rmfgen

August 16, 2005

Abstract

Creates an OGIP-compliant redistribution matrix file (RMF) for the source in question.

1 Instruments/Modes

Instrument	Mode
EPIC MOS	IMAGING,TIMING
EPIC PN	IMAGING,TIMING

2 Use

pipeline processing	no
interactive analysis	yes

3 Description

Further information aabout the scientific use and accuracy of this task may be found in the document, XMM-SOC-PS-TN-0043, (available from *http://xmm.vilspa.esa.es/calibration/*)

Reformat the detector response and energy bounds information provided by the CAL for the given instrument, corrects for instrumental effects specific to the event selection criteria used and writes the result to a specified dataset (the Redistribution Matrix File or RMF). The dataset conforms to the OGIP standard ([3]).

The RMF matrix generated by **rmfgen** describes the response of the instrument as a function of energy and PI channel; PHA channel responses can not be generated by this task.

This dataset can be used in conjunction with the Ancillary Response File (ARF) dataset (generated by **arfgen**) to perform spectral analysis, the most commonly used analysis package being XSPEC.

Note that details of individual CCF constituents and how the CAL processes such data are beyond the scope of this document, and the reader should refer to the Calibration Access and Data Handbook for such information ([1]).

The following list summarises the current and planned features of **rmfgen**:

Item	Description	Status
1	Grouping of response data above a threshold value	implemented
2	Channel rebinning	implemented
3	Spectrum-response channel range matching	implemented
4	Observing mode dependence	implemented
5	Spatial dependence and averaging	implemented
6	User-defined energy grid	implemented
7	DSS support (including eg pattern selection)	implemented
8	Modelling for spectral distortion due to pile-up	initial implementation, not available
9	Modelling of spectral distortion due to Charge Transfer Inefficiency (CTI)	to be coded

Items 1-7 are described in more detail below. The remaining items are discussed in the future developments section (**??**).

3.1 Grouping and threshold

rmfgen generates the o/p rmf in the format specified by OGIP. This format allows response data for a given energy range to be stored in the form of contiguous groups of elements whose values are above a particular threshold. The values below threshold are not written. The fitting package expands this format, replacing those elements in between groups with zero. This is extremely useful in the case of a very sparse response matrix, where values close to zero are of no interest.

The threshold value can be set by the user through the `threshold` parameter and allowed range is between 0 and 1 (the maximum possible response value).

3.2 Variable length column support

rmfgen writes out RMF matrices in variable-length vector column format. This results in smaller matrices than the alternative fixed length format and reduces the demands on machine resources. The default thresholding used is such that elements less than 1.0×10^{-6} are ignored by default.

3.3 Channel rebinning

Even with the grouping/thresholding mechanism described above, the RMF sizes can still be fairly large. Another possiblity is to resample the input spectrum, using an external task such as **evselect**. **rmfgen** will read in the attributes in the spectrum dataset describing that rebinning and offset (`SPECPIX`, `SPECVAL` and `SPECDELT`), and resample the response matix accordingly. The resampling affects the channel axis of the matrix only; the energy axis is unaffected.

It is expected that a rebinning factor of typically 10 is possible with EPIC MOS spectra, which corresponds to a similar size reduction in the RMF.

Table 1: **Observing mode support within rmfgen**

Camera	Observing Mode	Support
MOS	FullFrame	Yes
MOS	SmallWindow	Yes
MOS	LargeWindow	Yes
MOS	Timing	Yes
MOS	Burst	Yes
PN	FullFrame	Yes
PN	ExtendedFullFrame	Yes
PN	SmallWindow	Yes
PN	LargeWindow	Yes
PN	Timing	Yes
PN	Burst	Yes

3.4 Spectrum-response channel range matching

Fitting packages such as XSPEC do not allow the channel range of the spectrum and that defined in the RMF to differ. XSPEC assumes the spectral range to correspond to the legal range of the `CHANNEL` column, ie the attributes `TLMIN1` and `TLMAX1`.

If the channel range of the spectrum and that of the reponse data obtained from the CAL are found to differ, **rmfgen** raises a warning and performs one of the following corrective actions:

- If the spectrum contains channels that are outside the response range, **rmfgen** will pad out the response (prior to the grouping described above) to fit the spectral range.
- If the channel range of the spectrum is a subset of the response range, **rmfgen** will truncate the response data to match the channel range.

Note that the only accepted value for `TLMIN` is zero (which is the same as the lower bound for the CAL data, ensuring that the spectral channel range can only be a subset or superset of the response channel range, and not overlapping).

3.5 Observing mode support

This version of **r**mfgen supports all commonly used observing modes (Table 1). The mode is obtained from the keyword SUBMODE in the primary header of the input spectrum. If the mode contained in this header is not recognised then the PRIME_FULL_WINDOW mode is used by default. The PN response is different for each observing mode while the MOS response is equal in the window modes but varies due to threshold effects in the Timing and Burst modes.

3.6 Pattern support

The RMF of the Epic detectors varies depending on the event grades used to create the spectrum. Different RMFs may be created by the task for single (pattern 0), double (patterns 1–4), triple (patterns 5–8), quadruple (patterns 9-12), single plus double and single plus double plus triple plus quadruple events. The pattern information is read from the datasubspace stored in the spectrum header. The current standard practise is to use singles, doubles or singles plus doubles for Epic-PN and singles or

singles plus double plus triple plus quadruple for Epic-MOS but see XMM-SOC-PS-TN-0043 for more details.

3.7 Spatial dependence and averaging

The redistribution function of the EPIC-MOS and EPIC-pn cameras is spatially dependent. In the pn the spectral resolution of the detector improves for positions further away from the CAM-EX, such that the worst resolution is obtained at the RAW-Y=199 position. The MOS redistribution is uniform over the camera except for a small patch in each camera, coincident with the boresight, which shows a broadened response below 1 keV that has evolved with time. This spatial and temporal dependence is coded into CCFs and is treated automatically by rmfgen.

The task calculates an average RMF over the spatial region used to extract the source spectrum. For a correct treatment the task does need to know how the source flux is distributed within the extraction region. A detector map mechanism, successfully employed in arfgen, has been adopted for this purpose. Essentially a grid is placed over the source region and the redistribution parameters calculated at each point. These values are then averaged in some way. By means of command line parameters the averaging may be performed assuming that the flux distribution is that of the point spread function (PSF), a flat uniform distribution or it may be defined explicitly by supplying the task with an image containing the source region.

Point Sources:

For point sources the PSF option should be chosen by setting `detmaptype`=psf. The PSF for photons of energy 2 keV is used. For a single circular, annular or rectangular (BOX) region the EXTENDED mode PSF is used which is based upon a King profile and is the most accurate parameterisation. A source region with a different shape, for instance a region with circles excluded to remove other point sources, will use the HIGH mode implementation.

Extended Sources:

Large, non-pointlike, sources may be either approximated by a uniform distribution, by setting `detmaptype`=flat or modelled accurately by creating an image of the source region, using **evselect** or **xmmselect**, and supplying this to the task by:

```
rmfgen detmaptype=array detmaparray=<image_name>
```

The image MUST be created in detector coordinates (not X/Y coords) and should entirely cover the region used for extracting the spectrum. If the image does not cover the whole region then a warning (`detmapXBoundsExceeded` or `detmapYBoundsExceeded`) is issued. It should be noted that this process averages the parameters used to model the redistribution function rather than averaging the function itself. Hence there is little computational overhead associated with this step. Images should be created with sufficient resolution to sample the variation in flux distribution; a 100 by 100 pixel image may be sufficient.

For a more in-depth discussion of detector maps please see the **arfgen** user guide, section 7.3.

3.7.1 CCF dependence

For the PN, the redistribution function is simply dependent upon the mean RAW-Y value of the photons. This is time independent and averaging will work with any version of the EPN_REDIST CCFs. The MOS spatial dependence has been introduced explicitly into the EMOSn_REDIST CCFs from version

0052 onwards. Earlier CCFs do not support this feature and hence no averaging will be performed if these CCFs are used.

3.7.2 Bad pixels

Ideally the effect of bad pixels, bad columns and chip gaps would be taken into account in the spatial averaging. Such a level of sophistication has NOT yet been implemented.

3.8 User-defined energy grid

The performance of the task is directly proportional to the number of bins in the energy grid (ie the number of rows of the matrix), while the scientific accuracy of the results of using the matrix improves (to the limit of the energy resolution of the instrument) the more rows there are in the matrix. This task by default uses the energy grid defined by the CAL to construct the response matrix. This grid currently contains approx 1000 bins for both PN and MOS instruments.

rmfgen now allows the user to specify his/her own grid, thereby allowing the user to decide at which point to compromise performance for accuracy. In addition, a user-specified grid will allow the user to restrict the matrix to a specific energy range of the instrument, and thus allow improved accuracy and a smaller RMF file as output, without having to sacrifice performance.

A user-defined grid can at present be only such that the grid points are evenly distributed across the energy range. It can be specified by setting `withenergybins` = `true` and entering values for the parameters `energymin`, `energymax` and `nenergybins`.

3.9 Examples

3.9.1 A point source

The task defaults have been set for a point source and so specifying:

```
rmfgen spectrumset=spectrum.ds rmfset=myresp.rmf
```

will create a response function, called myresp.rmf, which has been spatially averaged using a PSF detector map, appropriate for a point source. The output filename defaults to *response.ds* if `rmfset` is not defined.

3.9.2 An extended source

Here a choice has to be made between a uniform spatial averaging, made by:

```
rmfgen spectrumset=spectrum.ds rmfset=myresp.rmf detmaptype=flat
```

and supplying an image to give the ultimate accuracy by:

```
rmfgen spectrumset=spectrum.ds rmfset=myresp.rmf detmaptype=dataset detmaparray=myimage.ds
```

NB: The image supplied MUST be in detector and NOT sky (X/Y) coordinates.

3.9.3 Using the `threshold` parameter

The example above would generate a dataset, with a threshold of 1.0×10^{-6}. The matrix size may be reduced at the expense of accuracy by lowering the threshold as in the following example:

```
rmfgen spectrumset=spectrum.pha threshold=1e-5 format='var'
```

4 Parameters

This section documents the parameters recognized by this task (if any).

Parameter	Mand	Type	Default	Constraints
spectrumset	yes	dataset	spectrum.pha	none

Name of the counts spectrum file for which the output RMF dataset is associated.

rmfset	no	dataset	responsefile.ds	none

Name of output response dataset.

threshold	no	real	1e-5	0 <= `threshold` <= 1

Threshold level for grouping. All elements within the matrix below this value are excluded in the formatted output.

withenergybins	no	boolean	false	none

If `true` use energy grid specified by `energymin`, `energymax`, and `nenergybins`, otherwise, use the grid defined in the CAL.

energymin	no	real	0	none

Used if `withenergybins` = true. Lower energy bound of matrix, in keV.

energymax	no	real	15	none

Used if `withenergybins` = true. Upper energy bound of matrix, in keV.

nenergybins	no	integer	30	none

Used if `withenergybins` = true. The number of bins in the energy grid = number of rows in RMF matrix.

format	no	choice	var	fixed, var

Format for `MATRIX` vector column. Set this to 'fixed' to include all matrix elements.

detmaptype	no	choice	psf	dataset flat psf

The origin of the detector map: if set to `dataset`, then the user must specify the name of that dataset, via `detmaparray`. If it is set to `flat` or `psf`, then the parameters `detxbins` and `detybins` define a detector map grid of the same size as the source extraction region. A larger region can be defined by setting `withdetbounds` true and setting the `detxoffset`, `detyoffset` parameters.

detmaparray	no	array	detmapfile.ds:	none

Name of detector map dataset and optionally array in the DAL compound notation. Only used if `detmaptype` is set to `dataset`.

withdetbounds	no	boolean	false	none

For `psf` or `flat` detector maps. If `true` then the bounds of the internal detector map are taken from the parameters `detxoffset` and `detyoffset`. Otherwise, the bounds are taken from the DSS information in the spectrum dataset.

detxoffset	no	real	1200	0 <= detxoffset <= 30000

Used if `withdetbounds` is set true and `detmaptype` is set to `psf` or `flat`. This defines the half-width of the detector map in DET coordinates. For example, if the source x-coordinate is x, the x-bounds are [x−`detxoffset`, x+`detxoffset`].

detxbins	no	integer	20	≥ 1

Used if `detmaptype` is set to `psf` or `flat`. This is the number of bins along the x-axis of the map.

detyoffset	no	real	1200	0 <= detyoffset <= 30000

Used if `withdetbounds` is set true and `detmaptype` is set to `psf` or `flat`. This defines the half-height of the detector map in DET coordinates. For example, if the source y-coordinate is y, the y-bounds are [y−`detyoffset`,y+`detyoffset`].

detybins	no	integer	20	≥ 1

Used if `detmaptype` is set to `psf` or `flat`. This is the number of bins along the y-axis of the map.

5 Errors

This section documents warnings and errors generated by this task (if any). Note that warnings and errors can also be generated in the SAS infrastructure libraries, in which case they would not be documented here. Refer to the index of all errors and warnings available in the HTML version of the SAS documentation.

incompatibleLowerChannelBounds *(warning)*
The channel range specified by the `TLMIN` and `TLMAX` attributes of the column `CHANNEL` differs from the intrinsic channel range of the response data. If the spectral channel range is greater than the response range, **rmfgen** will pad out the response to fit the spectral range. If the spectrum contains channels that are a subset of the response range, **rmfgen** will truncate the response data to match the channel range.
corrective action: Pad out or truncate response data to match channel bounds specified in spectral dataset.

invalidCCFChannels *(warning)*
The program reads the energy boundaries of the spectral channels from the CAL and attempts to rebin these boundaries to the number of bins in the input spectrum. If the ratio of spectral bins to CAL channel boundaries is not an integer the rebinning is not possible and so the boundaries are defined as a regularly increasing set of energies.
corrective action: assume that the channel energies are regularly spaced.

UnknownModeString *(warning)*
The PN rmf generator is dependent on the observing mode. If this is not available from the input spectrum or is not recognised then the task defaults to using full frame mode.
corrective action: Use PRIME_FULL_WINDOW mode

NoDSSPatterns *(warning)*
The RMF generated is dependent on the event patterns used to create the spectrum. If these can not be read from the DSS in the spectrum header then the task assumes that all patterns have been used.
corrective action: Assume all patterns selected

invalidChannelLowBound *(error)*
The `TLMIN` attribute value for column `CHANNEL` is not zero

invalidChannelBounds *(error)*
The values of attributes `TLMIN` and `TLMAX` for column `CHANNEL` and attribute `DETCHANS` are not such that DETCHANS $\leq$ TLMAX $-$ TLMIN $+ 1$.

invalidChannelOffset *(error)*
The channel offset specified by attributes `SPECPIX`, `SPECVAL` and `SPECDELT` does not correspond to an integer value.

incompatibleEnergyRanges *(error)*
If `withenergyranges` is set to true, and `responsedata` is set to 'cal', then this error is raised if the requested range specified by parameters `energymin` and `energymax` is outside the range over which response data is defined in the CAL.

invalidResponseChannels *(error)*
rmfgen is supposed to pad out/truncate the response data to match the spectral channel range. This error is raised if this particular algorithm fails. An SPR should be raised.

noPhaElements *(error)*
Prior to regrouping, no detector channels could be found in the response matrix above the

threshold level specified by the `threshold` parameter. This error is raised at a different point in the grouping algorithm to the error `noGroups` below - such an err

groupingInvalid *(error)*
No groups of contiguous response elements above `threshold` could be found in the response matrix, although the maximum number of PI elements found in any one channel is greater than one. This suggests a problem with the grouping algorithm, and an SPR should be raised.

tooFewDataPoints *(error)*
There are less than 2 data points in the dataset specified by the parameter `photondistset` (and `modelpileup` is set to true).

TooManyChannels *(error)*
The spectrum contains too many data channels. The EPIC calibration is defined for a certain PI channel binning. If this is exceeded then the task can not produced a meaningful matrix. Please recreate the spectrum using the standard number of channels or less. The current standard is 800 MOS bins from PI channels 0 to 11999 and 4096 PN bins from 0 to 20479.

InvalidPatterns *(error)*
The patterns in the datra subspace are not the standard selections. See the XMM SAS User Guide for more information about the supported patterns.

NoRawCoords *(error)*
Literally means an incorrect treatment of the CCF within the task. It can probably only arise if the REDIST CCF being used is corrupt.

For other warnings and errors please see the **arfgen** user guide.

6 Input Files

1. Input counts spectrum dataset specified by the parameter `spectrumset`. The format of this dataset is defined in the OGIP document [2], but for the purposes of **rmfgen**, it must contain the following information:
   - The global attributes `TELESCOP`, `INSTRUME`, `FILTER`.
   - A table named `SPECTRUM`, containing the attributes `SPECPIX`, `SPECDELT` and `SPECVAL`. This table in turn contains:
     - A column named `CHANNEL`, which contains the attributes `TLMIN` and `TLMAX`.
2. Input detector map. This is a file or array specified by the parameter `detmaparray` and containing the following information:
   - A 2-d array of any numerical type. By default this will be the primary array in an image dataset.
   - The WCS attributes `CTYPE1`, `CRPIX1`, `CRVAL1`, `CDELT1`, `CTYPE2`, `CRPIX2`, `CRVAL2`, `CDELT2`. This map must be generated in terms of [`DETX`,`DETY`] coordinates, so the axis types, ie `CTYPE1` and `CTYPE2`, must be DETX and DETY respectively.

   This map is only used by **rmfgen** if `detmaptype` is set to 'dataset'.
3. Input photon spectrum dataset specified by the parameter `photonspectrumset`. This is a DAL dataset containing:

- A table labelled DISTRIBUTION. This in turn contains
  - A 32-bit real column labelled ENERGY, containing the energy for each data point (keV).
  - A 32-bit real column labelled FLUX, containing the flux level (photons/s/cm2).

This describes the estimated incident photon spectrum. This is used during the pile-up correction process.

This dataset is not used if the parameter modelpileup is false.

7 Output Files

1. An OGIP-compliant response dataset. The format of this dataset is defined in the OGIP document [3], but in summary, this contains:

   - A table labelled MATRIX, containing the following columns
     - A 32-bit real column labelled ENERG_LO
     - A 32-bit real column labelled ENERG_HI
     - A 16-bit integer column labelled N_GRP
     - A 16-bit integer vector column labelled F_CHAN
     - A 16-bit integer vector column labelled N_CHAN
     - A 32-bit real vector column labelled MATRIX

     and the following attributes:
     - TELESCOP
     - INSTRUME
     - FILTER
     - RMFVERSN
     - CHANTYPE
     - DETCHANS
     - HDUCLASS
     - HDUCLAS1
     - HDUCLAS2
     - HDUCLAS3
     - HDUVERS1
   - A EBOUNDS table, containing the following
     - A 16-bit integer column labelled CHANNEL
     - A 32-bit real column labelled E_MIN
     - A 32-bit real column labelled E_MAX

     and the following attributes:
     - TELESCOP
     - INSTRUME
     - FILTER
     - RMFVERSN
     - CHANTYPE
     - DETCHANS
     - HDUCLASS
     - HDUCLAS1
     - HDUCLAS2
     - HDUCLAS3
     - HDUVERS1

8 Algorithm

This is divided into four stages:

1. Initialisation
2. Main stage
3. `MATRIX` table creation
4. `EBOUNDS` table creation
5. Cleanup

These are described in detail in the following subsections.

8.1 Initialisation

- Set state of internal RMF data server. Various actions include:
  1. Setting the CAL state to the default CCD chip and mode (Full Frame) for the instrument in question.
  2. Open input counts spectrum file and determine channel range in o/p matrix from the TLMIN and TLMAX attributes for the channel column.
  3. Use the spatial region information and selected detector map to calculate average redistribution parameters.
  4. If pile-up is requested, input photon spectrum file for input.
- Create output file

8.2 Main Stage - Generate response matrix

- Get energy ranges for matrix from the CAL
- Write these data out to matrix table in output file
- Foreach energy range
  - Obtain response data from CAL using the corresponding row value for elow and ehigh as arguments
  - Perform instrumental corrections. This is currently a placeholder, but will eventually include pile-up and CTI corrections,
  - Normalise response row
  - If necessary, pad out or truncate response row according to the values of TLMIN and TLMAX of the CHANNEL column.
  - If necessary, resample response row according using attributes SPECPIX, SPECVAL and SPECDELT of the spectral file.
  - Place corrected response row in memory.
- End loop

8.3 MATRIX table creation

- Go through corrected matrix to find out the following statistics:
  1. The largest number of channel groups in a row
  2. The largest number of channel elements in a row above threshold
- Create empty output table header and structure.
- Create a linked list (GROUP_LIST) to store groups of contiguous elements
- Foreach row in corrected matrix:
  - Create a linked list corresponding to a single group (GROUP).
  - Foreach element in row
    - If element is above threshold, add element to GROUP.
    - Else push GROUP onto GROUP_LIST, and create a new instance of the GROUP structure for the next group of contiguous elements.
  - End element loop
  - Write out GROUP_LIST to output table.
  - Create a new instance of GROUP_LIST for the next row.
- end row loop

8.4 EBOUNDS table creation

- Write header information for EBOUNDS table
- Get the channel ranges from the CAL and attempt to rebin them for the input spectrum. If this fails: For each spectrum channel, compute the values of the columns E_MIN and E_MAX from the PI channel range corresponding to that channel.
- Write EBOUNDS data to file

8.5 Cleanup

- Close output file, cleanup variables and close the CAL and DAL.

9 Comments

- The RGS is addressed by the task **rgsrmfgen**.
- The size of the o/p RMF is dependent on the spectrum binsize and the number of bins in the energy grid, and can be very large. Heavy consumption of disk space is expected. Memory consumption will also be high as the full RMF is held in memory before being written to disk. Note that the pile-up correction requires knowledge of the entire uncorrected response matrix during each row construction, but accesses this from the CAL, not from memory.
- The task directly uses a section of code from the **arfgen** task. The debug and warnings from this code may appear to be related to arfgen but they are equally relevant for this task.

References

[1] Christian Erd, Phillipe Gondoin, David Lumb, Rudi Much, Uwe Lammers, and Giuseppe Vacanti. Calibration Access and Data Handbook. Technical Report XMM-PS-GM-20, ESA/SSD, Jan 14 2000. Found at the URL: `http://xmm.vilspa.esa.es/docs/documents/CAL-MAN-0001-2-1.ps.gz`.

[2] I.M. George K.A. Arnaud and A.F. Tennant. The OGIP Spectral File Format. Technical Report OGIP/92-007, NASA/GSFC, Sept 1992. Found at the URL: `http://legacy.gsfc.nasa.gov/docs/heasarc/ofwg/docs/summary/ogip_92_007_summary.html`.

[3] I.M. George K.A. Arnaud and A.F. Tennant. The Calibration Requirements for Spectral Analysis. Technical Report OGIP/92-002, NASA/GSFC, Dec 1998.