



Publication Year	2016
Acceptance in OA	2020-05-20T15:18:33Z
Title	C3: Command-line Catalogue Crossmatch for modern astronomical surveys
Authors	RICCIO, GIUSEPPE, BRESCIA, Massimo, CAVUOTI, STEFANO
Handle	http://hdl.handle.net/20.500.12386/25011
Journal	Astrophysics Source Code Library



User Guide

Introduction

C³ is a command-line python script, designed and developed to perform general cross-matching among astrophysical catalogues. The tool is able to be easily executed as a stand-alone python process or integrated within any generic data reduction/analysis pipeline.

Its high-performance capability is ensured by making use of a multi-core parallel processing paradigm. It is designed to deal with massive catalogues in different formats, with the maximum flexibility given to the end user, in terms of choice about catalogue parameters, coordinates and cross-matching type.

In C³ different functional cases and matching criteria have been implemented, as well as the most used join types. It also works with the most common catalogue formats, with or without header, like FITS (Flexible Image Transport System, version tabular), ASCII (American Standard Code for Information Interchange, ordinary text, i.e. space separated values), CSV (Comma Separated Values), VOTable (Virtual Observatory Table, XML based) and with two coordinate systems, equatorial and galactic.

HW/SW Requirements

- PC/Workstation, x-core processor, RAM min. 4 GB
- OS: Linux Ubuntu, Windows 7/10, Mac OS, Fedora
- Python 2.x.x or 3.x.x, with mandatory libraries: Numpy, Pyfits
- STILTS, release 3.0-7

Execution

```
prompt>> python c3.py [-c, --config FILE] [-h, --help] [-d, --default FILE]
```

where:

- c, --config: Specify the name of the configuration file (optional, the default is config.ini)
- h, --help: Show help
- d, --default: It automatically generates a default configuration file (default name confi.ini). If the specified name or default name is already assigned to an existing file, the program exit with error



Configuration

The configuration file contains all information needed for cross-matching. Parameters may be mandatory, required only for specific use cases, or optional. Parameters are grouped in sections, and their name MUST NOT be changed. The use cases implemented by C³ are:

1. **sky**: the cross-match is done within sky areas defined by the catalogue parameters;
2. **exact value**: two objects matched if they have the same value for the specified column;
3. **row-by-row**: match done on a same row-ID of the two catalogues (the two catalogues must have same number of rows).

In Table 1 and Table 2 the configuration file parameters are described. In particular Table 2 is structured as follows:

- **column 1**: parameter name, as it has to be written in the file (**case insensitive**);
- **column 2**: parameter reference section of the configuration file (**within square brackets**);
- **column 3**: parameter condition (required/specific/optional);
- **column 4**: parameter allowed entries;
- **column 5**: default value (when present).

Name	Description
Input catalogue 1	catalogue 1 file name
Format catalogue 1	catalogue 1 file name
Input catalogue 2	catalogue 2 file name
Format catalogue 2	catalogue 2 file name
Output	output file name – Format CSV
Log file	log file name
Stilts directory	Directory containing the STILTS jar file
Working directory	Temporary working directory (automatically removed at the end of execution)
Algorithm	type of algorithm used for cross-matching (see Table 2 for details)
Input 1 column	name or number of column used for matching reference in the catalogue 1
Input 2 column	name or number of column used for matching reference in the catalogue 2
Area shape	Shape of the region where to perform matching
Size type	Type of size for the matching area: parametric in case of any dependence from a specific object column; fixed otherwise
Matching area first dimension	value of the first dimension (first axis for ellipse/rectangle) of the match region. Units: arcsec (in case of fixed size type); name or number of column containing the 1 th dimension value for each object in catalogue 1 (in case of parametric size type)
Matching area second dimension	value of the second dimension (second axis for ellipse/rectangle) of the match region. Units: arcsec (in case of fixed size type); name or number of column containing the 2 nd dimension value for each object in catalogue 1 (in case of parametric size type)
Parametric factor	multiplicative factor for the parametric values, useful for dimension units conversion or resizing
pa column/value	name or number of column related to object position angle of catalogue 1. It is also possible to insert an arbitrary numeric value. Units: degrees
pa settings	orientation of position angle and its degrees of shift from 0-point. The PA is referred always to the major axis of the match region
Catalogue 2 minimum partition cell size	minimum dimension of the unit cell used for sky partitioning (optional, default=100, user machine dependent). Units: arcsecs .
coordinate system	coordinate system used for catalogue 1



coordinate units	coordinates units for catalogue 1
glon/ra column	name or number of column related to the glon or ra coordinate for catalogue 1
glat/dec column	name or number of column related to the glat or dec coordinate for catalogue 1
designation column	name or number of column related to object Designation of catalogue 1 (if not set the row number will be used)
coordinate system	coordinate system used for catalogue 2
coordinate units	coordinates units for catalogue 2
glon/ra column	name or number of column related to glon or ra coordinate for catalogue 2
glat/dec column	name or number of column related to glat or dec coordinate for catalogue 2
designation column	name or number of column related to object Designation of catalogue 2 (if not set the row number will be used)
thread limit	maximum number of simultaneous processes (user machine dependent)
Match selection	match selection type (in case of multiple rows matching): all , every match is included; best , only the best pairs are stored
Join type	output joined table content specification: 1 and 2 : only rows having an entry in both input catalogues (correspondent SQL: <i>inner join</i>) 1 or 2 : every row, matched and unmatched, from both input catalogues (correspondent SQL: <i>full outer join</i>) all from 1 (all from 2) : All matched rows from catalogue 1 (or 2), together with the unmatched rows from catalogue 1 (or 2). Correspondent SQL: <i>left/right outer join</i> 1 not 2 (2 not 1) : all catalogue 1 (or 2) rows not having matches in the catalogue 2 (or 1) 1 xor 2 : The "exclusive or" of the match – i.e. only rows from the catalogue 1 not having matches in the catalogue 2 and vice versa.

Table 1 – Description of all configuration file parameters

Name	Reference section	Required/Optional	Valid entries	Default
Input catalogue 1	I/O Files	required	/<path>/<filename>	
Format catalogue 1		required	csv fits ascii votable	
Input catalogue 2		required	/<path>/<filename>	
Format catalogue 2		required	csv fits ascii votable	
Output		required	/<path>/<filename>	
Output format		required	csv fits ascii votable	
Log File		optional	/<path>/<filename>	./log.txt
Stilts directory		required	/<path>	
Working directory		required	/<path>	

Table 2a - Parameters of the configuration file: column 1, parameter name; column 2, section in the configuration file; column 3, required/optional; column 4, allowed values; column 5, default value.



Name	Reference section	Required/Optional	Valid entries	Default
Algorithm	Match Criteria	required	sky exact value row-by-row	
Area shape	"Sky" parameters	required	ellipse rectangle	
Size type		required	parametric fixed	
Matching area first dimension		required	<float> or <column name> <column number>	
Matching area second dimension		required	<float> or <column name> <column number>	
Parametric factor		required only for parametric type	<float>	
pa column/value		optional only for Algorithm: sky	<column name> <column number> or arbitrary number	PA value=0
pa settings		optional only for Algorithm: sky	clock/counter, <float>	clock, 0.0
Catalogue 2 minimum partition cell size		optional only for Algorithm: sky	<float>	100
Input 1 column		"Exact Value" parameters	required only for Algorithm: exact value	<column name> <column number>
Input 2 column	required only for Algorithm: exact value		<column name> <column number>	
	"Row-by-Row" parameters	no specific parameters are required for this type		

Table 3b - Parameters of the configuration file: column 1, parameter name; column 2, section in the configuration file; column 3, required/optional; column 4, allowed values; column 5, default value.



Name	Reference section	Required/Optional	Valid entries	Default
coordinate system	Catalogue 1 Properties	required only for Algorithm: sky	galactic icrs fk4 fk5	
coordinate units		required only for Algorithm: sky	degrees (deg) radians (rad) sexagesimal (sex)	
glon/ra column		required only for Algorithm: sky	<column name> <column number>	
glat/dec column		required only for Algorithm: sky	<column name> <column number>	
designation column		optional only for Algorithm: sky	<column name> <column number>	

Table 4c - Parameters of the configuration file: column 1, parameter name; column 2, section in the configuration file; column 3, required/optional; column 4, allowed values; column 5, default value.

Name	Reference section	Required/Optional	Valid entries	Default
coordinate system	Catalogue 2 Properties	required only for Algorithm: sky	galactic icrs fk4 fk5	
coordinate units		required only for Algorithm: sky	degrees (deg) radians (rad) sexagesimal (sex)	
glon/ra column		required only for Algorithm: sky	<column name> <column number>	
glat/dec column		required only for Algorithm: sky	<column name> <column number>	
designation column		optional only for Algorithm: sky	<column name> <column number>	

Table 5d - Parameters of the configuration file: column 1, parameter name; column 2, section in the configuration file; column 3, required/optional; column 4, allowed values; column 5, default value.

Name	Reference section	Required/Optional	Valid entries	Default
thread limit	Threads Properties	optional	<integer>	CPUcount – 1 or 1 (for single CPU)

Table 6e - Parameters of the configuration file: column 1, parameter name; column 2, section in the configuration file; column 3, required/optional; column 4, allowed values; column 5, default value.

Name	Reference section	Required/Optional	Valid entries	Default
Match selection	Output Rows	required	all best	
Join type		required	1 and 2 1 or 2 all from 1 all from 2 1 not 2 2 not 1 1 xor 2	

Table 7f - Parameters of the configuration file: column 1, parameter name; column 2, section in the configuration file; column 3, required/optional; column 4, allowed values; column 5, default value.

Matching area settings

In the “Sky” case, C³ cross-matches objects lying within an elliptical, circular or rectangular area, centered on the sources of the first input catalogue. The matching area is characterized by 6 parameters, to be set in the configuration, which define its shape, dimensions and orientation (Table 1 and Table 3b). In Figure 1 is depicted a graphical representation of two matching areas (one elliptical and one rectangular) with the indication of its parameters.

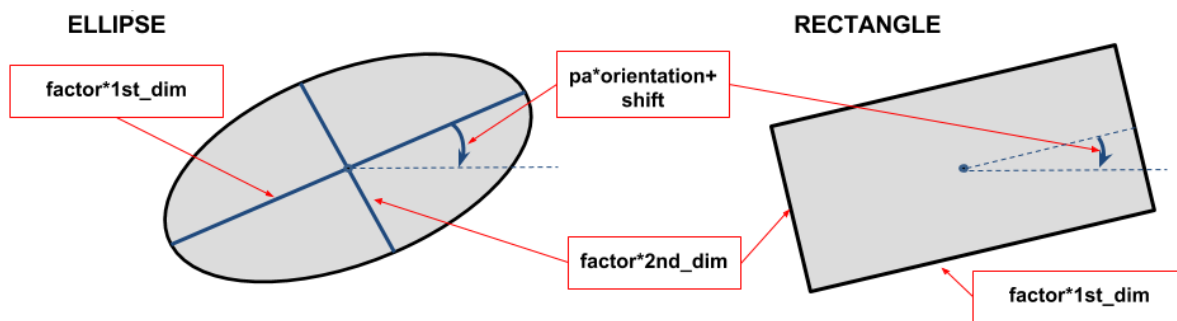


Figure 1 - Configuration of C3 Matching Area: it can be elliptical (circular as special case) or rectangular; its dimensions, defined in the configuration file as “matching area 1st dimension” and “matching area 2nd dimension”, represent the ellipse axes or width and height of the rectangle, multiplied, in the case of parametric size type, by a user defined “parametric factor”; the position angle is characterized by a value (in degree) and two offset parameters, “orientation” and “shift”.

In particular, to define the orientation of the matching area, C³ requires two further parameters, in addition to the value of the position angle, *orientation* and *shift*, defining the PA offset. The position angle, indeed, is referred, by default, to the main axis of the matching area (greatest) with a clockwise orientation. The two offset parameters give the possibility to indicate the correct orientation (clockwise/counterclockwise) and a shift angle (in degrees).



EXAMPLES OF CONFIGURATION FILE

1. SKY CONFIGURATION (Size Type: parametric – Match criteria: Best, 1 and 2)

```
[I/O Files]
Input catalogue 1:                ./input/cat1.csv
Format catalogue 1:              csv
Input catalogue 2:              ./input/cat1.fits
Format catalogue 2:              fits
Output:                          ./output/out.csv
Log file:                        ./out.log
Stilts directory:                ./libs
working directory:               tmp
[Match Criteria]
algorithm:                        sky
[Sky Parameters]
area shape:                       ellipse
size type:                        parametric
matching area first dimension:    FW500_1
matching area second dimension:   12
parametric factor:                1.5
pa column/value:                 PA500
pa settings:                      clock, 45.0
catalogue 2 minimum partition cell size: 125
[Catalogue 1 Properties]
coordinate system:                icrs
coordinate units:                 rad
glon/ra column:                  RA
glat/dec column:                 DEC
designation column:               1
[Catalogue 2 Properties]
coordinate system:                galactic
coordinate units:                 deg
glon/ra column:                  GLON
glat/dec column:                 GLAT
designation column:               SOURCEID
[Threads Properties]
thread limit:                     100
[Output Rows]
Match selection:                  best
Join type:                        1 and 2
```



2. SKY CONFIGURATION (Size Type: FIXED – Match criteria: All, 1 or 2):

[I/O Files]

Input catalogue 1:	./input/cat1.csv
Format catalogue 1:	csv
Input catalogue 2:	./input/cat1.fits
Format catalogue 2:	fits
Output:	./output/out.csv
Log file:	./out.log
Stilts directory:	./libs
working directory:	tmp

[Match Criteria]

algorithm:	sky
------------	-----

[Sky Parameters]

area shape:	rectangle
size type:	fixed
matching area first dimension:	20.0
matching area second dimension:	10.0
pa column/value:	30.0
pa settings:	clock, 45.0
catalogue 2 minimum partition cell size:	125

[Catalogue 1 Properties]

coordinate system:	icrs
coordinate units:	rad
glon/ra column:	RA
glat/dec column:	DEC
designation column:	1

[Catalogue 2 Properties]

coordinate system:	galactic
coordinate units:	deg
glon/ra column:	GLON
glat/dec column:	GLAT
designation column:	SOURCEID

[Threads Properties]

thread limit:	100
---------------	-----

[Output Rows]

Match selection:	all
Join type:	1 or 2



3. EXACT VALUE CONFIGURATION:

[I/O Files]

Input catalogue 1: ./input/cat1.csv
Format catalogue 1: csv
Input catalogue 2: ./input/cat1.fits
Format catalogue 2: fits
Output: ./output/out.csv
Log file: ./out.log
Stilts directory: ./libs
working directory: tmp

[Match Criteria]

algorithm: exact value

[Exact Value parameters]

Input 1 column: ID
Input 2 column: 1

[Threads Properties]

thread limit: 100

[Output Rows]

Match selection: all
Join type: 1 or 2

4. ROW-BY-ROW CONFIGURATION:

[I/O Files]

Input catalogue 1: ./input/cat1.csv
Format catalogue 1: csv
Input catalogue 2: ./input/cat1.fits
Format catalogue 2: fits
Output: ./output/out.csv
Log file: ./out.log
Stilts directory: ./libs
working directory: tmp

[Match Criteria]

algorithm: row-by-row

[Threads Properties]

thread limit: 100

Info & Questions

E-mail: helpdame@gmail.com

Website: <http://dame.dsf.unina.it/c3.html>