Should we keep columns adjacent to bad columns for spectral analysis with the EPIC-MOS ?

EPIC-SAP-JBB-02, Version 1.1

Jean BALLET

SAp, Bt 709, CEA Saclay, 91191 Gif sur Yvette Cedex, France
Email: jballet@cea.fr

June 30th, 2005

1 Introduction

Two years ago it was discovered (Ref. 1) that several columns per MOS CCD return an energy which is too low by a large amount (> 100 eV). The way this was addressed at the time was by declaring those columns dead in the BADPIX CCF. The SAS software then flags all events on those columns as ON_BADPIX. They are rejected by the standard #XMMEA_EM filter and are not included in the pipeline products.

The events on columns immediately adjacent to those "dead" columns are flagged as CLOSE_TO_DEADPIX. They are NOT rejected by the standard #XMMEA_EM filter and ARE included in the pipeline products. However, they are rejected by the stricter FLAG==0 selection, which is recommended in *emchain* for spectral analysis. This is because a fraction of all events are multiple (extending over several pixels) and therefore an event occurring on a pixel next to a bad column can return too low energy if part of the charge was released in that bad column.

Recently it was discovered that a number of additional columns show the same problem. Apparently they started returning too low energy after cooling (rev 533, Nov. 2002). Evidence exists (at least for some columns) that the energy decrement worsens with time. Work is in progress by D. Baskill to try to recover the energy in those columns, calibrating the decrement on CalClosed observations. But today a reliable recipe does not exist yet. So we are moving toward flagging those additional columns dead as well.

Together with the previously known "dead" columns and a few bright columns, the number of bad columns per CCD now becomes uncomfortably large. For example, it is around 15 in MOS 1 CCD 1. This is only 2.5% of the columns, but if the two adjacent columns next to a bad one are rejected as well (as in FLAG==0) it climbs to 7.5%. If the number of such bad columns keeps increasing, it will soon become a major issue.

As a first step, I decided to look in more detail into what happens next to a "dead" column, in order to decide whether it is really worth rejecting such events for spectral analysis. An obvious reason why it might not be so critical is that the fraction of multiple events is not that large (Fig. 1). In addition, the vertical doubles (not touched by the proximity of a bad column) form the majority of multiple events.

2 Simulation method

The idea is to start with normal data and try to simulate what a dead column (or a column shaving off part of the energy) would change. To that end I can use the distribution

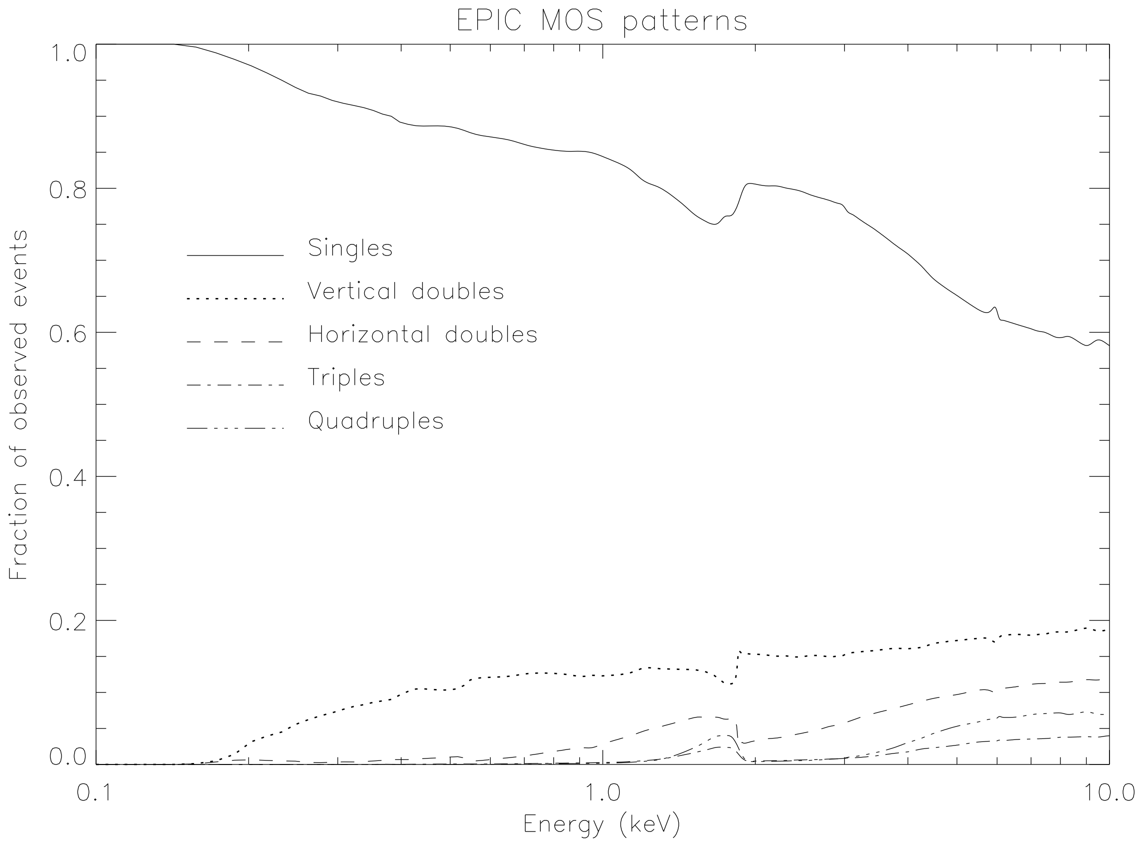


Figure 1: Calibrated pattern distribution for EPIC MOS (here MOS 1 CCD 1, but they are all very similar) as a function of measured energy.

of charge within the event, known via the PATTERN (shape of the event) and the ENERGYE1 / ENERGYE2 fields, which give the charge on the central pixel and the total charge on the other pixels above threshold in the event. The description of the patterns is reproduced in Fig. 2.

There is no information in the data, for events larger than double, on how charge is distributed between the pixels making up ENERGYE2. In this report I will simply assume it is equally distributed. One could probably do better (randomly assigning charge to the pixels) but I do not believe this is a significant issue, particularly as triple and quadruple events are less numerous than horizontal doubles.

All the analysis in this report is done on events with PATTERN ≤ 12 only. All larger events are ignored. Recorded events with $13 \leq$ PATTERN ≤ 25 are very few anyway. A sizable fraction of events at high energy are much larger (appearing as PATTERN 31), but those are flagged out by #XMMEA_EM and they will never migrate to PATTERN ≤ 12 even if close to a dead column.

2.1 Events next to a dead column: losing the minor part of an event

When an event next to a "dead" column encompasses pixels ON the dead column, those pixels will lose energy while the major part of the event will remain untouched. I model this by considering that the pixels ON the dead column are entirely lost. In theory, one could model an energy loss in each pixel rather than cutting them off entirely. I do that for simplicity and because it maximizes the effect, so I get an upper limit to the spectral change it incurs.

This obviously induces pattern migration as well as energy loss. For example, a PAT-

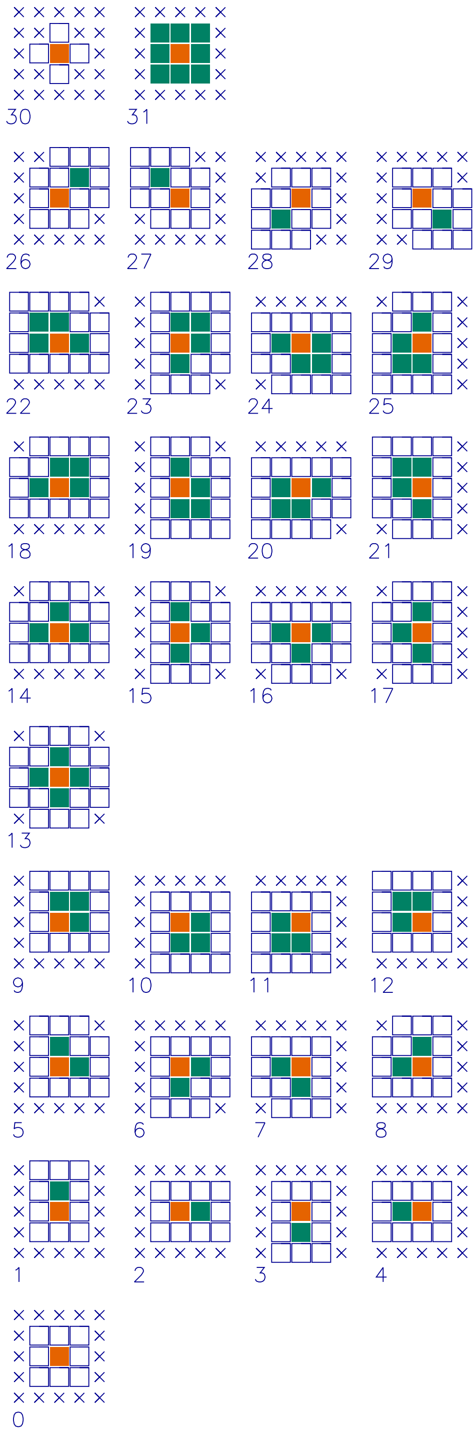


Figure 2: Geometry of EPIC-MOS patterns. The red (central) square is the position of the pixel with maximum charge, given by ENERGYE1. The green pixels are the other pixels above threshold in the event. Their total charge is ENERGYE2.

TERN 2 double event to the left of a column will become a single event with only ENERGYE1. A PATTERN 6 triple event will become a PATTERN 3 double event with ENERGYE2 halved.

2.2 Events on a dead column: losing the major part of an event

The fate of events occurring ON a "dead" column is somewhat more uncertain. I consider two simplified cases:

- The column is not entirely dead, but "grey". This means that charge measured in that column is not lost, but only shifted to lower energy. I model this simply by considering that events centered on a "grey" column are still recorded there (the pixels don't lose enough charge to change the peak position). Such events are then entirely removed by SAS flagging.
- The column is entirely dead (all the charge is lost). In that case, I assume that those pixels on the dead column disappear and the other ones remain. For example, a PATTERN 2 double event on the dead column will become a single event with RAWX shifted by +1 and ENERGYE1 = ENERGYE2 (original event). A PATTERN 11 quadruple event will become a PATTERN 3 double event with RAWX shifted by -1 and ENERGYE1+ENERGYE2 = 2/3 ENERGYE2 (original event). Such events (originally with pixels above threshold on neighbouring columns) are flagged as CLOSE_TO_DEADPIX by the SAS but not removed.

The end result is that a "dead" column leaves more events than a "grey" one (one for each multiple event on the dead column except vertical doubles). I am not sure this is true (and it is hard to check on real data). In any case, most of the columns we are interested in are not entirely dead, so the "grey" model is probably closer to the truth, but I keep the "dead" one also as an extreme case.

3 Data base

To investigate the issue, I have selected a number of data sets (Table 1).

Three were selected for simulating the effect on one column (Sect. 4). They are observations of bright (piled-up) point sources, in order to get a large number of events with identical spectrum over the entire window. Two of them (GX 13+1 and 4U 1624-49) are hard (very absorbed) sources to maximize the effect, whereas the third one (3C 273) is a softer (much less absorbed) source more representative of the average source.

Two were selected for simulating the effect on a source (Sect. 5). The first one is a long observation of a source at the pile-up limit (AU Mic). This source is rather soft (Fig. 6), so the effect I am looking for could be quite small. The second one was taken on a harder

Table 1: List of observations used. In GX 13+1, 3C 273 and 4U 1624-49, the central disk where the spectrum is modified by pile-up (within the radius defined in the "Piled-up" column) was excluded. The number of events ("Num events" column) is that after spatial selection. In AU Mic and NGC 4151 no spatial selection was performed.

Rev_Obs	Exposure	Mode	Piled-up	Num events	Comment
0061_0122340201	19 ks	FF	< 98"	703158	GX 13+1
0094_0126700301	55 ks	SW	< 7"	440413	3C 273
0155_0111420101	54 ks	SW	< 1"	142716	AU Mic
0190_0112310101	29 ks	SW	< 1"	55357	NGC 4151
0216_0098610201	56 ks	FF	< 28"	617802	4U 1624-49

source (Fig. 7) also at the pile-up limit (NGC 4151) to check what happens in a more difficult case. The exposure is not as long, but the number of recorded events is large enough.

The data was processed with SAS 6.1.

4 Results on a single column

First, I consider how the spectrum of a column next to a “dead” column is affected. Keeping only a single column to investigate that would result in few events (even on a bright source), so I simulate that as if ALL events were next to a “dead” column. The first step is to remove the piled-up part of the source (where the spectrum is different for another reason). This results in what I call the “original” (unmodified) events list.

Then all multiple events are modified following the recipe given in Sect. 2.1, assuming they are all to the right of a dead column. In other words, I cut off that part of all events to the left of the event peak. This results in what I call the “next to grey” events list, because it simulates what happens next to a “grey” column as defined in Sect. 2.2. This list contains the same number of events as the “original” one, but lower total energy.

Now if the bad column is truely dead (as defined in Sect. 2.2), the right part of all events occurring ON that bad column will end up as additional events on the column to the right of the bad one (*i.e.* the column I am considering here). To simulate that, I generate a second events list (from the same “original” events list, because the exposure varies very little from one column to its neighbour), now assuming that all events are ON the dead column. I follow the recipe given in Sect. 2.2, keeping only the events which migrate toward the right. In other words, I keep only that part of all events to the right of the peak. I then ADD that “on dead” events list to the “next to grey” one to simulate what happens next to a truely dead column. This results in what I call the “next to dead” events list. This list contains of course more events than the “original” and the “next to grey” ones (all rightward multiple events appear twice). It contains approximately the same total energy as the “original” one, because there is no statistical difference between the leftward and rightward multiple events (so that the part of leftward events cutoff in the “next to grey” is balanced by the part of rightward events kept in the “on dead”), and I don’t consider events extending over more than two columns. Using the very same “original” events list to generate the “next to grey” and “on dead” means that the events in the “next to dead” events list are not all independent, but that is not a serious issue here.

I can then compare the three spectra: the original spectrum, the “next to grey” spectrum, and the “next to dead” spectrum. Figs. 3, 4 and 5 show the results. As expected, the effect is not very large beyond the maximum of the spectrum (never larger than 10%). The ratio of the modified to original spectrum (bottom figure) decreases with energy as the fraction of horizontal doubles, triples and quadruples (which move to lower energy) increases (Fig. 1). The main effect is at low energy where counts originally at higher energy accumulate. This is not very visible in a typical spectrum with a sizable number of true photons at low energy (like 3C 273), but conspicuous on spectra with very few true photons at low energy (like GX 13+1 and 4U 1624-49).

The peak near 1 keV in the dotted lines of Figs. 3 and 4 comes specifically from multiple events originally just below the Si edge near 2 keV (peak in horizontal doubles, triples and quadruples on Fig. 1). The dotted line goes back to 1 at low energy because multiple events on the neighbouring column (which peak above 1.5 keV) always keep at least half their energy. The full lines (spectrum next to a dead column, including low energy remains of events on the dead column) show of course a much stronger effect at

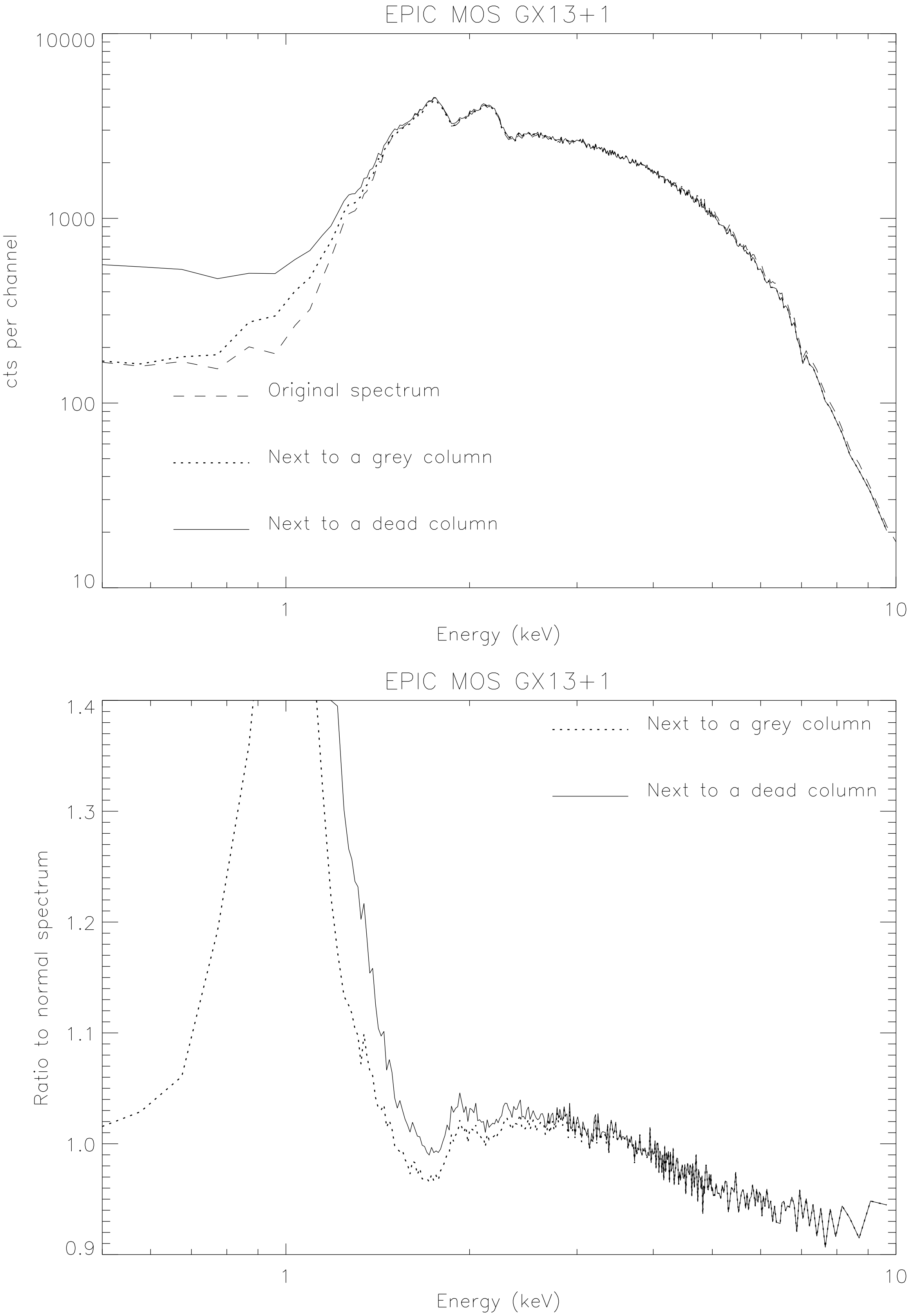


Figure 3: Observation of GX 13+1, with piled-up part excised. The dashed line on the top figure shows the total spectrum (PATTERN $<$ 13). The dotted line shows the same spectrum, after removing that part of all events (as in Sect. 2.1) located on the pixels to the left of the maximum (as if they occurred all to the right of a dead column). The full line shows the same spectrum as the dotted line, after adding the spectrum obtained by keeping only that part of all events (as in Sect. 2.2) located on the pixels to the right of the maximum (as if all events occurred on the dead column to the left of the column considered). The bottom figure shows the ratio between the modified spectra (dotted and full line on the top figure) and the original spectrum.

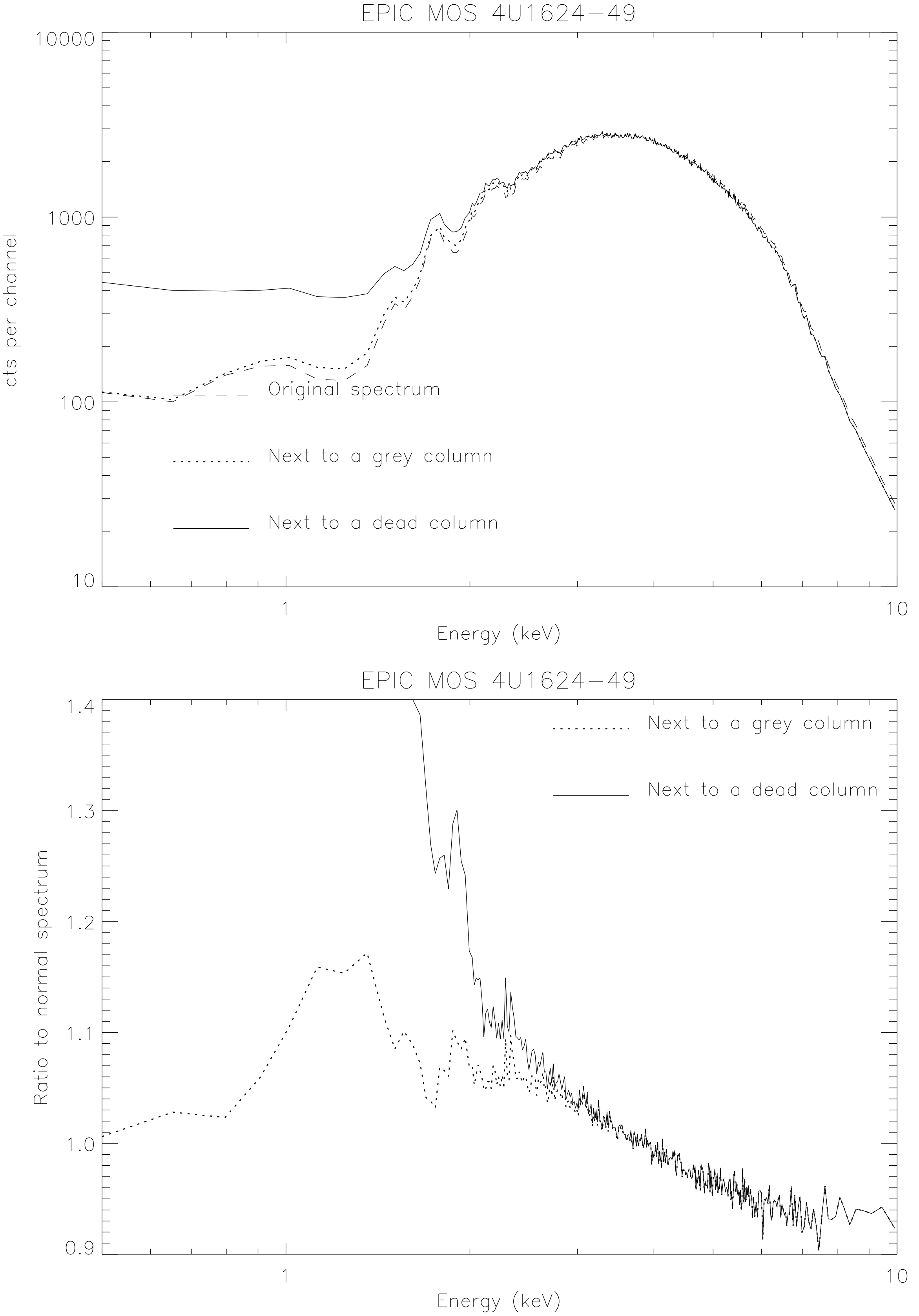
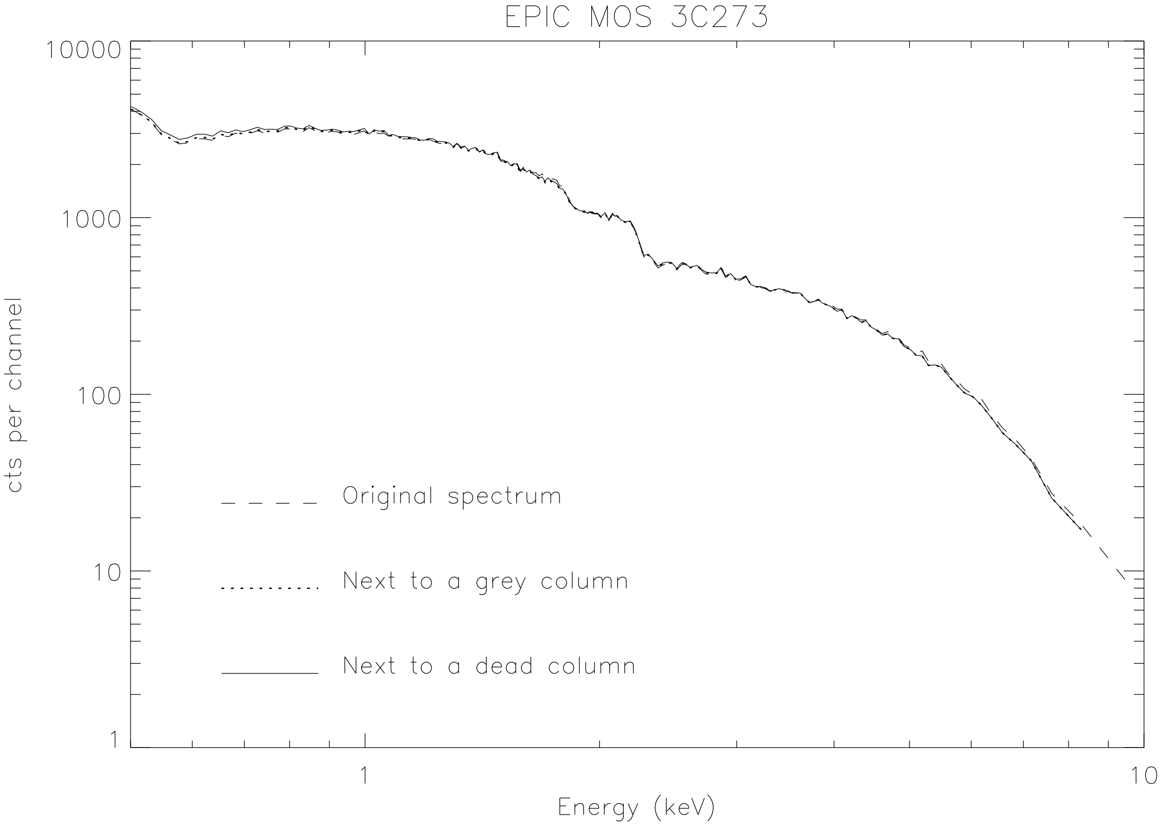


Figure 4: Observation of 4U 1624-49, with piled-up part excised. The figure is built in the same way as Fig. 3.



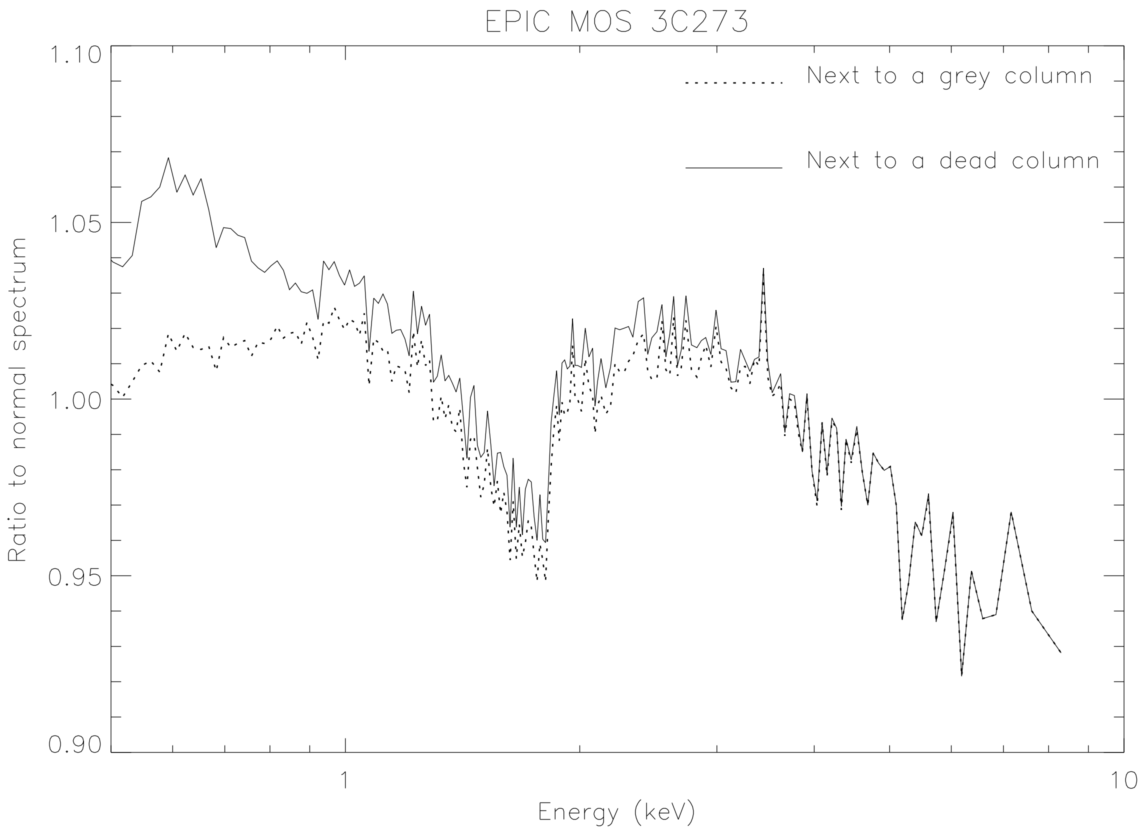


Figure 5: Observation of 3C 273, with piled-up part excised. The figure is built in the same way as Fig. 3. Note the very different vertical scale in the bottom plot.

Table 2: Model parameters for fits to the data in XSPEC over the 0.1 to 10 keV range, using the C statistic and no background. The model is wabs power (absorbed power law). The formal error bars are smaller than the last digit given. The second row is obtained assuming all events are next to a "grey" column (as in Sect. 2.1). The third row is obtained assuming all events are next to a "dead" column (as in Sect. 2.2). Because 3C 273 is much softer, the photon flux is compared in a broader band (starting at 0.1 keV) than for GX 13+1 and 4U 1624-49 (starting at 1 keV).

Data	N_H (cm^{-2})	Γ	Norm	1 to 10 keV rate
GX 13+1 Original	3.42 10^{22}	2.82	0.95	0.0963 ph/cm^2/s
GX 13+1 Next to grey	3.31 10^{22}	2.82	0.92	0.0959 ph/cm^2/s
GX 13+1 Next to dead	3.01 10^{22}	2.74	0.81	0.1002 ph/cm^2/s
Data	N_H (cm^{-2})	Γ	Norm	1 to 10 keV rate
4U 1624-49 Original	9.16 10^{22}	3.25	1.33	3.33 10^{-2} ph/cm^2/s
4U 1624-49 Next to grey	9.04 10^{22}	3.34	1.47	3.32 10^{-2} ph/cm^2/s
4U 1624-49 Next to dead	7.60 10^{22}	3.10	0.96	3.52 10^{-2} ph/cm^2/s
Data	N_H (cm^{-2})	Γ	Norm	0.1 to 10 keV rate
3C 273 Original	2.2 10^{19}	1.77	9.28 10^{-3}	5.44 10^{-2} ph/cm^2/s
3C 273 Next to grey	2.5 10^{19}	1.78	9.31 10^{-3}	5.41 10^{-2} ph/cm^2/s
3C 273 Next to dead	1.9 10^{19}	1.80	9.53 10^{-3}	5.95 10^{-2} ph/cm^2/s

low energy.

Table 2 shows the results in terms of a power law fit to the data. On a source emitting a large number of intrinsic low energy photons (like 3C 273), the effect is minimal, even in the extreme "next to dead" case. On a source emitting few intrinsic low energy photons (like GX 13+1 or 4U 1624-49) the effect is of course more significant. However even in that case the fit is not very strongly modified in the "next to grey" case, which is the most common.

5 Results on a source

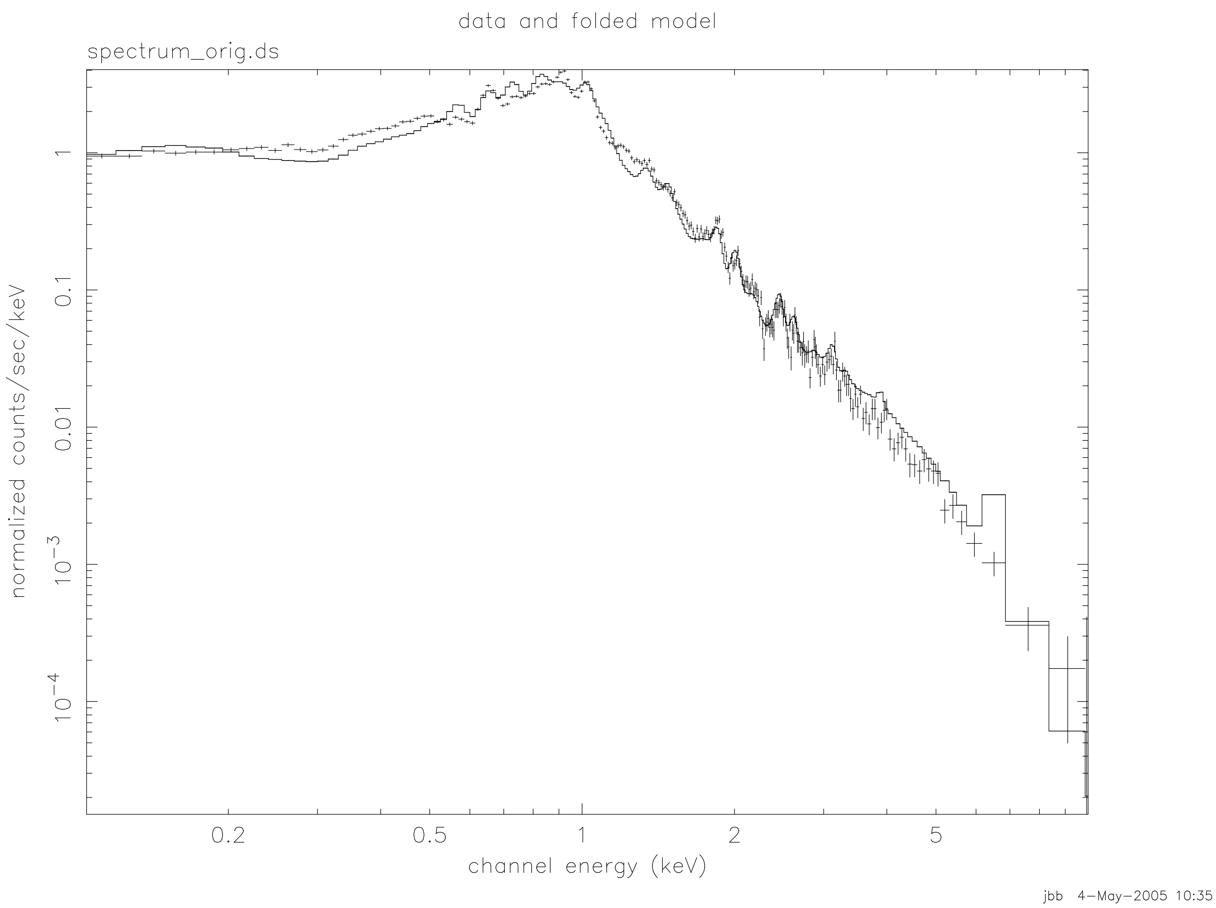
I now turn to what happens to the spectrum of a source, when a dead column crosses it. Of course the effect is less strong than in Sect. 4, because the source is not entirely contained in the dead column and its neighbours (the three columns centered on a source contain about 25% of the counts).

To simulate that, I look at data from two bright sources (to be able to see relatively small changes) not significantly piled-up (so that *arfgen* is expected to work reasonably well). I do not attempt to correct for the small pile-up fraction at all. Its main effect would be to reduce the count rate in the central pixel, which makes up less than 1% of the total counts.

I add a single dead column crossing the source right at its peak, and apply to the neighbouring events (on both sides) the recipe given in Sect. 2.1. Like in Sect. 4, I consider the two ideal cases of a "grey" or a "dead" column for modifying the events on the bad column. In the latter case I apply to the events on the "dead" column the recipe given in Sect. 2.2.

I then extract 4 spectra using evselect, in 15 eV bins:

- the original spectrum (untouched).
- the "grey" spectrum obtained with #XMMEA_EM when the bad column is "grey" and does not create new events on each side.
- the "dead" spectrum obtained with #XMMEA_EM when the bad column is "dead" and creates new low energy events on each side.



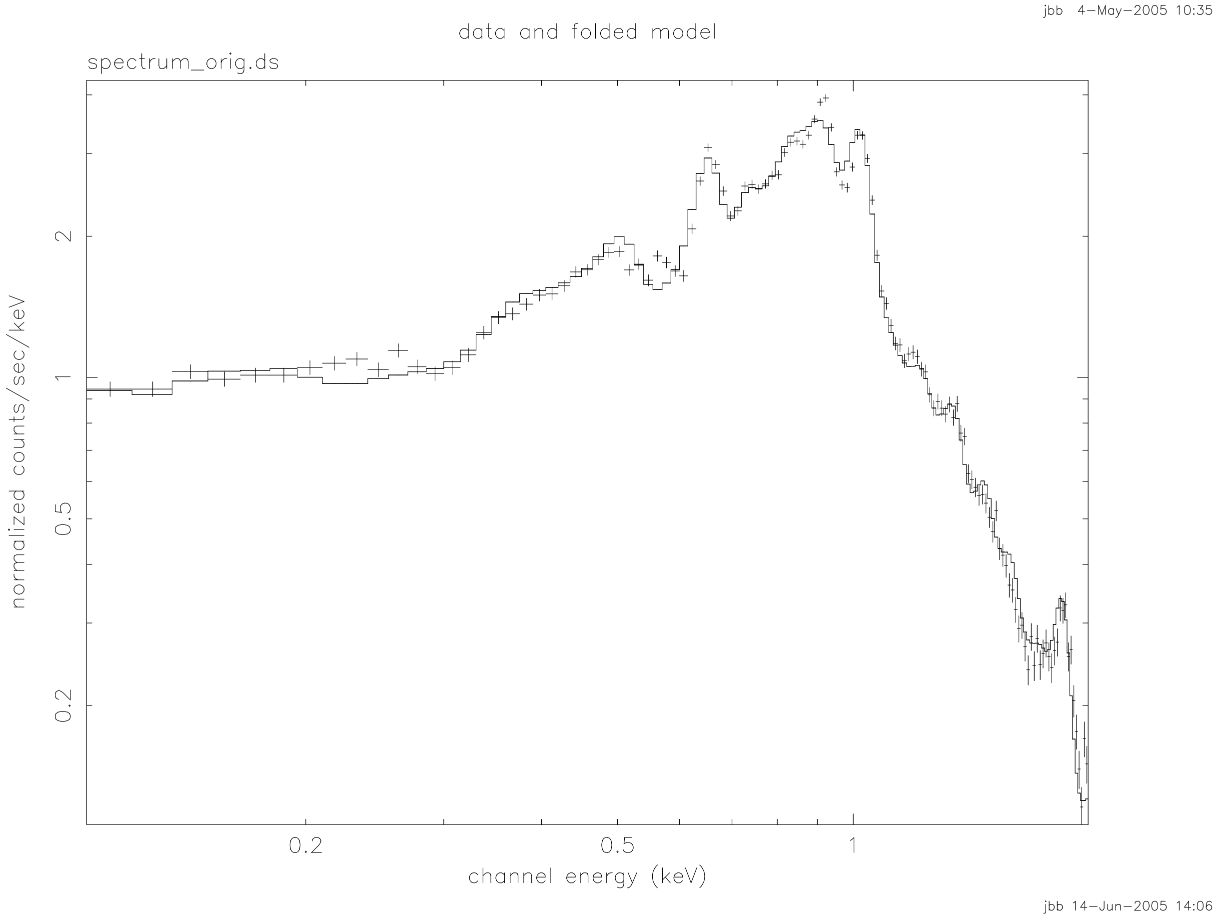
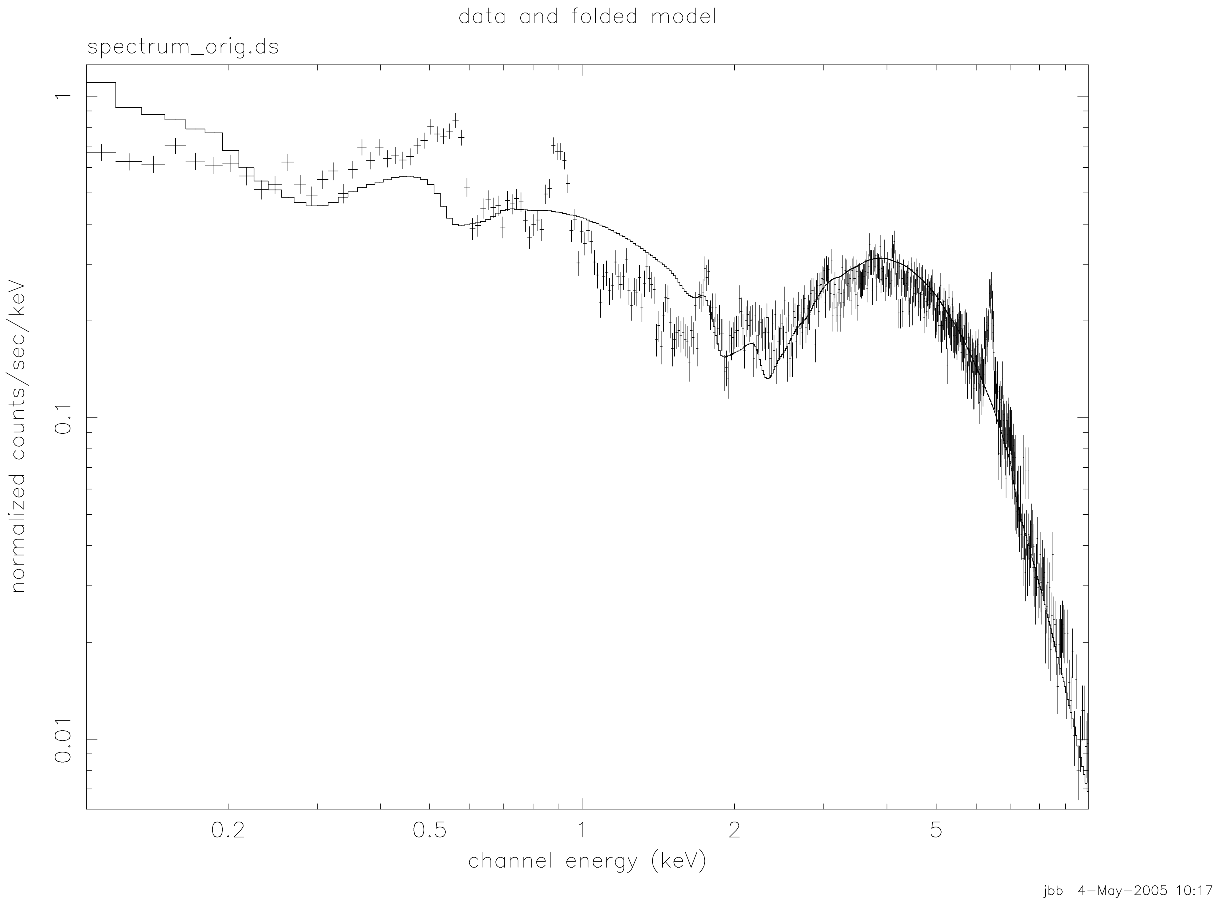


Figure 6: Spectrum of AU Mic, overlaid with the XSPEC spectral model used. Top: full range, wabs cemekl (absorbed multi-temperature plasma) emission model. Bottom: below 2 keV, vmekal (variable abundance plasma) emission model.



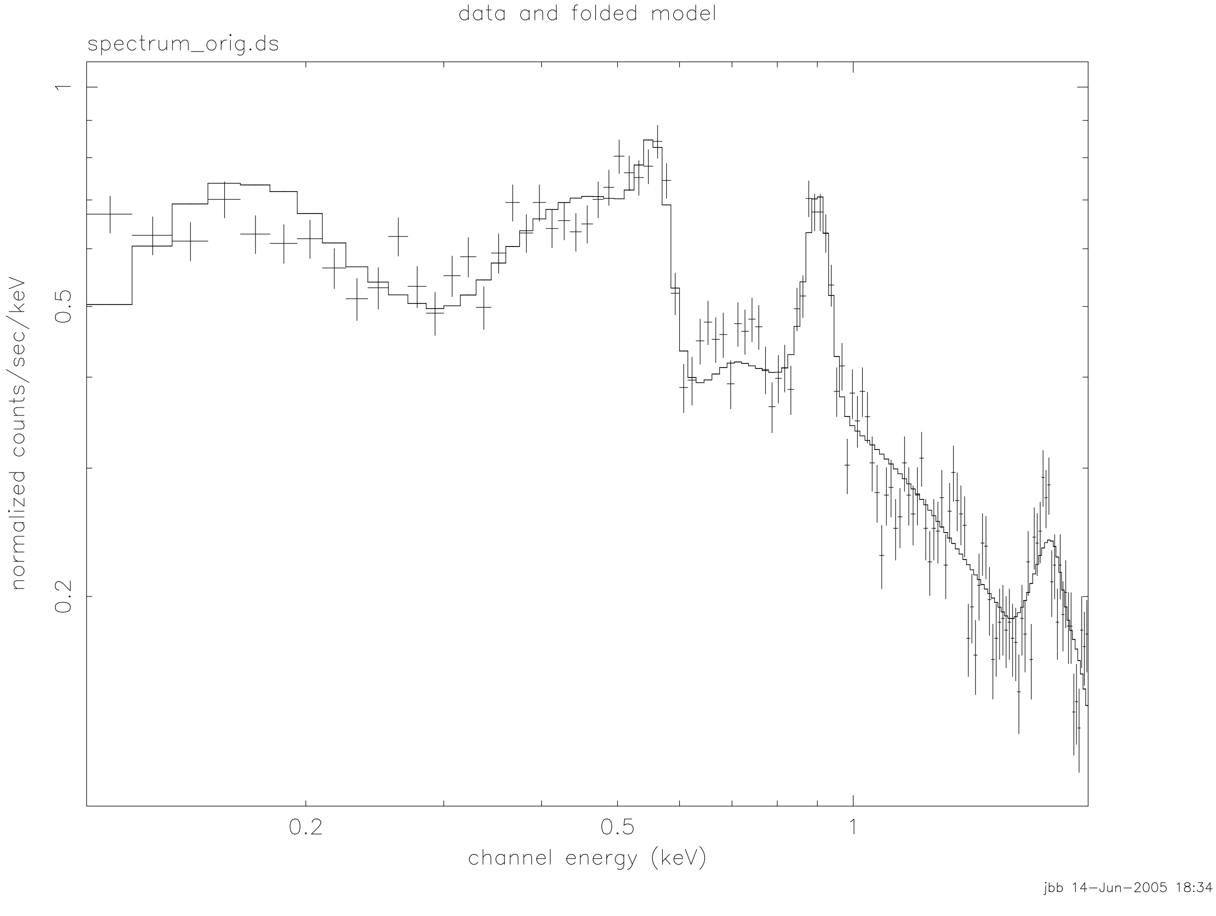


Figure 7: Spectrum of NGC 4151, overlaid with the XSPEC spectral model used. Top: full range, pcfabs powerlaw (partially absorbed power law) model. Bottom: below 2 keV, wabs powerlaw + gauss + gauss + gauss (absorbed power law + three gaussian lines) model.

Table 3: Model parameters for fits to the spectra in XSPEC over the 0.1 to 10 keV range, using the C statistic and no background. The model for AU Mic is wabs cemekl (absorbed multi-temperature plasma emission model). The model for NGC 4151 is pcfabs powerlaw (partially absorbed power law model). The formal error bars are smaller than the last digit given. The three bottom rows are obtained after declaring dead the column running through the center of the source. The "grey" row corresponds to the spectrum, obtained after removing that part of all CLOSE_TO_DEADPIX events located on the pixels of the newly "dead" column (as in Sect. 2.1). The "dead" row corresponds to the spectrum, obtained after cutting off the central part of events located on the newly "dead" column, building the spectrum of what remains of those events and adding that to the "grey" spectrum (as in Sect. 2.2).

Data (AU Mic)	N_H (cm^{-2})	α	T_{max}	Norm	0.5 to 10 keV rate
Original	$3.2\ 10^{19}$	0.09	2.85 keV	$8.0\ 10^{-3}$	$1.20\ 10^{-2}$ ph/cm^2/s
FLAG == 0	$3.9\ 10^{19}$	0.07	2.85 keV	$7.5\ 10^{-3}$	$1.16\ 10^{-2}$ ph/cm^2/s
#XMMEA_EM, grey	$3.9\ 10^{19}$	0.07	2.85 keV	$7.7\ 10^{-3}$	$1.19\ 10^{-2}$ ph/cm^2/s
#XMMEA_EM, dead	$3.3\ 10^{19}$	0.07	2.85 keV	$7.7\ 10^{-3}$	$1.19\ 10^{-2}$ ph/cm^2/s
Data (NGC 4151)	**N_H (cm^{-2})**	**Cov. frac**	**Γ**	**Norm**	**0.5 to 10 keV rate**
Original	$10.0\ 10^{22}$	0.944	1.81	$2.4\ 10^{-2}$	$7.33\ 10^{-3}$ ph/cm^2/s
FLAG == 0	$10.0\ 10^{22}$	0.943	1.82	$2.3\ 10^{-2}$	$7.12\ 10^{-3}$ ph/cm^2/s
#XMMEA_EM, grey	$9.9\ 10^{22}$	0.943	1.81	$2.4\ 10^{-2}$	$7.26\ 10^{-3}$ ph/cm^2/s
#XMMEA_EM, dead	$9.9\ 10^{22}$	0.942	1.82	$2.4\ 10^{-2}$	$7.31\ 10^{-3}$ ph/cm^2/s

- the spectrum obtained with FLAG==0 after modification. Since this removes the events on neighbouring columns altogether, it gives the same result in the "grey" and the "dead" cases.

In all cases I neglect the background (those are bright sources) and I use the same canned matrix (m1_110_im_pall_v1.2.rmf) adapted to spectra with all patterns, taken from the SOC web site (corresponding to the early days). *arfgen* is called to generate the effective area corresponding to the selection. Note that the "dead" and "grey" spectra have of course the same ancillary response file (they are identical from the SAS' point of view).

A difficulty arose because *arfgen* 1.65.16 (in SAS 6.1) does not remove the columns next to a bad column when FLAG==0 (SPR 2609, now corrected in *arfgen* 1.66). To make up for that I declared one more bad column on each side of the original one in the BADPIX extension.

The spectra are essentially undistinguishable from each other, so I don't show them all but I give spectral fitting results instead (Table 3). The spectral models do not represent the data very accurately over the full energy range (Figs. 6 and 7, top), but they capture the gross features of the continuum, which is expected to be most affected by the energy loss induced by the dead column.

The important point here is that the "grey" and "dead" spectra do not differ more from the original spectrum than the "FLAG == 0" spectrum. This indicates that the errors due to rejecting a large fraction of the source (via the imperfect PSF calibration) are as large as those incurred by keeping imperfect events.

Because the main effect is expected at low energy, I have also tried to model the spectra in XSPEC from 0.1 to 2 keV only (bottom of Figs. 6 and 7). I have used different spectral models to represent the data more accurately than over the entire energy range.

- For AU Mic, I used the vmekal (thermal emission with variable abundances) model, with kT = 0.65 keV, and variable abundances adjusted to N = O = 0.44, Ne = 0.64, Mg = 0.20, Si = 0.25 and Fe = Ni = 0.111. The other elements were left fixed to solar (no visible line in the spectrum).

- For NGC 4151, I used the wabs powerlaw + gauss + gauss + gauss (absorbed power law plus three lines), with $N_H = 1.1\ 10^{2O}$ cm^{-2}, $\Gamma = 2.26$, one broad (0.12 keV) line at 1.86 keV and two narrow lines at 0.56 and 0.90 keV.

The conclusion is exactly the same. No very strong variation of the parameters is observed between the four spectra, and the "grey" and "dead" spectra do not differ more from the original spectrum than the "FLAG == 0" spectrum.

I have not tested several bad columns on a source. This would certainly produce a larger effect, but I expect the conclusion would not be changed (that the "grey" and "dead" spectra are not any worse than the "FLAG == 0" spectrum).

6 Summary and recommendations

In Sect. 4, I have illustrated how the spectrum changes next to a dead column. The effect is quite significant on a hard spectrum. But it is diluted when averaged over an entire source. In Sect. 5, I have shown that keeping events next to a dead column running through a source with a standard spectrum does not affect the spectral fit in a significant way.

This does not mean that #XMMEA_EM is perfect, however, because it contains many other flags than CLOSE_TO_DEADPIX. I had originally thought to keep the FLAG==0 recommendation for spectral analysis but not flag any longer events CLOSE_TO_DEADPIX. This would have been a very simple change to *emenergy*, would not have changed anything for the users and would have kept exact parallelism (from the user's point of view) between MOS and pn. But it was remarked by R. Saxton that this would be ambiguous for the tasks dealing with exposure issues, like *arfgen*. There is currently no way for *arfgen* to know that events CLOSE_TO_DEADPIX were intentionally not flagged, and that consequently the pixels close to dead pixels should be counted in after all.

Therefore it was decided to keep the FLAG definition as it is currently, and add a new selection definition XMMEA_SM= '(FLAG & 0x766bae67) == 0' (the exact value may change if new flags appear). #XMMEA_SM takes out all flagged events except those flagged only as CLOSE_TO_DEADPIX (=0x100). This keeps all the information, provides all the necessary information to *arfgen*, but requires changing the recipe and breaks the parallel between MOS and pn for creating spectra. The new definition must be written in the *sasdal* part of the *saslib* package of the SAS, and *emenergy* will be modified to call it and write that definition as a keyword into the events file.

To improve over that, it would be worth investigating the spectrum next to truely dead columns (if there are any) over a large statistical sample to see whether an excess of low energy events exists there (due to pattern migration as described in Sect. 2.2, second option). If it does, then distinguishing between "dead" and "grey" columns would improve the spectral reliability without costing too much effective area (as long as truely dead columns are few).

Acknowledgements

Many thanks to Jean-Luc Sauvageot for a critical reading of the text.

References

[1] Pratt, Gabriel W., and Ballet, Jean, 2003, "Characterisation of MOS columns with shifted energies", EPIC-SAP-GWP-01