srcmatch

August 15, 2005

Abstract

This task makes the final EPIC summary source list. It reads the *emldetect* source list and correlates, if applicable, the X-ray sources with the sources in the OM source list. Additionally, this tasks allows more *emldetect* source lists to be merged into one summary list, and/or it correlates the summary source list with source lists from the earlier stages of the EPIC detection pipeline.

1 Instruments/Modes

Instrument	Mode
EPIC MOS:	IMAGING
EPIC PN:	IMAGING
OM:	IMAGING

2 Use

pipeline processing	yes
interactive analysis	no

3 Description

The task combines four different functions:

- A one line summary for each source detected by the EPIC detection tasks is created. The summary information is a subset of the information given in the **emldetect** source list table.
- Cross-identification of the summary source list with source lists from earlier stages of the detection pipeline (i.e to fill in the `PN_ML_ID_SRC`, `M1_ML_ID_SRC`,...,`M2_BOX_ID_SRC` columns).
- **emldetect** source lists from different detection runs can be merged into one summary source list.
- A cross-identification of the EPIC sources with the sources listed in the OM observation source list is performed. The cross-correlation radius can be specified by an input parameter. Up to five OM source identification numbers can be listed in the output set.

Cross-correlating EPIC source lists for either merging or for correlating earlier pipeline source lists is done in two stages. The first stage is similar for both the merging of source lists and cross-identifying sources with earlier pipeline source lists. In this stage (subroutine compare_srcs) correlations are investigated and pointers between matching sources are set. Always the closest companion (in terms of *sigmas*) is chosen. Only sources within 3 to 5 sigmas will be correlated (the exact number can be set using input parameter *maxerr*). In order to avoid problems due to boresight effects an input parameter *systerr* (for systematic error) has been introduced.

In the second stage the cross-identifications are either used to get the pointers to the source numbers from previous detection stages (e.g. the `PN_ML_ID_SRC` field), or the two lists are merged together. The first option works for the data sets provided with *crossidsets* command line parameter, whereas merging is done if more than one set is provided using the *inputlistsets* parameter. Both options can be used simultanuously.

Merging means that new sources will be added to the list and parameters of cross-identified sources will be averaged using weights based upon the statistical errors. The maximum likelihood parameters will be all listed seperately using columns with dimension larger or equal to one, in the same manner as the crossidentification columns (see developer's notes). The reason is that maximum likelihood values cannot be simply averaged or added without losing its meaning as a probability related parameter.

For both merging and crossidentfying the source lists are read one by one and are correlated seperately. This ensures that not more than 2 source lists will be stored in memory. Upon execution the number of sources that can be stored in the summary source list is three times the number of sources in the first input set. If during the execution this turns out not to be enough, the source list will be copied to a new array with a three times larger capacity.

The output EPIC observation source lists are provided both in FITS and in HTML format, *D.M.:* where the HTML list is intended to give a quick overview for each source detected, but does not provide instrument–specific data.

As of version 3.10.2 the output FITS table contains the colums IN_b, IN_b_ERR, IN_EXP, with IN=[PN,M1,M2] and b=[1,2,3,4,5,TOT,XID]. These columns contain count rates, (IN_b), count rate errors (IN_b_ERR), and exposure map values (IN_n_EXP, n=[1..5]) from the individual instruments and up to 5 energy bands. In the case of several source lists from one instrument, count rates are averaged and exposure map values are added.

As of version 3.10.5 the strings from the input columns VAL_FLAG, VAL_FLAG, VER_COMM are copied into the columns of the output file IN_VAL_FLAG, IN_VAL_FLAG, IN_VER_COMM (IN=[PN,M1,M2]).

From version 3.12 fluxes and flux errors from the input source lists are copied to the columns inst_b_FLUX and ERR_inst_b_FLUX (inst=[PN,M1,M2], b=[1,2,3,4,5,TOT,XID]).

4 Parameters

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

Parameter	Mand	Type	Default	Constraints
inputlistsets	yes	filename(s)	inputlistset.fits	

EPIC emldetect source list(s). Specifying more than one list, means these lists will be merged.

outputlistset	yes	filename	summarylist.fits	

FITS output observation source list

htmloutput	yes	filename	summarylist.html	

Output observation source list in HTML format

crossidsets	no	filename(s)	crossid.fits	

Source lists from earlier stages of the detection pipeline. These files will be used to fill in the pointers to the source identification numbers in these lists.

usecrossidsets	no	Boolean	no	

Controls whether a cross correlation with X-ray source list from earlier detection stages should be performed.

omlistset	no	filename	omlist.fits	

OM observation source list. Setting omlistset=- means that no OM data set is available.

useomlistset	no	Boolean	no	

Controls whether a cross correlation with an OM source list should be performed.

systerr	no	float	1.0	[0.0<param<6.0]

Systematic error for EPIC source matching [arcsec]

maxerr	no	float	1.0	[3.0<param<5.0]

Maximum allowed error (stat. + syst.) for EPIC source matching [sigmas].

omdist	no	float	15.0	[0.0<param<60.0]

Max. distance EPIC–OM for correlation [arcsec]

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.

noData *(error)*
This error occurs when the data set does not appear to be empty. Note that this may also be due to a wrongly formatted input set (e.g. no instrumental summary lines).

tooManySets *(warning)*
The number of input data sets to be merged exceeded the maximum number (currently set to 9).
corrective action: Only the first 9 number of input sets are processed.

noMatchingSources *(warning)*
Normally one expects at least some of the sources in two input data sets to overlap, if this is not the case this warning message appears.
corrective action: The input sets should be checked by hand/eye whether the two should actually be merged together.

6 Input Files

The following input files will be read:

1. EPIC **emldetect** source list(s), with a maximum of 9. (at least 1 file is mandatory)
2. EPIC source lists from an earlier stage of the pipeline. These will be used for cross identifications only. (optional)
3. OM observation source list

This task requires that the EPIC input sets contain summary lines for each source, except if the input set contains only data from one instrument. Summary information has `ID_INST` = 0 and `ID_BAND` = 0.

7 Output Files

The two output files are:

1. PPS product : final EPIC FITS observation source lists. Columns `RA_CORR` and `DEC_CORR` are only passed to the output summary source lists, when **srcmatch** is used on only one source list. The appropiate **eposcorr** are passed to the output source list (see below).
2. PPS product : final EPIC HTML observation source lists.

srcmatch allows more than one EPIC source list to be merged or to be cross correlated with the summary source list. In order to keep track of the original source identifications of task **emldetect** we allow the columns `EP_BOX_ID` and `EP_ML_ID` and the columns (`PN_ML_ID_SRC`, `M1_ML_ID_SRC`,.., `PN_BOX_ID_SRC`, etc.) to have more than 1 entry. Additionally we use also this scheme (but now with floating point data) for columns with maximum likelihood values (`PN_DET_ML`, etc, and `PN_EXT_ML`, etc.). The number of entries depends on the number of datasets, which will be stored using the keywords `NSETS` (for `EP_BOX_ID` , `EP_ML_ID` and the maximum likelihood parameters) and `NCROSSID` (the other ID columns). Additionally,

the number of entries per row for a given column can also be retrieved from the FITS keyword TDIMi (i being the column index).

srcmatch will pass on the columns RA_CORR and DEC_CORR (generated by **eposcorr**) from the input X-ray source list to the summary output list, *only if the number of input source lists is 1.* The reason for this restriction is that it is not obvious how the values should be averaged or what should be done if one of the input list was not modified by **eposcorr**. The keywords generated by **eposcorr** will also be passed on. In case the input file was not modified by **eposcorr** the above mentioned columns and associated keywords will not be written to the output set.

The datasets which were merged together and their order can be retrieved from the header keywords EP_SET1, EP_SET2, ... with a maximum up to the number stored in the internal parameter *max_nsets*. The datasets used for cross-identification can be retrieved from the CROSSID1, CROSSID2,..., keywords. The following additional header keywords are present (they correspond to the input parameters with the same names): SYSTERR, MAXERR and OMDIST.

Whether the input X-ray source list(s) was/were cross correlated with an OM source list, can be verified with the (logical) keyword OMCROSS. OMCROSS = T means it was cross correlated. The maximum number of associated optical sources (counterparts) from the OM source list is set to five (corresponding to five additional columns of the EPIC observation source list). If more correlations are found, the five counterparts with the smallest distance to the EPIC position will be listed and a label (column OMFLAG, see below) will be set accordingly.

In addition to the five columns giving the OM source numbers, a numeric column giving the total number of OM counterparts found for each EPIC source (which may be more than five) and a label column OMFLAG is added to the observation list.

The label OMFLAG can take the following string values:

- OM_SRC_FND: Counterparts found
- NO_OM_SRCS: OM data available, no counterpart found
- NO_OM_DATA: No OM data available at location of X-ray source

In order to determine the OMFLAG values NO_OM_SRCS and NO_OM_DATA an OM exposure map or detection mask would be required which at present is not included in the list of PPS data products. OMFLAG will therefore not be set in the first release software version.

8 Algorithm

```
subroutine srcmatch
  call subbroutine read_set in order to read set 1
  IF number of EMLDETECT datasets > 1 THEN
     LOOP over input EMLDETECT datasets
        call i/o subroutine for additional input sets
        call subroutine compare_src (correlates the two sets)
        call subroutine combine_src (merge two sets into one set)
        ! the first set is the main set and will finally constitute the
        ! summary set
     END LOOP
     LOOP over input CROSSID datasets
```

```
        call i/o subroutine for input sets
        call subroutine compare_src (correlates the two sets)
        call subroutine crossid_srcs (fills in cross id pointers)
     END LOOP
  END IF

  Copy source coordinate and source numbers from
    OM observation source lists to memory.

  LOOP over EMLDETECT source list

    IF (ID_INST=0).AND.(ID_BAND=0) THEN

       LOOP over OM source coordinates

         Find EPIC OM counterparts (distance < distance parameter)

         Increment 'number of counterparts' counter and copy
           to output record buffer

         Write OM source number to output record buffer

         Update OMFLAG and copy to output record buffer

       END LOOP

       Copy EMLDETECT source parameters to output record buffer.

    END IF

    IF (ID_INST>=0.OR.ID_BAND>=0) THEN
       Copy instrument and band specific EMLDETECT source parameters
         to output record buffer.
    END IF


    IF (ID_INST=0.AND.ID_BAND.AND..NOT.first pass through loop) THEN

       Write out output record buffer in FITS format
         (EPIC FITS observation source list).
       Write out output record buffer in HTML format
         (EPIC HTML observation source list).

    END IF

  END LOOP

end subroutine  srcmatch

subroutine compare_srcs
    LOOP over two source lists
      IF two sources are within max_err sigma of each other THEN
            IF sources have no associations assigned yet THEN
                    let the two source point to each other
            ELSE
```

```
                IF New association is better than previous one THEN
                     let the two source point to each other
                END
        END IF
      END LOOP
  end subroutine compare_srcs
```

9 Comments

Users have noticed that in some cases the number of output sources depend on the order of the input lists. This is due the fact that the first source list is used for position reference. Therefore, with more than 3 input lists the matching result can depend on the input order, if the distances between the candidates are near the matching radius.

10 Future developments

References