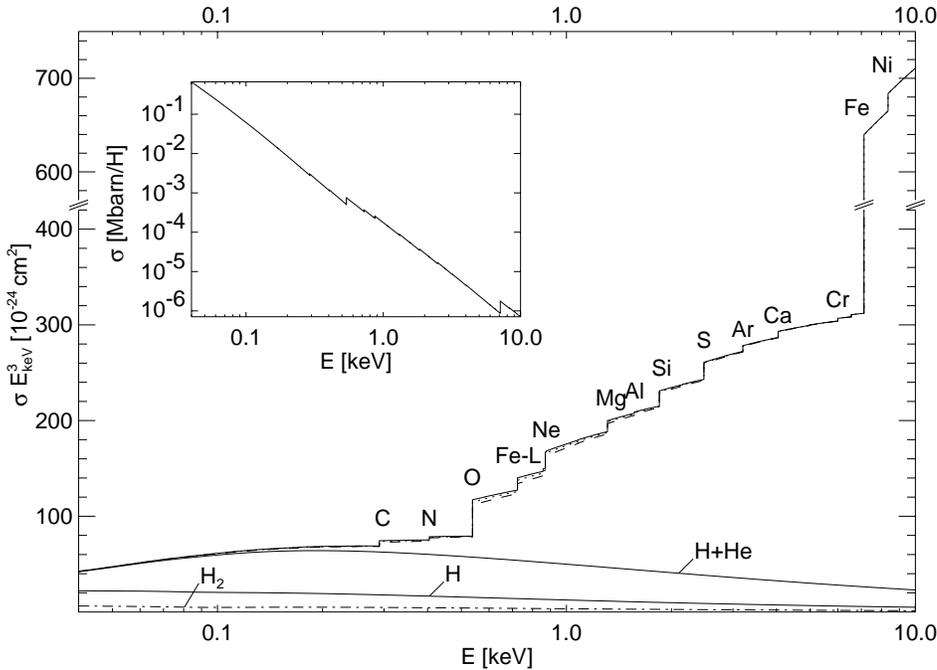
Fig. 2. Cross-section for interstellar absorption in X-rays (Wilms, Allen and McCray, 2000). The inset shows the absolute cross-section, illustrating the steep decline to larger energy. The large figure is the same curve multiplied by E_{keV}^3, showing the contributions of all elements. The steps correspond to the absorption thresholds for K electrons (also L, for Fe). This illustrates that hydrogen itself plays a very small role. The most important absorbers are He, O and Fe (above 7.12 keV). An obvious consequence is that a different metallicity can affect a lot the energy dependence of the X-ray absorption.

2 Spectral modelling

2.1 Typical X-ray spectra

Examples of X-ray spectra are shown throughout the text. Like the one shown on Fig.1 (Fujimoto et al., 1995), they all contain, in various proportions, spectral lines (broadened by the instrumental resolution) and continuum emission. The usual representation (the best one for data analysis, if not for science) is in cts/s/keV (or /Å). This is directly what is measured by the instruments. The spectrum always decreases toward high energy because the photons carry more energy there and hence are less numerous. X-ray spectra are nearly always limited by statistical precision at high energy. The spectrum also decreases toward low energy because of interstellar absorption. Both effects are amplified by the variations of the effective area of the instrument (also maximum in the middle of the spectral range).

2.2   *Interstellar absorption*

Interstellar absorption is a major issue in X-ray astronomy, at least below 1 keV (and further when the column density N_H gets into the 10^{22} cm^{-2} range). It has been discussed several times, most recently (and comprehensively) by Wilms, Allen and McCray (2000). The cross-section σ is a steeply decreasing function of energy (Fig.2). The spectrum is attenuated following

$$S(E) \;=\; \exp\left(-\sigma(E)N_H\right)\; S_0(E) \qquad\qquad (2.1)$$

X-ray absorption does not depend too much on the state of matter as long as it is cold (whether H is molecular or atomic, whether metals are gaseous or in grains). But it depends sensitively on

- abundances of the dominant species. Those are usually set to interstellar, but it is not necessarily a good choice when dealing with external galaxies, for example.

- ionisation. Whenever the K (and L, for Fe) electrons start being removed, the absorption edges shift in energy and depth. This is mostly important around active galactic nuclei (warm absorber).

- redshift. Obviously, when absorption occurs in an external galaxy, all the absorption edges follow the galaxy's redshift.

2.3   *Basic models*

The first thing to do when trying to interpret an observed sectrum is to compare it with basic radiation mechanisms (Marcowith and Tatischeff, this volume). The main categories of models include:

- black body emission (neutron star surface). The spectral shape depends on temperature only.

- coronal or hot thin plasma with collisional ionisation (stellar coronae, supernova remnants and hot interstellar medium, clusters of galaxies). The spectral shape depends mostly on temperature (and abundances).

- photoionisation or cold plasma irradiated by a hard continuum (around active galactic nuclei). The spectral shape depends on the number of ionising photons per atom.

- comptonisation or hot plasma irradiated by a soft continuum (X-ray binaries, active galactic nuclei). The spectral shape depends on temperature and optical depth.

- synchrotron or inverse Compton of relativistic electrons (supernova remnants, pulsars, blazars). The spectral shape is usually a power law defined by its slope.
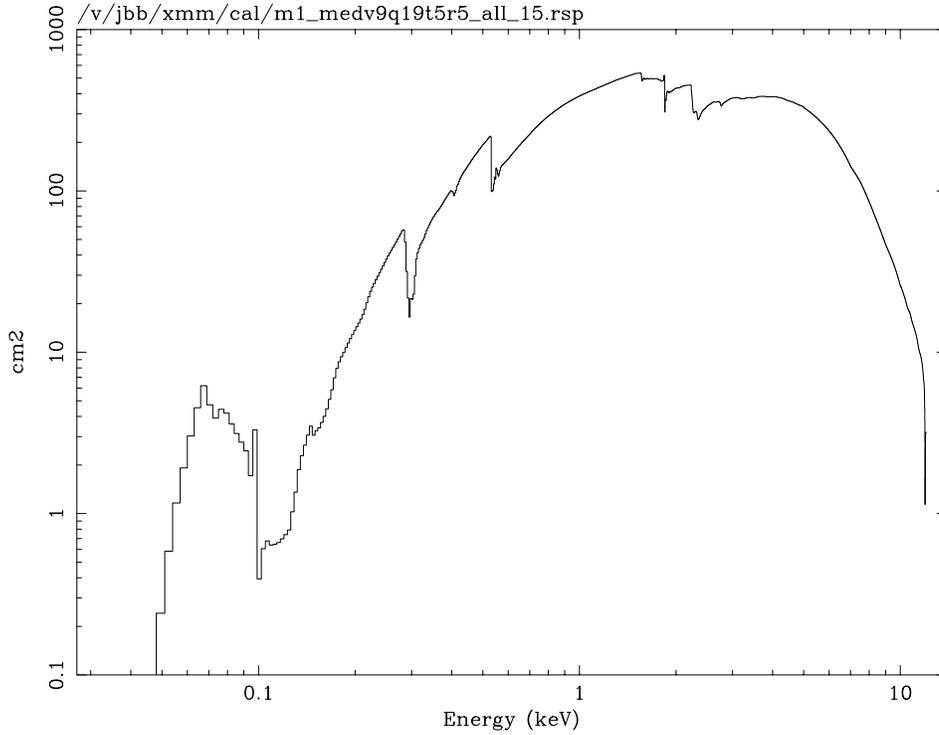
Fig. 3. Total effective area as a function of energy of the EPIC/MOS instrument on *XMM-Newton* used in the example of Sect.4. The absorption thresholds of C K (0.29 keV), O K (0.53 keV), Si K (1.84 keV) and Au L (2.25 keV) are due to intervening material in the mirror, the filter or the top layer of the CCD. The effective area decreases at low energy due to absorption along the light path, and at high energy because the detector becomes transparent.

2.4 Instrumental effects

The instrumental representation relates the model spectrum $M(E)$ (in ph cm^{-2} s^{-1} keV^{-1}) to a model observation $H(C)$ (in cts/channel) directly comparable to the data for a known observation time t:

$$H(C) = \left(\int_E R(E,C)\, A(E)\, M(E)\, \mathrm{d}E \ + \ B(C) \right) t \qquad (2.2)$$

$A(E)$ (in cm^2) is the effective area of the instrument in the direction of the source (.arf file, example on Fig.3). $R(E,C)$ is the energy redistribution matrix of the instrument (.rmf file, example on Fig.4) such that

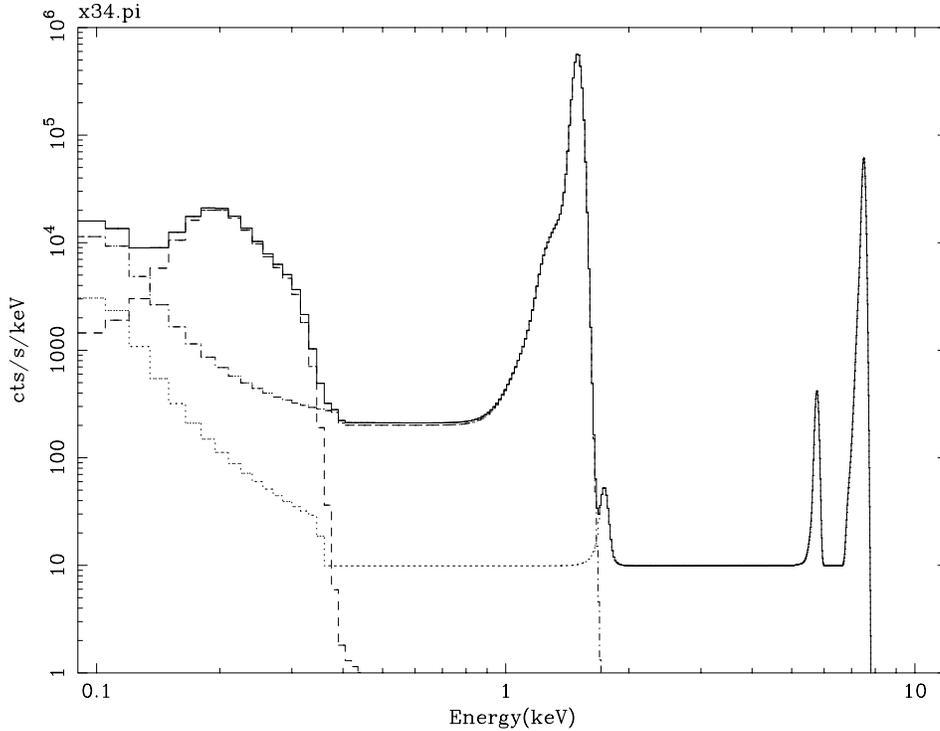$$\sum_C R(E,C) = 1, \quad \forall E \qquad (2.3)$$

Fig. 4. Spectral response of the EPIC/MOS instrument on *XMM-Newton* (used in the example of Sect.4) to three narrow lines at 0.3, 1.5 and 7.5 keV. At 0.3 keV, the energy resolution $E/\Delta E$ of a CCD detector is very bad, but it improves at higher energy. The secondary peaks in the response to the 7.5 keV line are due to Si K fluorescence (1.74 keV) and Si K escape (7.5 - 1.74 keV).

The product $A(E)\,R(E,C)$ is called the response matrix (.rsp file).

For visual representation, the channels are converted to energy, associating each channel with the energy E where the peak of the spectral response is at that channel. Because the detectors are quite linear, this is usually very close to a linear relationship. This allows, as in Fig.1, to plot the data as a function of energy (much easier to associate lines with their physical origin in that representation), and dividing by the channel width converted to keV avoids steps if channels are rebinned. But it should be kept in mind that this is only a visual representation. All numerical operations (such as the statistical fits of Sect.4) must be performed on the data in cts/channel.

The background $B(C)$ (in cts/channel/s) is usually a compound of instrumental and astrophysical origins. The instrumental contribution, due to energetic particles interacting with the detector, usually has constant spectral and spatial (in detector coordinates) shape. A classical way of estimating it is to use a long

reference observation and extract the spectrum from the same region on the detector as the source. It must be renormalised by a scale factor (normally close to the ratio of observing times) obtained by comparing images outside the field of view or spectra at very high energy (where the effective area for true X-rays tends to 0). This reference contains some astrophysical background as well, but not necessarily the right one for the current field.

In addition to the instrumental background, the residual astrophysical component (positive or negative) may be estimated if it can be assumed to be uniform over the field of view, and the source does not cover the entire field of view. The background spectrum should be extracted in an area outside (if possible around) the source, larger than the source area to avoid losing statistical precision. It must be corrected by the ratio of effective areas between the source and the background areas (vignetting).

Another important instrumental characteristic is the precision on calibration. This is represented as a certain fraction of the effective area (or of the signal) $Sys\_err(C)$ (systematic error). It dominates over statistical precision for bright sources at the peak of the spectrum (usually around 1 keV).

2.5 Fitting packages

A number of integrated fitting packages are freely available. The best known (and most used) is XSPEC (Arnaud, 2001). Another one (specialised in line emission) is SPEX (Kaastra, 2001). They are particularly useful for fitting a whole spectrum (Sect.4).

Those packages allow to visualise models and data, both in the observation space and in the source space (related by Eq.2.2).

They allow to fit several observations at a time. Each data set must be associated with a .arf and .rmf file (Sect.2.4), representing the instrument.

They propose many standard models, which may be combined (added or multiplied) to represent complicated situations. Each model depends on a number of parameters. Those parameters may be linked with one another (for example, one may force the iron abundance to be twice the silicon abundance in a coronal model).

The core of the package is a minimisation engine, working on a choice of statistical criteria. It provides the best fitting set of parameters, as well as confidence intervals or contours on parameters (Sect.4).

3 Spectral lines

Gratings on *Chandra* (LETG, HETG) and *XMM-Newton* (RGS) are the most sensitive high resolution X-ray spectrometers in use today. They allow to reach sources down to $F_X = 10^{-12}$ erg cm^{-2} s^{-1}, and offer spectral resolution $E/\Delta E$ of several hundred. Because the spectral width of a grating plate is nearly constant in wavelength (but not in energy), grating spectra are usually presented as a
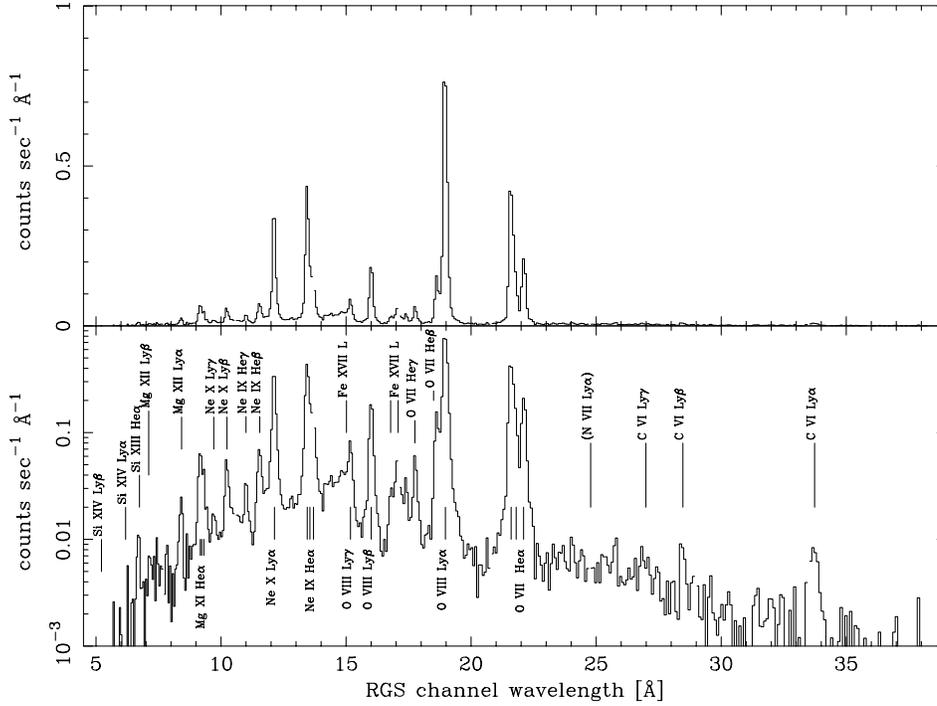
Fig. 5. Spectrum of the supernova remnant 1E 0102-72.3 in the Small Magellanic Cloud, obtained with the RGS instrument on *XMM-Newton* (Rasmussen et al., 2001). The top and bottom halves are the same spectrum in linear and logarithmic representations, respectively. K lines of elements C, O, Ne and Mg, as well as Fe L lines, are resolved.

function of wavelength. For example, the *XMM-Newton* RGS has $\Delta\lambda \simeq 0.04$ Å. 1 keV corresponds to 12.399 Å(and vice-versa).

3.1 Emission lines

Many X-ray sources, particularly the coronal sources like on Figs.5 (Rasmussen et al., 2001) and 6 (Güdel et al., 2001b) are rich in emission lines which can be resolved by grating spectrometers. A good on-line reference for line energies and fluxes is APEC/APED (Smith et al., 2000).

When the lines are clearly resolved, the simplest approach is to extract each one separately (by the same methods described in Sect.4), modelling each line by a constant (background) plus a gaussian (flux, energy and width of the line).

The line must first be identified, on the basis of its energy (or wavelength) and flux. This is often better done collectively (for example, a set of lines arising in the same source at unknown redshift are all shifted by the same energy ratio). Flux is used to eliminate unlikely identifications, on the basis of abundances.

Once this is done, the comparison between the expected and measured energies
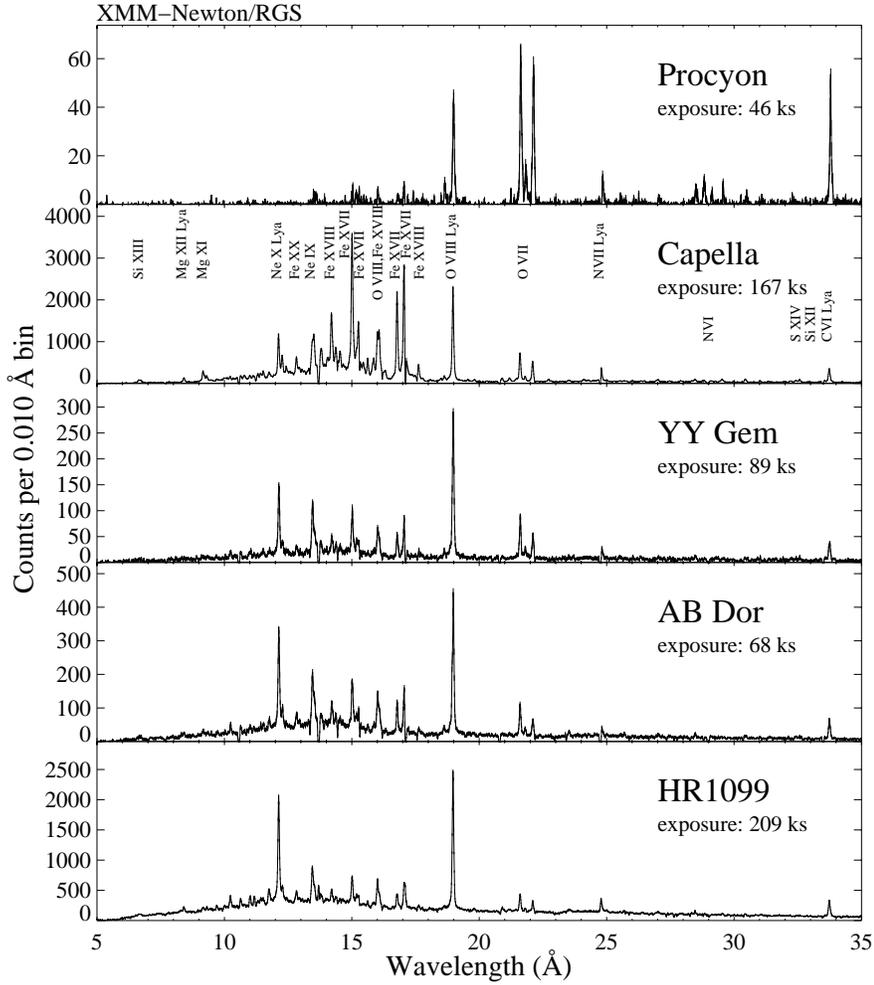
Fig. 6. Spectra of five stellar coronae observed with the *XMM-Newton* RGS (Güdel et al., 2001a). The average coronal temperature increases from top (Procyon, 1-2 10^6 K) to bottom (HR1099, 10-30 10^6 K). The inferred Ne abundance is much larger and the Fe abundance much lower than in the solar corona (less active than those five).

gives the velocity or redshift of the source. In some cases, it may also be interpreted as an ionisation effect (for example, the centroid of the Fe K line slowly increases with ionisation stage above Fe$_{XVII}$).

The intrinsic width of the line (beyond the instrumental width) is related to Doppler broadening and/or the spatial structure of the source for gratings (Sect.3.5).

Finally, the photon flux in the line is related to the astrophysical quantities by

$$
\begin{aligned}
F_{i,l} \,(\mathrm{ph/cm^2/s}) \;&=\; \int \frac{n_e\, n_{ion}}{4\pi D^2}\, \epsilon_{i,l}(T)\, d\tau \qquad\qquad\qquad\qquad (3.1)\\[2mm]
&=\; \frac{8.63\ 10^{-6}}{4\pi D^2}\, \frac{n_Z}{n_H} \int n_e\, n_H\, X_i\, \frac{Br_{i,l}\, \Omega_{i,l}(T)}{\omega_0 \sqrt{T}}\, \exp\!\left(\frac{-E_{i,l}}{kT}\right)\, d\tau
\end{aligned}
$$

where the integration runs in the emitting volume along the line of sight. n_e and n_H are the electron and hydrogen densities, T the electron temperature, D the distance to the source. i denotes ionisation stage and l a particular line. X_i is the ionic fraction of the emitting ion from element Z. ω_0 is the statistical weight of the ground level (from which excitation starts). $Br_{i,l}$ is the branching ratio of line l from its upper level. $\epsilon_{i,l}(T)$ is the excitation coefficient of the upper level of the transition, and $\Omega_{i,l}(T)$ the associated collision strength, factoring cascade excitation. $E_{i,l}$ is the line energy. When T is constant, the flux is proportional to the emission measure $\int n_e n_H d\tau$.

3.2   Plasma diagnostics by line ratios

This has been the workhorse of optical astronomy for decades. It has not been used as much in X-ray astronomy because it requires good spectral resolution, but this is now changing with the advent of sensitive grating instruments. Line ratios allow to forget about unknown normalisation factors in Eq.3.1 which are the same for both lines. For that reason ratios between lines of the same element are preferred (they do not depend on abundances). The ratios must be corrected for extinction (as in the optical). They can be related (via functions depending only on atomic physics) to the parameters of the plasma. What follows is written for coronal plasmas, but it can be transposed to the case of photoionised plasmas as well. In the limit of low-density plasmas ($n < 10^8$ cm^{-3}), the equilibrium is between collisional excitation and radiative deexcitation, and the line fluxes are determined by the excitation mechanism only. All excitations are from the ground state. The line ratios (in photons, not ergs) may be written, following Eq.3.1:

$$
\frac{F_{i,l}}{F_{j,m}} \;=\; \frac{X_i}{X_j}\, \frac{Br_{i,l}}{Br_{j,m}}\, \frac{\epsilon_{i,l}(T)}{\epsilon_{j,m}(T)} \;=\; \frac{X_i}{X_j}\, \frac{Br_{i,l}}{Br_{j,m}}\, \frac{\Omega_{i,l}(T)}{\Omega_{j,m}(T)}\, \exp\!\left(\frac{E_{j,m}-E_{i,l}}{kT}\right) \quad (3.2)
$$

The notations are the same as in Eq.3.1. The exponential term is usually varying more rapidly than the Ω ratio.

The most obvious line ratio (involving the brightest lines) is the Lyα to Heα ratio between K lines of the H-like and He-like ions of the same element. This depends mostly on the ionic fractions of those ions (and also on temperature). When collisional ionisation equilibrium (CIE) is reached, that ratio depends on temperature only. Because CIE is not always assured in hot tenuous plasmas, that ratio is used mostly to check whether CIE is reached after T has been estimated by other means. That ratio decreases with Z (toward heavier elements), as is clearly seen on Fig.1, because the ionisation threshold of the He-like ions increases as Z^2.
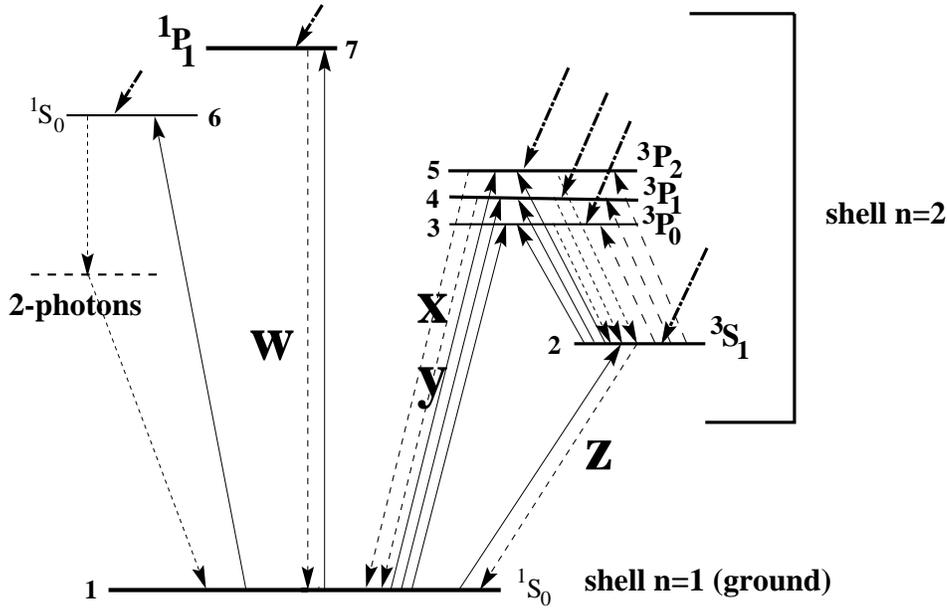
Fig. 7. Simplified level scheme for Helium-like ions, from Porquet et al. (2001): resonance (**w**), intercombination (**x + y**), and forbidden (**z**) lines. *Upward arrows*: collisional excitation transitions; *Dashed arrows*: radiative transitions; *Dot-dashed arrows*: recombination (radiative and dielectronic) plus cascade processes.

The line ratios of the same ion (Kα to Kβ and K to L) are preferred as temperature diagnostics, because they depend only on that parameter. The K to L line ratio of Fe is in theory the most sensitive (because of the large energy difference in the exponential term of Eq.3.2). But in practice it is more difficult to use because the K lines of different ions are often unresolved, and the L lines are more complex and difficult to calculate. The use of Kα to Kβ (and Kγ) is illustrated in Fig.9 (Rasmussen et al., 2001).

Finally, it must be stressed that plasma diagnostics rigorously apply only to single temperature plasmas. When gas at several temperatures contributes along the line of sight (a rather common occurrence, unfortunately), the ratios will represent an average and different line ratios may hint at different temperatures even if atomic physics is correct. They also assume that the lines are optically thin (not always true for the resonance lines).

3.3   The fine structure of the Heα line

The Heα line of all elements is a blend of three lines: resonance (w), intercombination (x+y) and forbidden (z). The Grotrian diagram for He-like ions (Porquet et al., 2001) is shown in Fig.7. When spectral resolution allows to measure them,
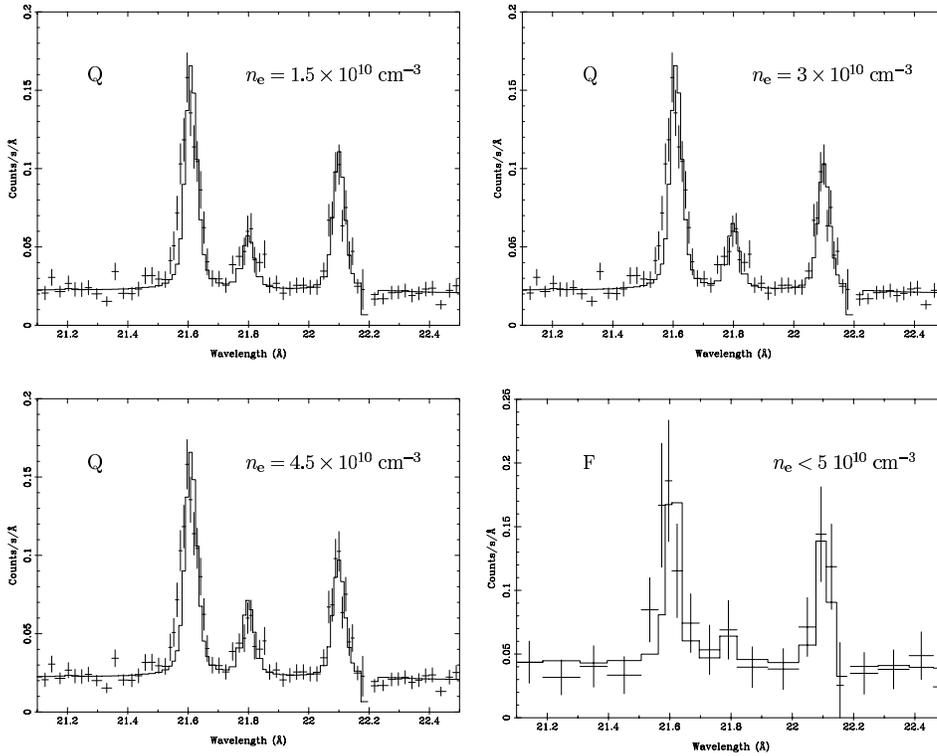
Fig. 8. Density estimate of the stellar corona of AB Dor (Güdel et al., 2001a). The first three quadrants show the same spectrum (quiescent state) around the O Heα triplet. The best fit is at top right. The lower limit on density (90% confidence interval) is at top left, the upper limit at bottom left. The bottom right spectrum was obtained during two large flares.

the ratio $G = (x + y + z)/w$ is temperature dependent and $R = z/(x + y)$ is density dependent around a critical density (10^9 cm^{-3} for C to 10^{14} cm^{-3} for Si). As density increases, the ^{3}S level (metastable) is collisionally depleted toward the ^{3}P level, and R decreases to 0. In particular, R is an important density estimate in stellar coronae as illustrated in Fig.8 (Güdel et al., 2001a), because the density there is precisely of that order of magnitude. The ratio G is also very sensitive to photoionisation (Porquet and Dubau, 2000). In photoionised plasmas it may go up to 4, whereas it is around 1 in collisional plasmas.

3.4 Abundances

Measuring abundances is the other major goal of spectroscopy. This may be done on emission lines only after the emitting conditions (temperature, most impor-
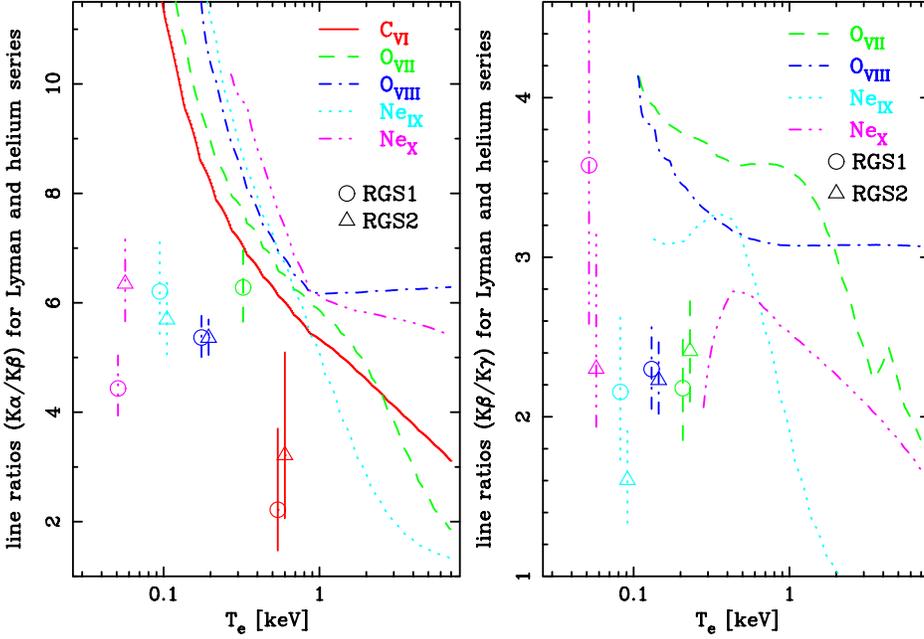
Fig. 9. Line ratios extracted from the *XMM-Newton* RGS observation of 1E 0102.2-72.3 (Rasmussen et al., 2001, Fig.5). The left figure shows the Kα to Kβ ratios of several ions, the right figure shows the Kβ to Kγ ratios. The curves are the predictions (as a function of electron temperature) of the APEC (Smith et al., 2000) line model assuming collisional ionisation equilibrium. Because this is not true in young supernova remnants (where the plasma is still ionising), no consistent solution may be obtained for O$_{\rm VII}$, for example.

tantly) are known from plasma diagnostics (Sect.3.2).

The measurable quantity is the equivalent width, the ratio of the line flux to the underlying continuum intensity. The abundance is related to the equivalent width by:

$$W = \frac{F_{Z,i,l}}{F_\nu} = \frac{n_Z}{n_H} X_{Z,i} \frac{\epsilon_{Z,i,l}(T)}{\epsilon_\nu(T)} \qquad (3.3)$$

The ϵ are provided by atomic physics. Of course if several mechanisms contribute to the continuum emission (like nonthermal mechanisms), they must be modelled and only that part of the continuum attributed to the same origin as the line should be counted in F_ν. The ionic fraction $X_{Z,i}$ is known from atomic physics as a function of temperature (Arnaud and Rothenflug, 1985) if CIE applies. If CIE is not reached then $X_{Z,i}$ must be modelled.

Stellar coronae (Fig.6) are a good example of relatively straightforward application. CIE is reached because of the high densities, the abundances are well-defined (constant within the corona). The difficulty is to model the temperature distribu-
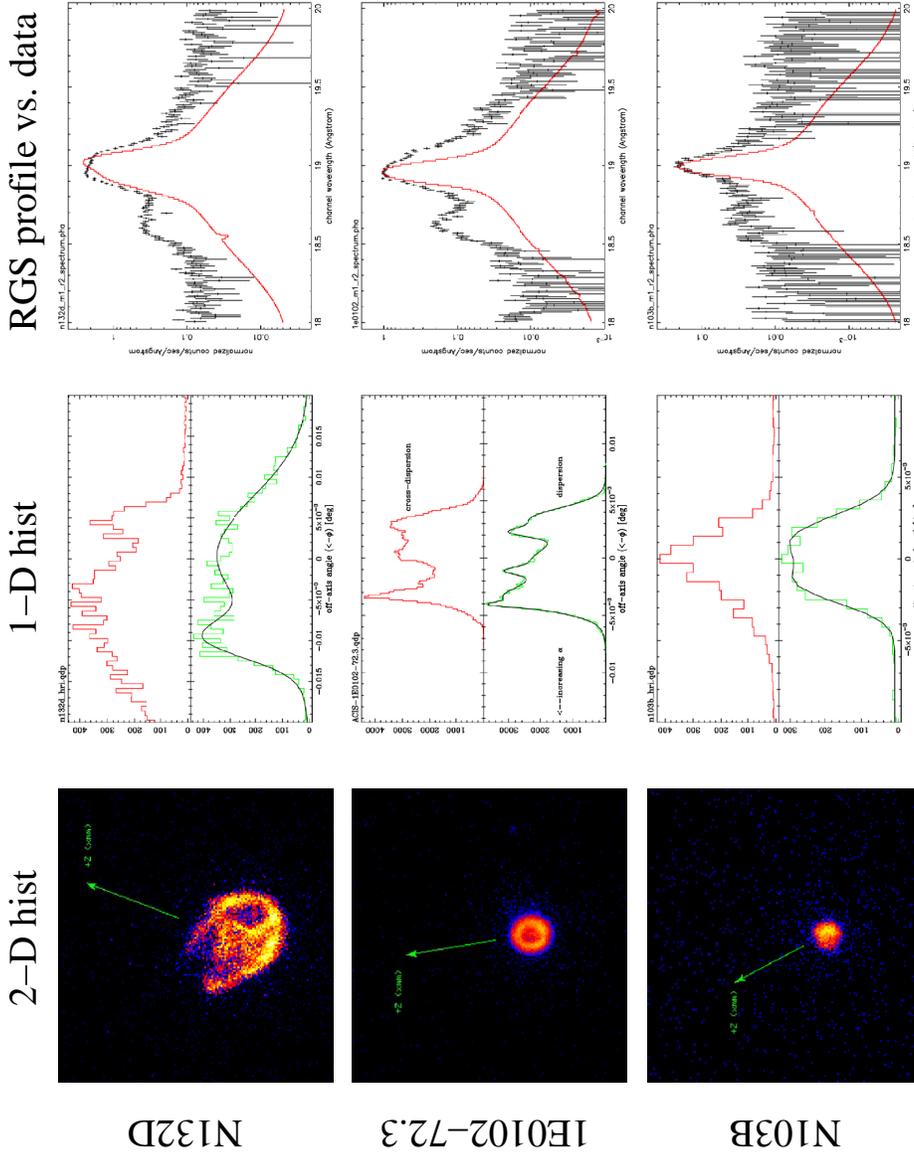
Fig. 10. *XMM-Newton* RGS observation of three supernova remnants in the Magellanic Clouds (Rasmussen et al., 2002). The central column shows the image profiles (using *Chandra* broad-band images) projected onto the dispersion (below) and cross-dispersion (above) axes of the gratings. The source extension along the dispersion axis broadens the spectral response. The right column compares the observed spectrum around the O Lyα line at 18.967 Å (with the O Heβ shoulder at 18.627 Å) to the line spread function broadened with the profile along the dispersion axis. This works well in N103B (bottom). The observed line is broader in 1E 0102-72.3 (centre) because of Doppler broadening ($\simeq$ 1500 km/s). The model profile does not fit the line shape in N132D (top). This is because the oxygen map does not look like the broad-band map.

tion (a corona is not a single temperature plasma). But in bright stars, this may be reasonably constrained by the data directly (Fig.18).

On the other hand, young supernova remnants are a good example of very difficult application. CIE is not reached because of the low densities, the abundances vary within the remnants (reflecting the nucleosynthesis), temperature varies as well, and the electronic temperature is often lower than the hydrodynamic temperature. So everything must be modelled (Ballet, this volume).

3.5   Line broadening

The intrinsic atomic width of the lines is still beyond the reach of current X-ray spectrometers. The thermal broadening

$$\frac{\Delta E}{E} \;=\; \frac{\Delta v}{c} \;=\; \frac{1}{c}\sqrt{\frac{kT}{2Am_p}} \;=\; 7.3\,10^{-4}\sqrt{\frac{kT_{keV}}{A}} \tag{3.4}$$

is also too small to be measured, particularly because the X-ray lines are from heavy elements ($A \geq 12$). Only Doppler velocities larger than 1000 km/s give rise to measurable broadening.

Another reason for broadening exists with dispersive spectrometers. The X-ray gratings are slitless, and as such they integrate everything shining within their field of view. This is of course a serious problem along the dispersion axis, because the same axis mixes spatial and spectral structure.

A good example of what happens for extended sources is shown in Fig.10 (Rasmussen et al., 2002), when the source's extent is small enough that the spatial broadening is comparable to the instrumental broadening. Then the spatial broadening may be folded into the spectral response (if the source's image is known). After this is done, true Doppler broadening can be measured (Fig.10, centre).

If the map used to model the source (usually from a broad band) is not representative of the image in the line, the model will not fit the line shape (Fig.10, top). In that case, if the line is isolated (not blended with other lines), the procedure may be inversed and an image of the source may be extracted in the line. The instrumental (and Doppler) broadening will then translate into a spatial resolution.

3.6   Absorption lines

On a strong continuum source, absorption lines may be seen. Because the absorbing gas is usually in coronal conditions (whether collisional or photoionised) and not at local thermal equilibrium, the vast majority of ions are at their fundamental level, so only lines arising from the fundamental level are observed. The most common site of X-ray absorption is next to active galactic nuclei, as illustrated in Fig.11 (Sako et al., 2001).

Like emission lines, the absorption lines must be identified. Then their equivalent width, velocity shift and broadening can be measured. The same procedures apply as in Sect.3.1.
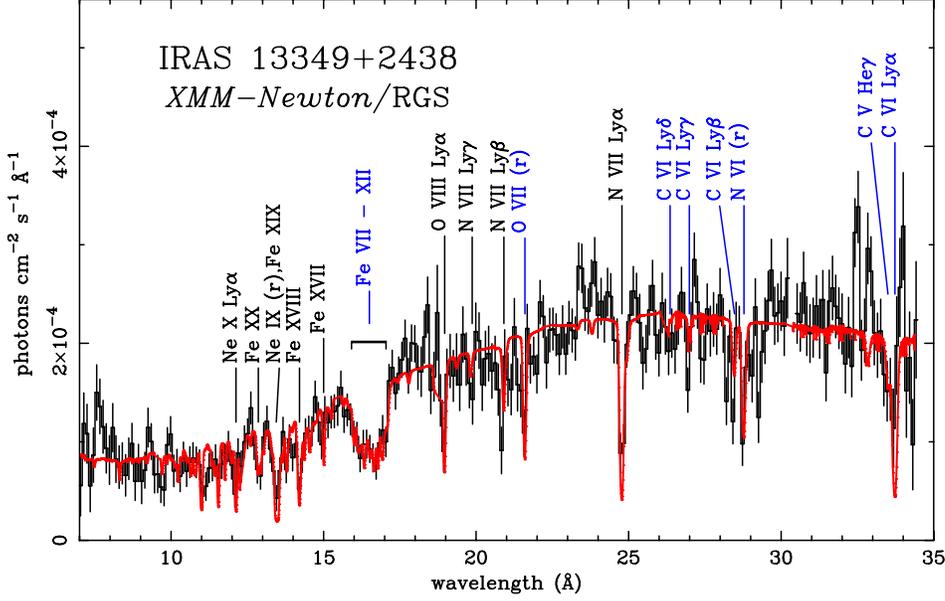
Fig. 11. *XMM-Newton* RGS spectrum of the quasar IRAS 13349+2438 (Sako et al., 2001). Two absorption line systems are detected. The first one (labelled in black) is strongly ionized ($\xi \simeq 10^2$ erg cm/s). The second one (labelled in blue) is less ionized ($\xi \simeq 1$ erg cm/s) and moving toward us (like a wind) at 400 km/s. The width of the lines points at turbulent broadening at $v_{turb} \simeq 1500$ km/s.

In the limit of unsaturated absorption (low optical depth), the measured equivalent width is directly related to the column density of the absorbing ion $N_{Z,i}$:

$$W = 1 - \frac{\int_{\nu \simeq \nu_0} F_\nu \, d\nu}{F0_{\nu_0}} = 2.65 \, 10^{-2} \, N_{Z,i} \, f_{Z,i,l} \text{ Hz} \qquad (3.5)$$

F_ν is the observed intensity, and $F0_{\nu_0}$ is the expected continuum intensity without absorption (interpolated from nearby frequencies). $\nu_0(Z, i, l)$ is the central frequency of the transition. $f_{Z,i,l}$ is the oscillator strength of the transition (known from atomic physics). This assumes that all the continuum is emitted behind the absorbing gas.

4   Global spectral analysis

CCDs on *Chandra* (ACIS) and *XMM-Newton* (EPIC) are the most sensitive X-ray spectrometers in use today. They allow to reach sources down to $F_X = 10^{-14}$ erg cm^{-2} s^{-1}, and offer an instrumental width ΔE around 100 eV.

To illustrate the general concepts described in Sect.2, I use as an example the *XMM-Newton* observation of the starburst galaxy NGC 253 (Pietsch et al., 2001).
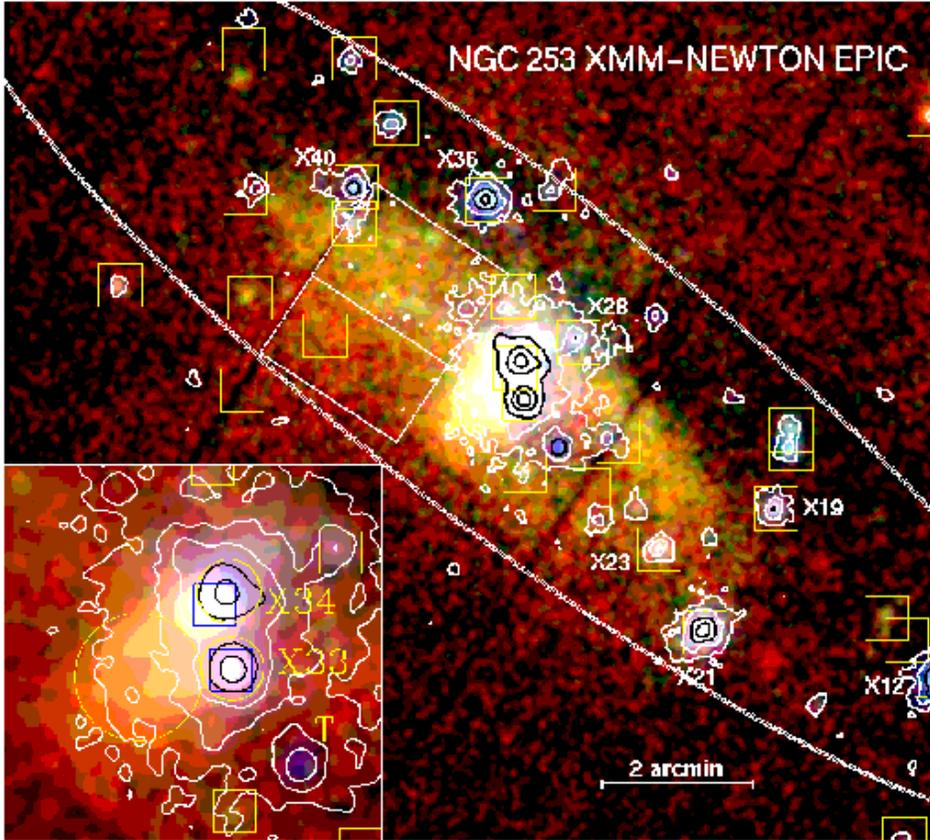
Fig. 12. Image of the starburst galaxy NGC 253 (Pietsch et al., 2001). The spectrum used as an example in Sect.4 is that of the nucleus (X34 in the close-up, bottom left).

The list of photons from the EPIC instrument is first projected on the sky to build an image (Fig.12). A spatial selection is applied to keep only the nucleus, and the energies are histogrammed to build the source spectrum (Fig.13). Another spatial selection is used to build the background spectrum for Eq.2.2.

4.1 Model fitting

The environment I have used is XSPEC (Arnaud, 2001). Following Pietsch et al. (2001), I apply a coronal model (mekal) absorbed by interstellar gas (either in our Galaxy or in NGC 253). mekal is an example of an additive model. phabs (interstellar absorption) is an example of a multiplicative model. Because NGC 253 is nearby, the redshift is not an issue. For a start I use a single temperature model.

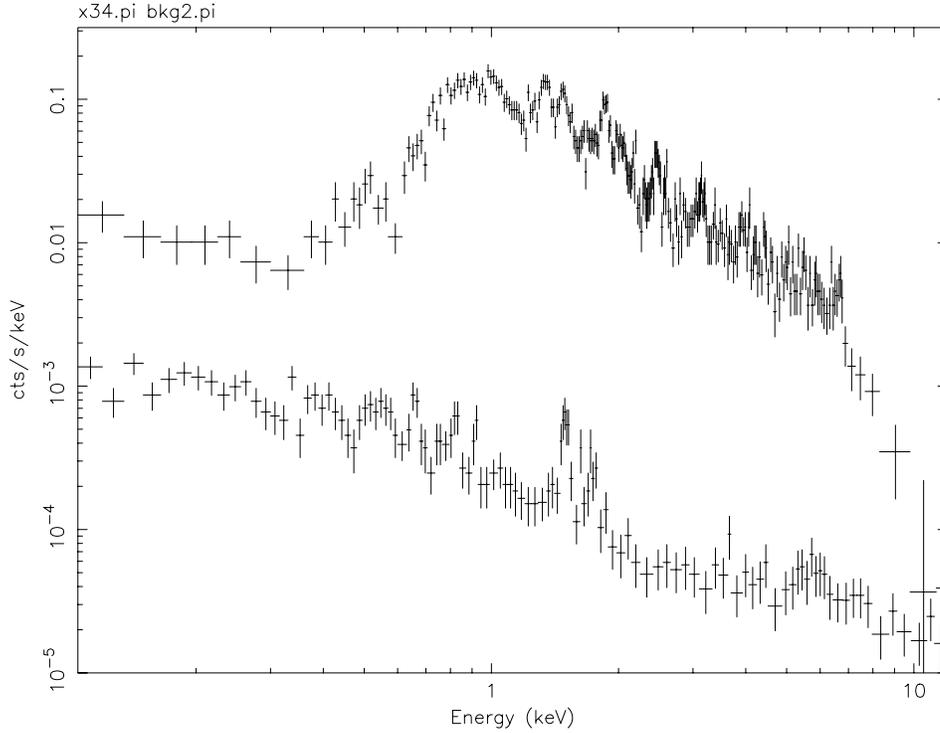$$M = \mathrm{phabs}(N_H) \, A \, \mathrm{mekal}(T, Z) \qquad (4.1)$$

Fig. 13. *XMM-Newton* EPIC/MOS spectrum of the nucleus of NGC 253 (above), after spatial selection in the image (Fig.12). Lines of Mg Heα (1.33 keV), Mg Lyα (1.47 keV), Si Heα (1.84 keV), S Heα (2.43 keV), Fe Heα (6.70 keV) are visible. The lower spectrum is the background. Here it was estimated from a larger area surrounding the nucleus (then scaled down to the same area as the nucleus). This ensures that the statistical accuracy on the background is larger than that on the signal. The line at 1.49 keV is due to Al Kα fluorescence on the aluminium shield protecting the CCD from particles and optical light coming from the sides. The spectrum shown on Figs.14 and 15 is background subtracted.

Z is the metallicity (abundances of elements heavier than He with respect to solar). A is the normalisation (related to the emission measure as in Eq.3.1).

4.2 Statistical criteria

The first choice to be made is which statistic to use when comparing the data $D(C)$ with the model (via Eq.2.2). The default statistic used in XSPEC is the χ^2 statistic.

$$S \; = \; \sum_C \left(\frac{D(C) - H(C)}{\sigma(C)} \right)^2 \quad (\chi^2, \text{ many cts/channel}) \qquad (4.2)$$

This applies when the errors $\sigma(C)$ on the data points follow a Gaussian distribution. It has the advantage of allowing to merge seamlessly errors of different nature (like measurement and calibration errors) by adding them quadratically. It also provides a measurement of the overall fit quality, but this value is difficult to use because it depends critically on the spectral binning (one doesn't get the same answer when the spectral channels are made twice bigger, for example).

But in X-ray astronomy the major cause of measurement uncertainty is the Poisson nature of the radiation we measure. The number of counts n measured in channel C is related to the model observation (Eq.2.2) by the Poisson probability

$$P(n) \;=\; \frac{H(C)^n}{n!}\, e^{-H(C)} \tag{4.3}$$

The Poisson law looks like a Gaussian distribution of width $\sigma(C) = \sqrt{H(C)}$ in the limit of large $H(C)$. The match is good for $H(C) > 30$ and passable for $H(C) > 10$. Actually the χ^2 formula applies only when $\sigma(C)$ does not depend on the model, so $\sqrt{H(C)}$ is replaced by $\sqrt{n}$ (or 1 when $n = 0$). This is the approximation made in XSPEC when $\sigma(C)$ is not explicitly provided.

In practice it is often the case (particularly at the high energy end of the X-ray band) that there is 0 or 1 photon per spectral channel. In that limit, the appropriate thing to do is go back to the original maximum likelihood formulation (Cash, 1979). For the Poisson law, this results in the following statistic (called C statistic in XSPEC), equal to - 2 ln(likelihood):

$$S \;=\; 2 \sum_C \left(H(C) - D(C) \ln H(C) \right) \quad (C_{stat},\ \text{few cts/channel}) \tag{4.4}$$

It may be applied to any data following the Poisson law (whatever the number of counts). The 2nd order development of each term in H around $H = D$ (best local fit) agrees with that of Eq.4.2 if $\sigma = \sqrt{H}$, so that both formulae agree in the limit of large H. Of course Eq.4.4 must be applied on true counts, and this is the reason why the background is entered in the model observation in Eq.2.2 instead of subtracted from the data. The C statistic is not much more difficult to use than the χ^2 statistic (it is not quadratic, though). It does not provide a number measuring the global quality of the fit, but its variations as a function of parameters allow to extract confidence intervals and contours exactly like the χ^2 statistic does. It does not allow to include easily the precision on calibrations, so the χ^2 statistic should be used when calibration precision is an issue (for bright sources).

Finally, the same formalism (Cash, 1979) may be used to form the statistic which should be used on unbinned data (list of photons):

$$S \;=\; -2 \sum_i \ln H(C_i) \quad (C_{stat},\ \text{unbinned counts}) \tag{4.5}$$

where the sum now runs over all detected photons. This is used a lot in γ-rays where the photons are even scarser.
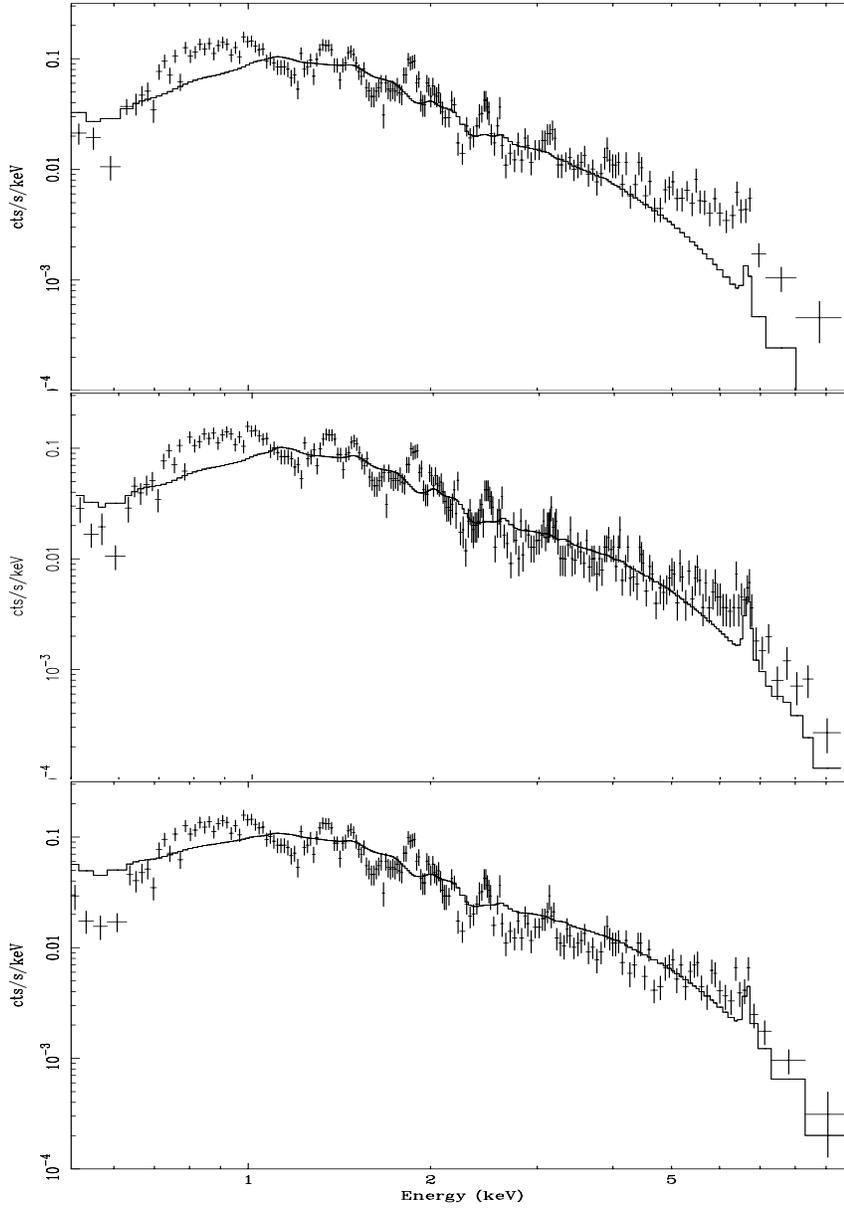
Fig. 14. Same spectrum, same model (Eq.4.1) but different statistics. The first fit (above) uses the χ^2 statistic directly on the original bins (the plotted bins are the result of a rebinning performed by XSPEC for plotting purposes only). The second fit (centre) uses the χ^2 statistic on bins grouped such that each holds at least 10 counts. Here the true bins are plotted. The third fit (below) is the correct one, using the C statistic (Eq.4.4). See Sect.4.2 for details.

Fig.14 illustrates the dangers of not using the right statistic. The top figure shows the fit using the χ^2 statistic (Eq.4.2) on the original channels (15 eV width). The binning shown by XSPEC (using the setplot rebin command) is for plotting only, but is not used in the fit. The best fit parameters are $N_H = 0.21\ 10^{22}$ cm^{-2}, $kT = 2.70$ keV, $Z = 0.11$ and $A = 1.56\ 10^{-3}$. Choosing that statistic is a very bad idea because many bins above 3 keV have only 0 or 1 photon per bin, so the Gaussian approximation to the Poisson law is very poor. The contribution of high energy bins to the χ^2 statistic is underestimated.

The middle figure applies the same χ^2 statistic, but after binning the data so that each spectral channel contains at least 10 counts. The best fit parameters are $N_H = 0.16\ 10^{22}$ cm^{-2}, $kT = 4.39$ keV, $Z = 0.35$ and $A = 1.24\ 10^{-3}$. This is better, but still not good because XSPEC sets σ to $\sqrt{n}$ instead of $\sqrt{H}$ in Eq.4.2. This procedure biases towards a lower model where there are few counts (at high energy particularly), because the low points have an artificially smaller σ. It also tends to blur the spectral features (like the Fe K line) because of the binning.

The bottom figure applies the C statistic (Eq.4.4) on the original channels. This is the appropriate statistic in that case. The best fit parameters are $N_H = 0.10\ 10^{22}$ cm^{-2}, $kT = 5.58$ keV, $Z = 0.29$ and $A = 1.31\ 10^{-3}$. The value at minimum is $S_{min} = 1261.0$. The model spectrum at high energy is more reasonable.

4.3 Refining the model

Even with the right statistic, the model obtained in Sect.4.2 is clearly a bad fit to the data. So refining it is in order. The ultimate model proposed by Pietsch et al. (2001) is the sum of three coronal spectra absorbed by different column densities but with the same metallicity Z. Abundances in the absorbing gas are assumed to be Galactic in all components.

$$M = \sum_{i=1}^{3} \text{phabs}(N_{Hi})\ A_i\ \text{mekal}(T_i, Z) \qquad (4.6)$$

Z is an example of linked parameters (it is forced to be the same in all components). Each component has three specific parameters (N_{Hi}, A_i, T_i).

To judge whether a refinement is significant, the question to answer is: is the model without this refinement (which is a subset of the refined model) significantly less likely than the refined model to generate the data set at hand ? Say the refined model has p additional parameters. S_0 is the minimum value of the statistic without refinement. S_p is the minimum value of the statistic with refinement (always smaller, because the simple model is a particular case of the refined one). Then, in the hypothesis where the true model is the simple one (without refinement), it was shown (Lampton et al., 1976; Cash, 1979) that the difference $S_0 - S_p$ follows a χ^2 law with p degrees of freedom, for any of the three statistics presented in Sect.4.2. Reciprocally, it may be said that the refined model is significant at confidence level α if $S_0 - S_p > \beta$, where $\text{Prob}(\chi^2 > \beta, p) = \alpha$.

This procedure is illustrated in Fig.15. All fits use the C statistic (Eq.4.4). The top figure introduces a second coronal component. The best fit parameters
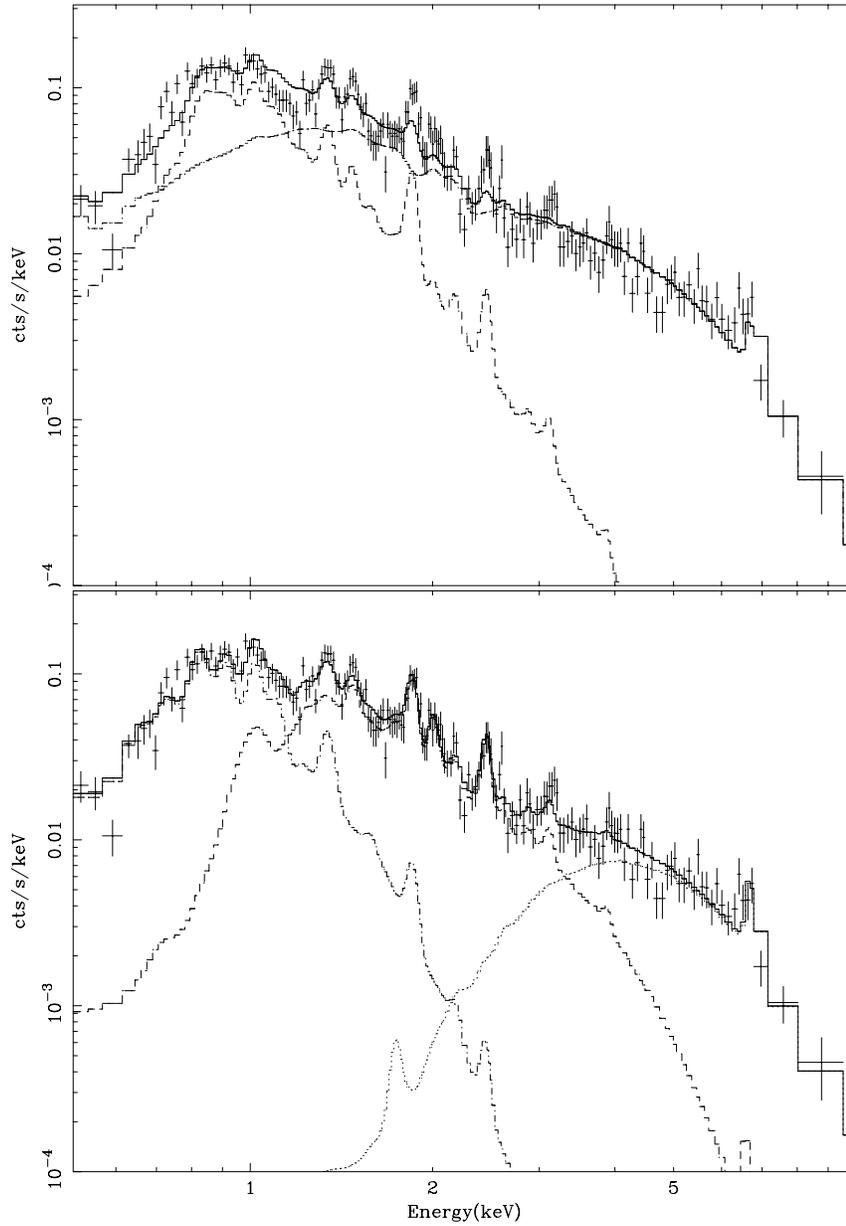
Fig. 15. Refining the fit of Fig.14 (bottom). The first fit (above) has two thermal components with different absorption. The second fit (below) has three. Both use the C statistic and show significant improvement over the simpler model. The dashed, dot-dashed and dotted lines show the contributions of the individual components. See Sect.4.3 for details.

are $N_{H1} = 0.75\ 10^{22}$ cm^{-2}, $kT_1 = 0.59$ keV, $A_1 = 8.4\ 10^{-4}$, $N_{H2} = 0.20\ 10^{22}$ cm^{-2}, $kT_2 = 14$ keV, $A_2 = 8.6\ 10^{-4}$ and $Z = 1.14$. The value at minimum is S_{min} = 875.3. This is clearly a much better representation of the data, and indeed the difference in S_{min} is 385.7, whereas 11.35 would be significant at 99% confidence for 3 additional parameters.

The bottom figure introduces a third coronal component. The best fit parameters are $N_{H1} = 0.66\ 10^{22}$ cm^{-2}, $kT_1 = 0.28$ keV, $A_1 = 3.5\ 10^{-3}$, $N_{H2} = 1.70$ 10^{22} cm^{-2}, $kT_2 = 0.92$ keV, $A_2 = 4.4\ 10^{-3}$, $N_{H3} = 10.6\ 10^{22}$ cm^{-2}, $kT_3 = 8.7$ keV, $A_3 = 1.4\ 10^{-3}$ and $Z = 0.64$. The value at minimum is $S_{min} = 686.8$. This is again a much better representation of the data (particularly of the lines), and indeed the difference in S_{min} is 188.5, much larger than 11.35. Of course now the model without refinement against which to judge progress is the intermediate two temperature model.

The astrophysical interpretation is that the spectrum is a composite of several media at increasing temperature toward the very nucleus, blanketed by more and more local gas. Nothing prevents the adjunction of even more components, but the convergence of the minimisation engine is then very slow and local minima are numerous.

The framework described above does not naturally allow to compare unrelated models. It is nevertheless commonplace to apply qualitatively the same criterion, favoring the model with lower S_{min} after accounting for the number of parameters of both models. This is more or less similar to assigning identical prior probabilities to both models in a Bayesian framework (Sect.4.5).

4.4   Confidence intervals and contours

The best fit parameters found by minimising the statistic are not the true values (even if the underlying model is indeed correct, like in a Monte-Carlo simulation). Because of the noise, another set of parameters (hopefully close to the true one) will always give a slightly smaller value of the statistic.

If the underlying model is correct, then the difference $S_{true}(p_1, ..., p_q) - S_{min}$ as a function of the parameters deemed of interest $p_1, ..., p_q$ (the other parameters left free) follows a χ^2 law with q (the number of interesting parameters) degrees of freedom, for any of the three statistics presented in Sect.4.2 (Lampton et al., 1976; Cash, 1979). Reciprocally, it may be said that the probability that the true parameters are within the region defined by $S(p_1, ..., p_q) - S_{min} < \beta$ is Prob$(\chi^2 < \beta, q)$.

This is applied on Fig.16 to the final fit found in Sect.4.3 (Fig.15, bottom). The top plot shows $S(kT_2) - S_{min}$. The 90% confidence interval on that single parameter is obtained by cutting at 2.70 (full line), resulting in [0.861,0.980]. Fig.8 illustrates the difference between models at both extremes of the confidence interval on density in the context of the density-sensitive Heα triplet.

The bottom plot shows isocontours of $S(kT_1, kT_2) - S_{min}$ at 2.30, 4.61 and 9.21, corresponding to 68%, 90% and 99% confidence contours on 2 parameters. The dashed line on the top figure corresponds to $S(kT_2) - S_{min} = 4.61$ (the middle
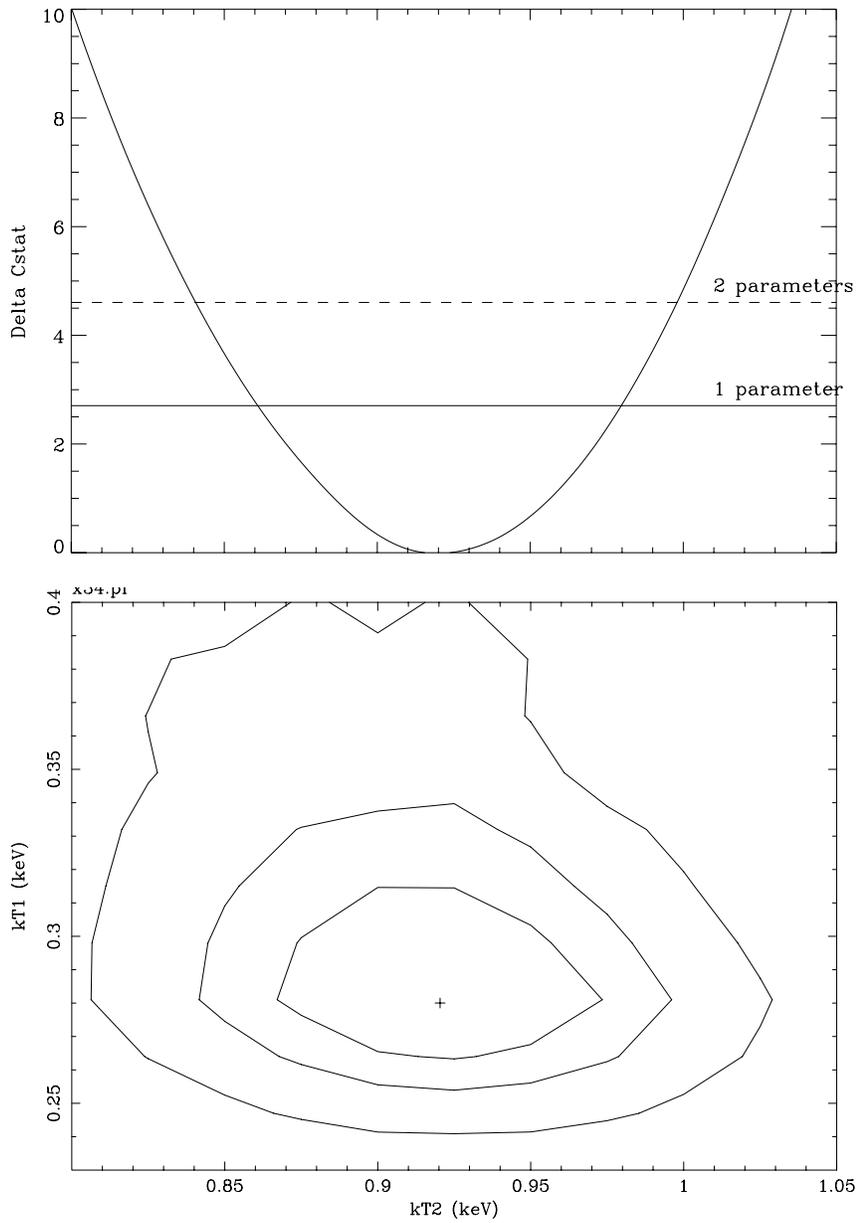
Fig. 16. Uncertainties on parameters for the model fit of Fig.15 (below). 90 % confidence interval on k T_2 (above), considering either that parameter alone or two parameters. 68 %, 90 % and 99 % confidence contours on k T_1 and k T_2 (below). The + is the best fit. The angles are due to the relatively poor sampling. See Sect.4.4 for details.

contour below). It is larger than the 90% confidence interval on kT_2 alone because it also incorporates a constraint on kT_1. Similarly, the region defined on the bottom plot by two vertical lines at $kT_2 = 0.861$ and 0.980 (unbounded in kT_1) is also a 90% confidence area (but less compact than the middle contour).

The validity of the χ^2 law for determining confidence regions starts breaking down when they approach infinity or a hard limit (0 for kT) on a parameter, or when secondary regions (around secondary minima) appear. A detailed discussion of those issues is given by Feldman and Cousins (1998).

4.5   The Bayesian approach

The framework described above does not allow in a quantitative sense to compare unrelated models (for example a coronal model and a comptonisation model). Also it does not easily allow to factor in prior information on parameters (other than hard limits).

This is in principle possible within the framework of Bayesian theory, which writes an explicit probability law on a set of models M_i:

$$p(M_i|D) \;=\; K\,p(M_i)\,p(D|M_i) \tag{4.7}$$

where $p(M_i)$ is the prior probability of model M_i (with respect to the other models) and $p(D|M_i)$ is the probability of obtaining the data set at hand if model M_i is true (this is the likelihood, as used in Eq.4.4). K is just a constant such that the sum over i is 1. Procedures also exist to address parametric models.

This formalism has been very little used in spectral X-ray analysis, but a good example of how it works is given by Gregory and Loredo (1992) in the context of timing analysis.

The main difficulty of Bayesian theory is how to set the prior probability on models (or parameters within models) when no information already exists (*i.e.* how to start the process). There is no general rule, so this is often the subject of debate.

4.6   Other examples

One important limitation of current imaging spectrometers is their energy band (limited to 10 keV or so). Several important classes of X-ray sources (most conspicuously, accretion-powered compact objects) emit a lot at higher energy, and important spectral characteristics (such as the thermal cut-off, if any) appear above 10 keV. In that case it is crucial to extend the energy band.

Another case when this is essential is highly absorbed sources. For example, it was shown on the basis of *ASCA* and *OSSE* data that NGC 4945 is brighter by a factor 30 in hard X-rays than below 10 keV (Done et al., 1996). This is because what we see in X-rays is only that small fraction of the original radiation which is scattered by a halo of ionised gas surrounding the nucleus. The primary spectrum is completely absorbed and shines through only at higher energy. Fig.17 shows a more recent *RXTE* spectrum of the same source (Madejski et al., 2000).

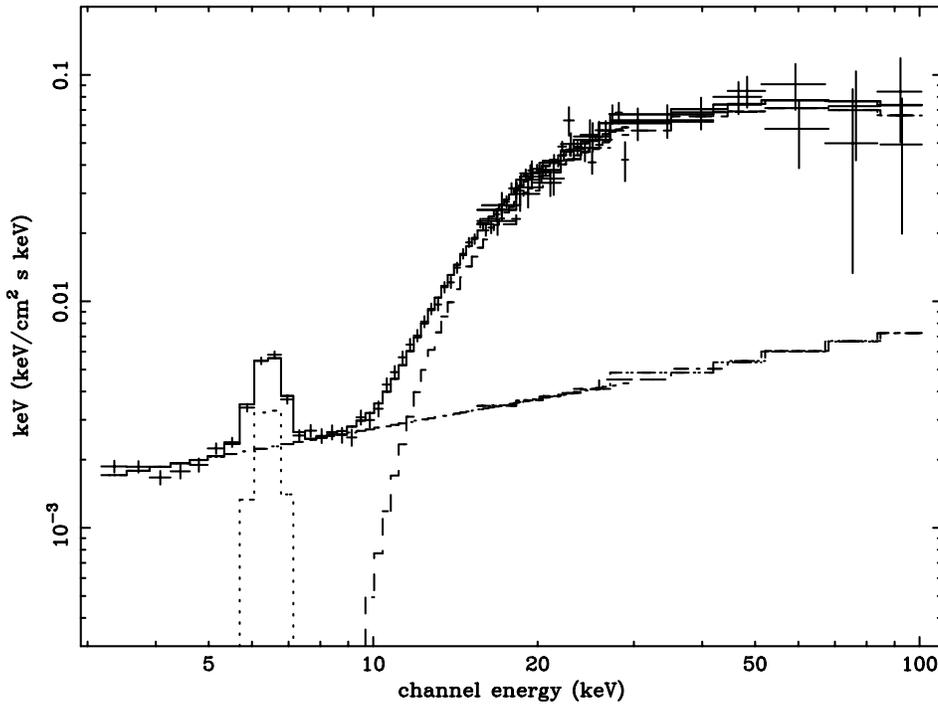**RXTE PCA and HEXTE unfolded spectrm of the Seyfert 2 galaxy NGC 4945**

Fig. 17. RXTE spectrum of the active galactic nucleus NGC 4945 (Madejski et al., 2000). Below 10 keV the direct spectrum is entirely absorbed surrounding molecular gas, with $N_H \simeq 4.5\ 10^{24}$ cm^{-2}. This illustrates the importance of a broad spectral band. What is seen below 10 keV is scattered light on tenuous gas above and below the nucleus.

4.7 Inversion procedures

The generic term of inversion refers to non-parametric methods using a regularisation term to converge on a solution. The same statistics as in Sect.4.2 are used to describe the quality of the fit to the data.

The first application is called deconvolution. It aims at removing the secondary maxima in the response function, setting the spectrum straight (*i.e.* removing the effect of the effective area), and narrowing the spectral response (this is possible only in the large signal to noise limit). The way to do that is to leave the intensity $M(E)$ in all energy bins free, but prevent unphysical fluctuations via the regularisation term.

Mathematically, this is done by minimising a function

$$L(M) = S(H, D) + \lambda\, Q(M) \qquad (4.8)$$

where $H(C)$ is related to $M(E)$ via Eq.2.2 and $S(H, D)$ is any of the statistics
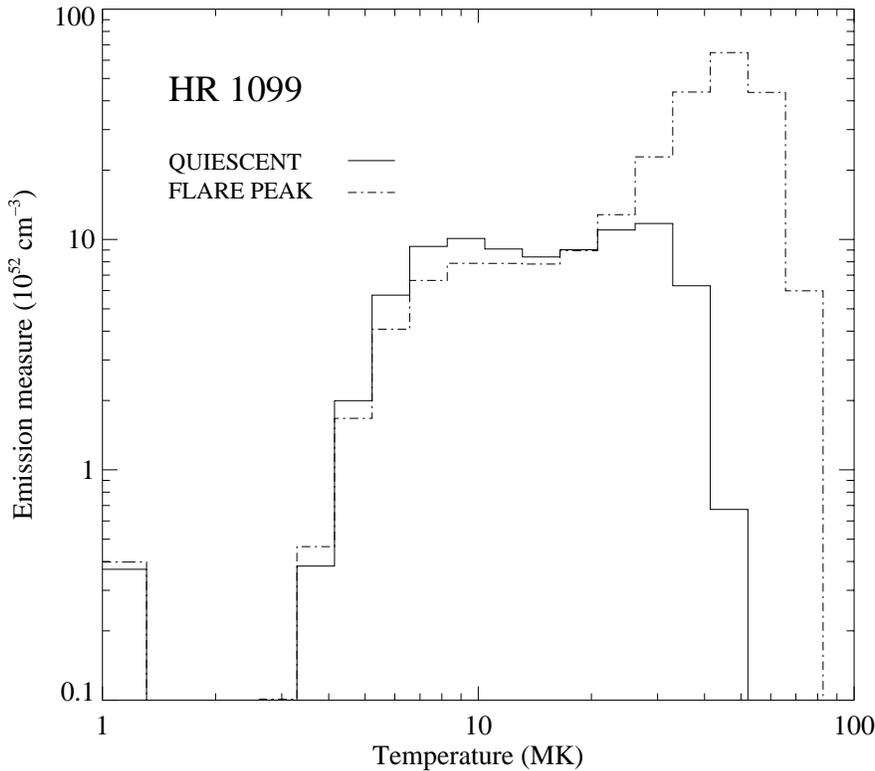
Fig. 18. Modelling the *XMM-Newton* RGS observations of the corona of HR 1099 (Audard, Güdel and Mewe, 2001). The spectrum is modelled with a continuous distribution of temperatures, weighted with the differential emission measure $d(n_e n_H V)/d(lnT)$. The full line is the distribution in quiescence. The dashed line is the distribution during a flare, showing a new hot component on top of the quiescent distribution.

described in Sect.4.2. λ is a free parameter (Lagrange multiplier). The most popular choices for the regularisation term Q are a quadratic tridiagonal form limiting local curvature, or the maximum entropy prescription

$$Q(M) = \sum_E p(E)\ln(p(E)) \quad \text{with} \quad p(E) = K\frac{M(E)}{r(E)} \quad \text{and} \quad \sum_E p(E) = 1 \quad (4.9)$$

where $r(E)$ is a reference spectrum chosen to be as non-informative as possible, whose main aim is to reduce the large dynamic range between low and high energies if $M(E)$ is expressed in photons.

This has not been used very much in X-ray spectroscopy, mostly because it does not directly provide hard numbers, and we are usually able to propose a

reasonable model on the basis of physical arguments. Deconvolution is mostly
good at qualitative detection of features when one has no idea of what to look for.
A hands-on discussion of deconvolution in the context of hard X-ray astronomy is
given by Bouchet (1995).

Another possible use of inversion methods is to apply them not on the spectrum
itself but on a continuous parameter in a spectral model. For example, spectra of
stellar coronae are often represented as a sum (mathematically discrete, but con-
ceptually continuous) of many thermal components as a function of temperature.
The normalisation of each component (the differential emission measure) is the
quantity of interest. Fig.18 (Audard et al., 2001) shows an example.

Fig.18 was actually obtained via model-fitting (the differential emission mea-
sure was parameterised as a Chebyshev polynomial). But an inversion procedure
might be applied as well, following Eq.4.8, by minimising $L(\mathbf{A})$:

$$M(E) \quad = \quad \sum_i A_i \, \mathrm{mekal}(T_i, Z) \tag{4.10}$$

$$L(\mathbf{A}) \quad = \quad S(H, D) \, + \, \lambda \, \mathbf{A}^t \, Q \, \mathbf{A} \tag{4.11}$$

where a quadratic regularisation has been assumed.

5   Complications

5.1   More specific models

It is often necessary to go beyond the simple radiation models (which assume
a homogeneous emitting medium), and account for the geometry of the source.
Standard examples (many already coded within XSPEC or SPEX) are

- multi-temperature models, where the differential emission measure is ob-
  tained from hydrodynamics: supernova remnants in the Sedov phase, opti-
  cally thick accretion disks, cooling flows in galaxy clusters.

- shock models, where electrons are progressively heated and heavy elements
  ionised behind a shock: young supernova remnants

- reflection models, where a hard continuum is partly reflected on an accretion
  disk or gas surrounding the black hole: active galactic nuclei

Such models may be used in exactly the same way as the simpler models, they
simply have different parameters (and take longer to compute !).

5.2   Spectro-imaging

The spectral analysis of extended sources (or clustered point sources) is often a
tricky business. On some instruments the point spread function $O_E(r)$ of the tele-
scope (the image obtained when a point source is observed) depends significantly
on energy (it usually has broader wings at high energy). This was particularly

a problem for *ASCA*. Then the spectrum extracted in an area close to a bright source is more contaminated by the bright source at high than at low energy and an additional term must be added to background spectrum (for Eq.2.2):

$$B(C) = B_0(C) + \int_E \int_\mathbf{u} R(E, C) A_{\mathbf{u}_0}(E) M_0(E) O_E(|\mathbf{u} - \mathbf{u}_0|) dE d\mathbf{u} \qquad (5.1)$$

where the spatial integration (on $\mathbf{u}$) runs over the extraction area, $\mathbf{u}_0$ is the position of the bright source and $M_0(E)$ is its spectrum. $B_0(C)$ is the background unrelated to the bright source.

Even if the point spread function does not depend significantly on energy, care must be taken when the spectrum of an extended source varies on scales comparable to the spatial resolution. The observed brightness $M(\mathbf{v}, E)$ will be related to the true one $B(\mathbf{u}, E)$ via

$$M(E, \mathbf{v}) = \int_\mathbf{u} B(E, \mathbf{u}) O_E(|\mathbf{v} - \mathbf{u}|) d\mathbf{u} \qquad (5.2)$$

where the spatial integration runs over the whole source (this assumes a constant point spread function over the source). In that way lines emitted only at particular places may creep up into spectra of other places.

When a spectrum is extracted in an area which is not much smaller than the field of view, an additional difficulty is that the effective area decrease toward the edges of the field of view (the vignetting) often depends on energy as well, so that the source regions within the extraction area are not weighted in the same way at all energies. Finally, the point spread function also changes within the field of view (it is better at the centre). The most general formula relating the model observation to the sky brightness (generalising Eq.2.2 for an extended source) is:

$$H(C, \mathbf{v}) = \iint_{E, \mathbf{u}} R_\mathbf{v}(E, C) Q_\mathbf{v}(E) A_\mathbf{u}(E) B(E, \mathbf{u}) O_E(\mathbf{v}, \mathbf{u}) dE d\mathbf{u} + B(C, \mathbf{v})$$

where the response matrix may depend on the position on the detector $\mathbf{v}$. The quantum efficiency $Q_\mathbf{v}(E)$ also depends on $\mathbf{v}$ rather than the direction in the sky $\mathbf{u}$. This is why it was separated from the effective area of the telescope $A_\mathbf{u}(E)$ here.

5.3 Spectro-timing

In general time is known with a finite precision, but is not completely redistributed in an instrument like energy or position is. This simplifies a lot the analysis. Time-resolved (or phase-resolved, for pulsars) spectroscopy means accounting for several independent spectra at the same time (related only via the model). Frequency-resolved spectroscopy, *i.e.* Fourier transform of the light curve in a number of spectral bins, is also possible.

References

Arnaud, K. A.: 2001, *XSPEC*, http://heasarc.gsfc.nasa.gov/docs/xanadu/xspec/

Arnaud, M. and Rothenflug, R.: 1985, *A&AS* **60**, 425

Audard, M., Güdel, M., and Mewe, R.: 2001, *A&A* **365**, L318

Bouchet, L.: 1995, *A&AS* **113**, 167

Cash, W.: 1979, *ApJ* **228**, 939

Done, C., Madejski, G. M., and Smith, D. A.: 1996, *ApJ* **463**, L63

Feldman, G. C. and Cousins, R. D.: 1998, *Phys. Rev. D* **57**, 3873

Fujimoto, R., Tanaka, Y., Inoue, H., et al.: 1995, *PASJ* **47**, L31

Gregory, P. C. and Loredo, T. J.: 1992, *ApJ* **398**, 146

Güdel, M., Audard, M., Briggs, K., Haberl, F., Magee, H., Maggio, A., Mewe, R., Pallavicini, R., and Pye, J.: 2001a, *A&A* **365**, L336

Güdel, M., Audard, M., den Boggende, A. J., et al.: 2001b, in R. Giacconi, L. Stella, and S. Serio (eds.), *X-ray astronomy 2000 (Palermo)*, ASI Conf. Ser. **234**, p. 73, San Francisco: ASP

Kaastra, J. S.: 2001, *SPEX*, http://www.sron.nl/divisions/hea/spex/

Lampton, M., Margon, B., and Bowyer, S.: 1976, *ApJ* **208**, 177

Madejski, G., Zycki, P., Done, C., Valinia, A., Blanco, P., Rothschild, R., and Turek, B.: 2000, *ApJ* **535**, L87

Pietsch, W., Roberts, T. P., Sako, M., et al.: 2001, *A&A* **365**, L174

Porquet, D. and Dubau, J.: 2000, *A&AS* **143**, 495

Porquet, D., Mewe, R., Dubau, J., Raassen, A. J. J., and Kaastra, J. S.: 2001, *A&A* **376**, 1113

Rasmussen, A. P., Behar, E., Kahn, S. M., den Herder, J. W., and van der Heyden, K.: 2001, *A&A* **365**, L231

Rasmussen, A. P. et al.: 2002, *in preparation*

Sako, M., Kahn, S. M., Behar, E., et al.: 2001, *A&A* **365**, L168

Smith, R. K., Brickhouse, N. S., Liedahl, D. A., and Desai, P.: 2000, *APEC/APED*, http://hea-www.harvard.edu/APEC/

Wilms, J., Allen, A., and McCray, R.: 2000, *ApJ* **542**, 914