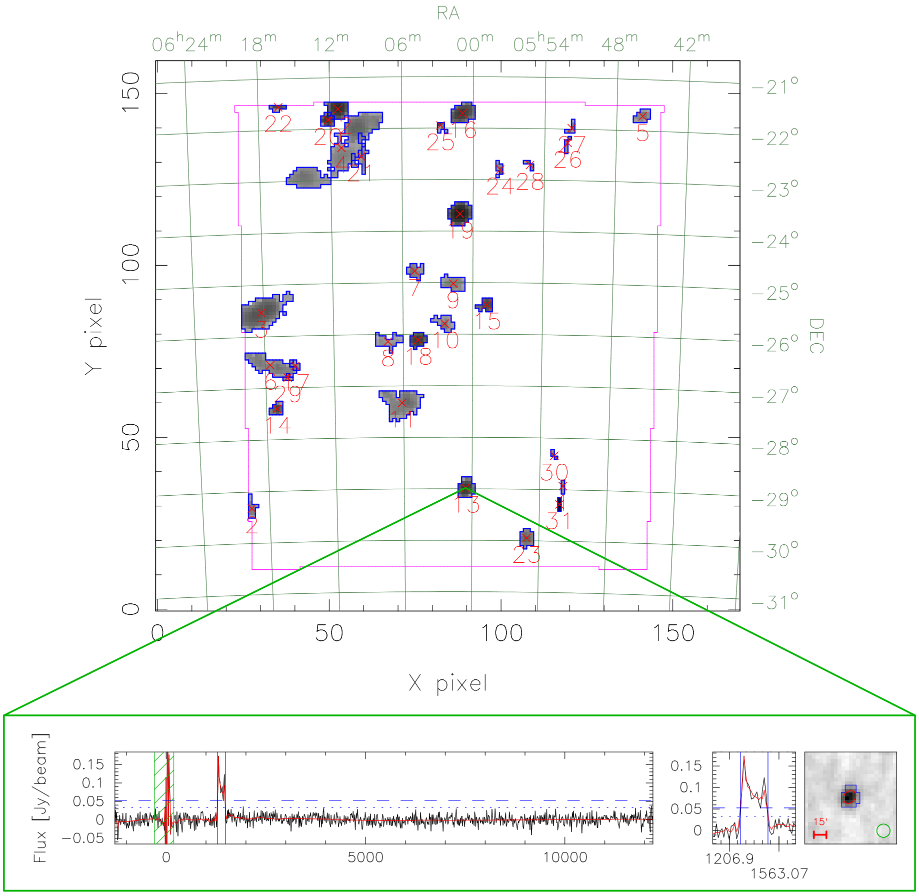
Source Detection with *Duchamp*

A User's Guide

Matthew Whiting
Australia Telescope National Facility
CSIRO Astronomy & Space Science

Duchamp version 1.6.1
May 1, 2014



Contents

1 Introduction and getting going quickly

1.1 About *Duchamp*

This document provides a user's guide to *Duchamp*, an object-finder for use on spectral-line data cubes. The basic execution of *Duchamp* is to read in a FITS data cube, find sources in the cube, and produce a text file of positions, velocities and fluxes of the detections, as well as a postscript file of the spectra of each detection.

Duchamp has been designed to search for objects in particular sorts of data: those with relatively small, isolated objects in a large amount of background or noise. Examples of such data are extragalactic HI surveys, or maser surveys. *Duchamp* searches for groups of connected voxels (or pixels) that are all above some flux threshold. No assumption is made as to the shape of detections, and the only size constraints applied are those specified by the user.

Duchamp has been written as a three-dimensional finder, but it is possible to run it on a two-dimensional image (i.e. one with no frequency or velocity information), or indeed a one-dimensional array, and many of the features of the program will work fine. The focus, however, is on object detection in three dimensions, one of which is a spectral dimension. Note, in particular, that it does not do any fitting of source profiles, a feature common (and desirable) for many two-dimensional source finders. This is beyond the current scope of *Duchamp*, whose aim is reliable detection of spectral-line objects.

Duchamp provides the ability to pre-process the data prior to searching. Spectral baselines can be removed, and either smoothing or multi-resolution wavelet reconstruction can be performed to enhance the completeness and reliability of the resulting catalogue.

1.2 Acknowledging the use of *Duchamp*

Duchamp is provided in the hope that it will be useful for your research. If you find that it is, I would ask that you acknowledge it in your publication by using the following: "This research made use of the Duchamp source finder, produced at the Australia Telescope National Facility, CSIRO, by M. Whiting."

Additionally, *Duchamp* has been described in a journal paper (Whiting 2012). This paper covers the key algorithms implemented in the software, and provides some simple completeness and reliability comparisons of different modes of operation. Users of *Duchamp* are encouraged to read the paper in conjunction with this user guide, as while some things are repeated herein, not everything is. Whiting (2012) should be cited when describing the use of *Duchamp* in your research.

1.3 What to do

So, you have a FITS cube, and you want to find the sources in it. What do you do? First, you need to get *Duchamp*: there are instructions in Appendix A for obtaining and installing it. Once you have it running, the first step is to make an input file that contains the list of parameters. Brief and detailed examples are shown in Appendix C. This file provides the input file name, the various output files, and defines various parameters that control the execution.

The standard way to run *Duchamp* is by the command

```
> Duchamp -p [parameter file]
```

replacing `[parameter file]` with the name of the file listing the parameters.

An even easier way is to use the default values for all parameters (these are given in Appendix B and in the file `InputComplete` included in the distribution directory) and use the syntax

```
> Duchamp -f [FITS file]
```

where `[FITS file]` is the file you wish to search.

The default action includes displaying a map of detected objects in a PGPLOT X-window. This can be disabled by setting the parameter `flagXOutput = false` or using the `-x` command-line option, as in

```
> Duchamp -x -p [parameter file]
```

and similarly for the `-f` case.

Once a FITS file and parameters have been set, the program will then work away and give you the list of detections and their spectra. The program execution is summarised below, and detailed in §3. Information on inputs is in §2.1 and Appendix B, and descriptions of the output is in §5.

1.4 Guide to terminology and conventions

First, a brief note on the use of terminology in this guide. *Duchamp* is designed to work on FITS "cubes". These are FITS[1] image arrays with (at least) three dimensions. They are assumed to have the following form: the first two dimensions (referred to as x and y) are spatial directions (that is, relating to the position on the sky – often, but not necessarily, corresponding to Equatorial or Galactic coordinates), while the third dimension, z, is the spectral direction, which can correspond to frequency, wavelength, or velocity. The three dimensional analogue of pixels are "voxels", or volume cells – a voxel is defined by a unique (x, y, z) location and has a single value of flux, intensity or brightness (or something equivalent) associated with it.

Sometimes, some pixels in a FITS file are labelled as BLANK – that is, they are given a nominal value, defined by FITS header keywords BLANK (and potentially BSCALE and BZERO), that marks them as not having a flux value. These are often used to pad a cube out so that it has a rectangular spatial shape. *Duchamp* has the ability to avoid these: see §3.3.1.

Note that it is possible for the FITS file to have more than three dimensions (for instance, there could be a fourth dimension representing a Stokes parameter). Only the two spatial dimensions and the spectral dimension are read into the array of pixel values that is searched for objects. All other dimensions are ignored[2]. Herein, we discuss the data in terms of the three basic dimensions, but you should be aware it is possible for the FITS file to have more than three. Note that the order of the dimensions in the FITS file does not matter.

[1] FITS is the Flexible Image Transport System – see Hanisch et al. (2001) or websites such as http://fits.cv.nrao.edu/FITS.html for details.

[2] This actually means that the first pixel only of that axis is used, and the array is read by the `fits_read_subsetnull` command from the CFITSIO library.

With this setup, each spatial pixel (a given (x, y) coordinate) can be said to be a single spectrum, while a slice through the cube perpendicular to the spectral direction at a given z-value is a single channel, with the 2-D image in that channel called a channel map.

Detection involves locating contiguous groups of voxels with fluxes above a certain threshold. *Duchamp* makes no assumptions as to the size or shape of the detected features, other than allowing user-selected minimum or maximum size criteria. Features that are detected are assumed to be positive. The user can choose to search for negative features by setting an input parameter – which will invert the cube prior to the search (see §3.7.2 for details).

1.5 A summary of the execution steps

The basic flow of the program is summarised here – all steps are discussed in more detail in the following sections.

1. The necessary parameters are recorded.

   How this is done depends on the way the program is run from the command line. If the `-p` option is used, the parameter file given on the command line is read in, and the parameters therein are read. All other parameters are given their default values (listed in Appendix B).

   If the `-f` option is used, all parameters are assigned their default values, with the flux threshold able to be set with the `-t` option.

2. The FITS image is located and read in to memory.

   The file given is assumed to be a valid FITS file. As discussed above, it can have any number of dimensions, but *Duchamp* only reads in the two spatial and the one spectral dimensions. A subset of the FITS array can be given (see §3.1 for details).

3. If requested, a FITS file containing a previously reconstructed or smoothed array is read in.

   When a cube is either smoothed or reconstructed with the *à trous* wavelet method, the result can be saved to a FITS file, so that subsequent runs of *Duchamp* can read it in to save having to re-do the calculations.

4. If requested, BLANK pixels are trimmed from the edges, and the baseline of each spectrum is removed.

   BLANK pixels, while they are ignored by all calculations in *Duchamp*, do increase the size in memory of the array above that absolutely needed. This step trims them from the spatial edges, keeping a record of the amount trimmed so that they can be added back in later.

   A spectral baseline (or bandpass) may optionally be removed at this point as well. This may be necessary if there is a ripple or other large-scale feature present that will hinder detection of faint sources.

5. If the reconstruction method is requested, and the reconstructed array has not been read in at Step 3 above, the cube is reconstructed using the *à trous* wavelet method.

This step uses the multi-resolution *à trous* method to determine the amount of structure present at various scales. A simple thresholding technique then removes random noise from the cube, leaving the significant signal. This process can greatly reduce the noise level in the cube, enhancing the reliability of the resulting catalogue.

6. Alternatively (and if requested), the cube is smoothed, either spectrally or spatially.

   This step presents two options. The first considers each spectrum individually, and convolves it with a Hanning filter (with width chosen by the user). The second considers each channel map separately, and smoothes it with a Gaussian kernel of size and shape chosen by the user. This step can help to reduce the amount of noise visible in the cube and enhance fainter sources, increasing the completeness and reliability of the output catalogue.

7. A threshold for the cube is then calculated, based on the pixel statistics (unless a threshold is manually specified by the user).

   The threshold can either be chosen as a simple $n\sigma$ threshold (i.e. a certain number of standard deviations above the mean), or calculated via the "False Discovery Rate" method. Alternatively, the threshold can be specified as a simple flux value, without care as to the statistical significance (e.g. "I want every source brighter than 10mJy").

   By default, the full cube is used for the statistics calculation, although the user can nominate a subsection of the cube to be used instead.

8. Searching for objects then takes place, using the requested thresholding method.

   The cube is searched either one channel-map at a time ("spatial" search) or one spectrum at a time ("spectral" search). Detections are compared to already detected objects and either combined with a neighbouring one or added to the end of the list.

9. The list of objects is condensed by merging neighbouring objects and removing those deemed unacceptable.

   While some merging has been done in the previous step, this process is a much more rigorous comparison of each object with every other one. If a pair of objects lie within requested limits, they are combined.

   After the merging is done, the list is culled (although see comment for the next step). There are certain criteria the user can specify that objects must meet: minimum (or maximum) numbers of spatial pixels and spectral channels, and minimum separations between neighbouring objects. Those that do not meet these criteria are deleted from the list.

10. If requested, the objects are "grown" down to a lower threshold, and then the merging step is done a second time.

    In this case, each object has pixels in its neighbourhood examined, and if they are above a secondary threshold, they are added to the object. The merging process is done a second time in case two objects have grown over the top of one another. Note that the rejection part of the previous step is not done until the end of the second merging process.

11. The baselines and trimmed pixels are replaced prior to output.

    This is just the inverse of step #4.

12. The details of the detections are written to screen and to the requested output file.

    Crucial properties of each detection are provided, showing its location, extent, and flux. These are presented in both pixel coordinates and world coordinates (e.g. sky position and velocity). Any warning flags are also printed, showing detections to be wary of. Alternative output options are available, such as a VOTable or annotation files for image viewers such as kvis, ds9 or casaviewer.

13. Maps showing the spatial location of the detections are written.

    These are 2-dimensional maps, showing where each detection lies on the spatial coverage of the cube. This is provided as an aid to the user so that a quick idea of the distribution of object positions can be gained e.g. are all the detections on the edge?

    Two maps are provided: one is a 0th moment map, showing the 0th moment (i.e. a map of the integrated flux) of each detection in its appropriate position, while the second is a "detection map", showing the number of times each spatial pixel was detected in the searching routines (including those pixels rejected at step 9 and so not in any of the final detections).

    These maps are written to postscript files, and the 0th moment map can also be displayed in a PGPLOT X-window.

14. A pixel mask is written to a FITS file.

    A FITS file of the same size as the input file can be written. Here, each pixel has a value indicating whether or note it was detected and falls in one of the catalogue sources. Different objects can be traced by different non-zero pixel values.

15. The integrated or peak spectra of each detection are written to a postscript file.

    The spectral equivalent of the maps – what is the spectral profile of each detection? Also provided here are basic information for each object (a summary of the information in the results file), as well as a 0th moment map of the detection.

16. If requested, a text file containing all spectra is written.

    This file will contain the peak or integrated spectra for each source, as a function of the appropriate spectral coordinate. The file is a multi-column ascii text file, suitable for import into other software packages.

17. If requested, FITS files are written containing the reconstructed, smoothed, baseline or mask arrays.

    If one of the preprocessing methods was used, the resulting array can be saved as a FITS file for later examination or use (for instance, reading in as described at step #3). The FITS header will be the same as the input file, with a few additional keywords to identify the file.

1.6 Why "*Duchamp*"?

Well, it's important for a program to have a name, and the initial working title of *cubefind* was somewhat uninspiring. I wanted to avoid the classic astronomical approach of designing a cute acronym, and since it is designed to work on cubes, I looked at naming it after a cubist. *Picasso*, sadly, was already taken (Minchin 1999), so I settled on naming it after Marcel Duchamp, another cubist, but also one of the first artists to work with "found objects".

2 User Inputs

2.1 Parameter file input

Duchamp allows a large degree of control over the way the different algorithms work. This is done by means of input parameters, specified through a parameter file. The parameter file is provided at runtime, via

```
> Duchamp -p [parameter file]
```

The parameter file simply contains a list of parameter names followed by the value that should be assigned to them. The syntax used is

```
parameterName value.
```

Parameter names are not case-sensitive, and lines in the input file that start with # are ignored. If a parameter is listed more than once, the latter value is used, but otherwise the order in which the parameters are listed in the input file is arbitrary. Example input files can be seen in Appendix C.

If a parameter is not listed, the default value is assumed. The defaults are chosen to provide a good result (a simple 5σ search with no pre-processing), so the user doesn't need to specify many new parameters in the input file. Note that the image file **must** be specified! The parameters that can be set are listed in Appendix B, with their default values in parentheses.

The parameters with names starting with `flag` are stored as `bool` variables, and so are either `true = 1` or `false = 0`. They can be entered in the file either in text or integer format – *Duchamp* will read them correctly in either case.

An example input file is included in the distribution tar file. It is as follows:

```
imageFile       your-file-here
logFile         logfile.txt
outFile         results.txt
spectraFile     spectra.ps
minPix          2
flagATrous      1
snrRecon        5.
snrCut          3.
minChannels     3
flagBaseline    1
```

You would, of course, replace the "`your-file-here`" with the FITS file you wanted to search. Further examples are given in Appendix C.

2.2 Command-line control

Duchamp provides the ability to run without constructing a parameter file first. Using the `-f` command-line flag to specify an image will make use of the default values for all parameters:

```
> Duchamp -f [FITS image]
```

It is possible to specify a flux threshold as well on the command line, using the `-t` flag. This allows the user to quickly search a given image to a given depth (i.e. give me all sources in this image above 1mJy.

```
> Duchamp -f [FITS image] -t [THRESHOLD]
```

The `-t` flag can also be used with the `-p` option – it sets the `threshold` parameter, and overrides the value provided in the parameter file specified. The flux threshold should be in the same brightness units as specified in the FITS image.

The other command-line flag that can be used is `-x`. This turns off the X-window output (that graphically shows where the detected objects are). This makes scripted use of *Duchamp* somewhat easier.

3 What *Duchamp* is doing

Each of the steps that *Duchamp* goes through in the course of its execution are discussed here in more detail. This should provide enough background information to fully understand what *Duchamp* is doing and what all the output information is. For those interested in the programming side of things, *Duchamp* is written in C/C++ and makes use of the CFITSIO, WCSLIB and PGPLOT libraries.

3.1 Image input

The image cube to be searched is provided to *Duchamp* in FITS format. The `cfitsio` library will read `.fits` files, as well as compressed `.fits.gz` files directly.

The cube is read in using basic CFITSIO commands, and stored as an array in a special C++ class. This class keeps track of the list of detected objects, as well as any reconstructed arrays that are made (see §3.4). The World Coordinate System (WCS)[3] information for the cube is also obtained from the FITS header by WCSLIB functions (Calabretta and Greisen 2002; Greisen and Calabretta 2002; Greisen et al. 2006), and this information, in the form of a `wcsprm` structure, is also stored in the same class. See §3.2 for more details.

A sub-section of a cube can be requested by defining the subsection with the `subsection` parameter and setting `flagSubsection = true` – this can be a good idea if the cube has very noisy edges, which may produce many spurious detections.

There are two ways of specifying the `subsection` string. The first is the generalised form `[x1:x2:dx,y1:y2:dy,z1:z2:dz,...]`, as used by the CFITSIO library. This has one set of colon-separated numbers for each axis in the FITS file. In this manner, the x-coordinates run from `x1` to `x2` (inclusive), with steps of `dx`. The step value can be omitted, so a subsection of the form `[2:50,2:50,10:1000]` is still valid. In fact, *Duchamp* does not make use of any step value present in the subsection string, and any that are present are removed before the file is opened.

If the entire range of a coordinate is required, one can replace the range with a single asterisk, e.g. `[2:50,2:50,*]`. Thus, the subsection string `[*,*,*]` is simply the entire cube. Note that the pixel ranges for each axis start at 1, so the full pixel range of a 100-pixel axis would be expressed as 1:100. A complete description of this section syntax can be found at the FITSIO web site[4].

Making full use of the subsection requires knowledge of the size of each of the dimensions. If one wants to, for instance, trim a certain number of pixels off the edges of the cube, without examining the cube to obtain the actual size, one can use the second form of the subsection string. This just gives a number for each axis, e.g. `[5,5,5]` (which would trim 5 pixels from the start *and* end of each axis).

All types of subsections can be combined e.g. `[5,2:98,*]`.

Typically, the units of pixel brightness are given by the FITS file's BUNIT keyword. However, this may often be unwieldy (for instance, the units are Jy/beam, but the values are around a few mJy/beam). It is therefore possible

[3] This is the information necessary for translating the pixel locations to quantities such as position on the sky, frequency, velocity, and so on.

[4] http://heasarc.gsfc.nasa.gov/docs/software/fitsio/c/c_user/node91.html

to nominate new units, to which the pixel values will be converted, by using the `newFluxUnits` input parameter. The units must be directly translatable from the existing ones – for instance, if BUNIT is Jy/beam, you cannot specify mJy, it must be mJy/beam. If an incompatible unit is given, the BUNIT value is used instead.

3.2 World Coordinate System

Duchamp uses the WCSLIB package to handle the conversions between pixel and world coordinates. This package uses the transformations described in the WCS papers (Calabretta and Greisen 2002; Greisen and Calabretta 2002; Greisen et al. 2006). The same package handles the WCS axes in the spatial plots. The conversions used are governed by the information in the FITS header – this is parsed by WCSLIB to create the appropriate transformations.

For the spectral axis, however, *Duchamp* provides the ability to change the type of transformation used, so that different spectral quantities can be calculated. By using the parameter `spectralType`, the user can change from the type given in the FITS header. This should be done in line with the conventions outlined in Greisen et al. (2006). The spectral type can be either a full 8-character string (e.g. 'VELO-F2V'), or simply the 4-character "S-type" (e.g. 'VELO'), in which case WCSLIB will handle the conversion.

The rest frequency can be provided as well. This may be necessary, if the FITS header does not specify one and you wish to transform to velocity. Alternatively, you may want to make your measurements based on a different spectral line (e.g. OH1665 instead of HI-21cm). The input parameter `restFrequency` is used, and this will override the FITS header value.

Finally, the user may also request different spectral units from those in the FITS file, or from the defaults arising from the WCSLIB transformation. The input parameter `spectralUnits` should be used, and Greisen and Calabretta (2002) should be consulted to ensure the syntax is appropriate.

3.3 Image modification

Several modifications to the cube can be made that improve the execution and efficiency of *Duchamp* (their use is optional, governed by the relevant flags in the parameter file).

3.3.1 BLANK pixel removal

If the imaged area of a cube is non-rectangular (see the example in Fig. 3, a cube from the HIPASS survey), BLANK pixels are used to pad it out to a rectangular shape. The value of these pixels is given by the FITS header keywords BLANK, BSCALE and BZERO. While these pixels make the image a nice shape, they will take up unnecessary space in memory, and so to potentially speed up the processing we can trim them from the edge. This is done when the parameter `flagTrim = true`. If the above keywords are not present, the trimming will not be done (in this case, a similar effect can be accomplished, if one knows where the "blank" pixels are, by using the subsection option).

The amount of trimming is recorded, and these pixels are added back in once the source-detection is completed (so that quoted pixel positions are applicable to the original cube). Rows and columns are trimmed one at a time until the first non-BLANK pixel is reached, so that the image remains rectangular. In practice, this means that there will be some BLANK pixels left in the trimmed image (if the non-BLANK region is non-rectangular). However, these are ignored in all further calculations done on the cube.

3.3.2 Negative-feature detection

Duchamp searches strictly for positive sources – that is, pixels that are **above** the detection threshold. If one instead wishes to search for negative features (such as absorption lines), set the parameter `flagNegative=true`.

This inverts the cube (i.e. multiplies all pixels by -1) prior to searching, and then re-inverts the cube, and the fluxes of any detections (so that the detections will have, for instance, a negative peak flux) after searching is complete. Any reconstructed or smoothed array that has been read from disk is also inverted.

This works fine for the simple case of inverting the cube. If, however, a spectral baseline needs to be removed (see the next section §3.3.3), then special care needs to be taken. This is described in more detail in §3.3.4.

3.3.3 Baseline removal

The user may request the removal of baselines from the spectra, via the parameter `flagBaseline`. This may be necessary if there is a strong baseline ripple present, which can result in spurious detections at the high points of the ripple, or if there is continuum emission present in the cube that is not interesting for the search.

There are two ways in which the baseline can be calculated. The first makes use of the *à trous* wavelet reconstruction procedure (see §3.4), where only the two largest scales are kept.

The second method uses a median filter combined with a sliding box of some specified width. For each pixel, the baseline is determined to be the median value of all pixels in a region of that width centred on the pixel of interest. The box is truncated at the edges of the spectrum (so that fewer pixels are included), but there will always be at least half the width of the box present.

The choice between these methods is made using the `baselineType` parameter – only values of `atrous` (the default) or `median` are accepted. If the `median` method is being used, the full width of the box is specified with `baselineBoxWidth`.

The baseline calculation is done separately for each spatial pixel (i.e. for each spectrum in the cube), and the baselines are stored and added back in before any output is done. In this way the quoted fluxes and displayed spectra are as one would see from the input cube itself – even though the detection (and reconstruction if applicable) is done on the baseline-removed cube.

When the `atrous` method is used, the presence of very strong signals (for instance, masers at several hundred Jy) could affect the determination of the baseline, and would lead to a large dip centred on the signal in the baseline-subtracted spectrum. To prevent this, the signal is trimmed prior to the reconstruction process at some standard threshold (at 5σ above the mean). The baseline determined should thus be representative of the true, signal-free baseline. Note that this trimming is only a temporary measure which does not affect the source-detection.

The baseline values can be saved to a FITS file for later examination. See §5.4.4 for details.

3.3.4 Combining negative detections with baseline removal

When both `flagNegative=true` and `flagBaseline=true`, care needs to be taken when parameterising the sources. They are strictly negative on the baseline-subtracted spectrum, and so parameterisation is done on this spectrum, prior to adding the baseline back.

This ensures that the peak flux and location correspond to the maximum absorption (the peak flux will be quoted as a negative value), and the spectral plots (see §5.3.1) will properly show the peak spectrum (when used in `spectralMethod=peak` mode).

This method is typical of the approach that would be used for detecting absorption lines against a continuum source. For now, the parameterisation is the same as for other detections, but a future version of *Duchamp* will have parameters more relevant for absorption lines, such as optical depth and equivalent width.

3.3.5 Flagging channels

Finally, it is possible to flag particular channels so that they are not included in the search. This is an extension of the old "Milky Way" channel range. That allowed the specification of a single contiguous block of channels, and was aimed at excluding Galactic emission in extragalactic HI cubes.

The new flagging approach allows the specification of a series of channels and channel ranges. This allows the user to block the detection of known regions of RFI, or known but uninteresting emission (e.g. Galactic HI emission if you are searching for extragalactic sources).

Flagged channels are specified using the `flaggedChannels` parameter, and can be given by a comma-separated list of single values or ranges. For instance: `flaggedChannels 5,6,12-20,87`. These channels refer to channel numbers in the **the full cube**, before any subsection is applied. Also note that **the channel numbering starts at zero**, that is, channel 0 is the first channel of the cube.

The effect is to ignore detections that lie within these channels. If a spatial search is being conducted (i.e. one channel map at a time), these channels are simply not searched. If a spectral search is being conducted, those channels will be flagged so that no detection is made within them. The spectral output (see Fig. 2) will ignore them as far as scaling the plot goes, and the channel range will be indicated by a green hatched box.

Note that these channels will be included in any smoothing or reconstruction that is done on the array, and so will be included in any saved FITS file (see §3.6).

3.4 Image reconstruction

The user can direct *Duchamp* to reconstruct the data cube using the multi-resolution *à trous* wavelet algorithm. A good description of the procedure can be found in Starck and Murtagh (2002). The reconstruction is an effective way of removing a lot of the noise in the image, allowing one to search reliably to fainter levels, and reducing the number of spurious detections. This is an optional step, but one that greatly enhances the reliability of the resulting catalogue, at the cost of additional CPU and memory usage (see §6 for discussion).

3.4.1 Algorithm

The steps in the *à trous* reconstruction are as follows:

1. The reconstructed array is set to 0 everywhere.

2. The input array is discretely convolved with a given filter function. This is determined from the parameter file via the `filterCode` parameter – see Appendix B for details on the filters available. Edges are dealt with by assuming reflection at the boundary.

3. The wavelet coefficients are calculated by taking the difference between the convolved array and the input array.

4. If the wavelet coefficients at a given point are above the requested reconstruction threshold (given by `snrRecon` as the number of σ above the mean and adjusted to the current scale – see Appendix J), add these to the reconstructed array.

5. The separation between the filter coefficients is doubled. (Note that this step provides the name of the procedure[5], as gaps or holes are created in the filter coverage.)

6. The procedure is repeated from step 2, using the convolved array as the input array.

7. Continue until the required maximum number of scales is reached.

8. Add the final smoothed (i.e. convolved) array to the reconstructed array. This provides the "DC offset", as each of the wavelet coefficient arrays will have zero mean.

The range of scales at which the selection of wavelet coefficients is made is governed by the `scaleMin` and `scaleMax` parameters. The minimum scale used is given by `scaleMin`, where the default value is 1 (the first scale). This parameter is useful if you want to ignore the highest-frequency features (e.g. high-frequency noise that might be present). Normally the maximum scale is calculated from the size of the input array, but it can be specified by using `scaleMax`. A value ≤ 0 will result in the use of the calculated value, as will a value of `scaleMax` greater than the calculated value. Use of these two parameters can allow searching for features of a particular scale size, for instance searching for narrow absorption features.

The reconstruction has at least two iterations. The first iteration makes a first pass at the wavelet reconstruction (the process outlined in the 8 stages above), but the residual array will likely have some structure still in it, so the wavelet filtering is done on the residual, and any significant wavelet terms are added to the final reconstruction. This step is repeated until the relative change

[5] *à trous* means "with holes" in French.

in the measured standard deviation of the residual (see note below on the evaluation of this quantity) is less than some value, given by the `reconConvergence` parameter.

It is important to note that the *à trous* decomposition is an example of a "redundant" transformation. If no thresholding is performed, the sum of all the wavelet coefficient arrays and the final smoothed array is identical to the input array. The thresholding thus removes only the unwanted structure in the array.

Note that any BLANK pixels that are still in the cube will not be altered by the reconstruction – they will be left as BLANK so that the shape of the valid part of the cube is preserved.

3.4.2 Note on Statistics

The correct calculation of the reconstructed array needs good estimators of the underlying mean and standard deviation (or rms) of the background noise distribution. The methods used to estimate these quantities are detailed in §3.7.3 – the default behaviour is to use robust estimators, to avoid biasing due to bright pixels.

When thresholding the different wavelet scales, the value of the rms as measured from the wavelet array needs to be scaled to account for the increased amount of correlation between neighbouring pixels (due to the convolution). See Appendix J for details on this scaling.

3.4.3 User control of reconstruction parameters

The most important parameter for the user to select in relation to the reconstruction is the threshold for each wavelet array. This is set using the `snrRecon` parameter, and is given as a multiple of the rms (estimated by the MADFM) above the mean (which for the wavelet arrays should be approximately zero). There are several other parameters that can be altered as well that affect the outcome of the reconstruction.

By default, the cube is reconstructed in three dimensions, using a three-dimensional filter and three-dimensional convolution. This can be altered, however, using the parameter `reconDim`. If set to 1, this means the cube is reconstructed by considering each spectrum separately, whereas `reconDim=2` will mean the cube is reconstructed by doing each channel map separately. The merits of these choices are discussed in §6, but it should be noted that a 2-dimensional reconstruction can be susceptible to edge effects if the spatial shape of the pixel array is not rectangular.

The user can also select the minimum and maximum scales to be used in the reconstruction. The first scale exhibits the highest frequency variations, and so ignoring this one can sometimes be beneficial in removing excess noise. The default is to use all scales (`minscale = 1`).

The convergence of the *à trous* iterations is governed by the `reconConvergence` parameter, which is the fractional decrease in the standard deviation of the residuals from one iteration to the next. *Duchamp* will do at least two iterations, and then continue until the decrease is less than the value of this parameter.

Finally, the filter that is used for the convolution can be selected by using `filterCode` and the relevant code number – the choices are listed in Appendix B. A larger filter will give a better reconstruction, but take longer and use more memory when executing. When multi-dimensional reconstruction is selected, this filter is used to construct a 2- or 3-dimensional equivalent.

3.5 Smoothing the cube

An alternative to doing the wavelet reconstruction is to smooth the cube. This technique can be useful in reducing the noise level (at the cost of making neighbouring pixels correlated and blurring any signal present), and is particularly well suited to the case where a particular signal size (i.e. a certain channel width or spatial size) is believed to be present in the data.

There are two alternative methods that can be used: spectral smoothing, using the Hanning filter; or spatial smoothing, using a 2D Gaussian kernel. These alternatives are outlined below. To utilise the smoothing option, set the parameter `flagSmooth=true` and set `smoothType` to either `spectral` or `spatial`.

3.5.1 Spectral smoothing

When `smoothType = spectral` is selected, the cube is smoothed only in the spectral domain. Each spectrum is independently smoothed by a Hanning filter, and then put back together to form the smoothed cube, which is then used by the searching algorithm (see below). Note that in the case of both the reconstruction and the smoothing options being requested, the reconstruction will take precedence and the smoothing will *not* be done.

There is only one parameter necessary to define the degree of smoothing – the Hanning width a (given by the user parameter `hanningWidth`). The coefficients $c(x)$ of the Hanning filter are defined by

$$c(x) = \begin{cases} \frac{1}{2}\left(1+\cos(\frac{\pi x}{a})\right) & |x| < (a+1)/2 \\ 0 & |x| \geq (a+1)/2. \end{cases}, \quad a, x \in \mathbb{Z}$$

Note that the width specified must be an odd integer (if the parameter provided is even, it is incremented by one).

3.5.2 Spatial smoothing

When `smoothType = spatial` is selected, the cube is smoothed only in the spatial domain. Each channel map is independently smoothed by a two-dimensional Gaussian kernel, put back together to form the smoothed cube, and used in the searching algorithm (see below). Again, reconstruction is always done by preference if both techniques are requested.

The two-dimensional Gaussian has three parameters to define it, governed by the elliptical cross-sectional shape of the Gaussian function: the FWHM (full-width at half-maximum) of the major and minor axes, and the position angle of the major axis. These are given by the user parameters `kernMaj, kernMin` & `kernPA`. If a circular Gaussian is required, the user need only provide the `kernMaj` parameter. The `kernMin` parameter will then be set to the same value, and `kernPA` to zero. If we define these parameters as a, b, θ respectively, we can define the kernel by the function

$$k(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y} \exp\left[-0.5\left(\frac{X^2}{\sigma_X^2} + \frac{Y^2}{\sigma_Y^2}\right)\right]$$

where (x, y) are the offsets from the central pixel of the gaussian function, and

$$\begin{aligned} X &= x\sin\theta - y\cos\theta & Y &= x\cos\theta + y\sin\theta \\ \sigma_X^2 &= \frac{(a/2)^2}{2\ln 2} & \sigma_Y^2 &= \frac{(b/2)^2}{2\ln 2} \end{aligned}$$

The size of the kernel is determined both by these FWHM parameters and the cutoff parameter `spatialSmoothCutoff`. The width is determined by

$$W = \sigma_{\rm maj}\sqrt{(2\log(C))}$$

where C is the cutoff value. The value of W is rounded up to get the half-width of the kernel in pixels. The default value of `spatialSmoothCutoff` is 1.e-10, which gives full kernel widths of 31 pix for `kernMaj=5` and 19 pix for `kernMaj=3`.

The algorithm provides different ways, controlled by the input parameter `smoothEdgeMethod`, to handle those pixels that lie within a kernel half-width from the edge of the image and so cannot have the full kernel placed on top of them. The default behaviour, with `smoothEdgeMethod=equal`, simply treats these pixels in the same way, adding up the product of the kernel pixels with the image pixels for all cases that lie within the image boundary. A more drastic alternative, `smoothEdgeMethod=truncate`, is to simply not evaluate the convolution at these pixels – they will be set at some null value that will not contribute to any detection. Thirdly, setting `smoothEdgeMethod=scale` will evaluate the convolution as for the `'equal'` case, but scale down the edge pixels by summing over only those kernel pixels contributing.

3.6 Input/Output of reconstructed/smoothed arrays

The smoothing and reconstruction stages can be relatively time-consuming, particularly for large cubes and reconstructions in 3-D (or even spatial smoothing). To get around this, *Duchamp* provides a shortcut to allow users to perform multiple searches (e.g. with different thresholds) on the same reconstruction/smoothing setup without re-doing the calculations each time.

To save the reconstructed array as a FITS file, set `flagOutputRecon = true`. The file will be saved in the same directory as the input image, so the user needs to have write permissions for that directory.

The name of the file can given by the `fileOutputRecon` parameter, but this can be ignored and *Duchamp* will present a name based on the reconstruction parameters. The filename will be derived from the input filename, with extra information detailing the reconstruction that has been done. For example, suppose `image.fits` has been reconstructed using a 3-dimensional reconstruction with filter #2, thresholded at 4σ using all scales from 1 to 5, with a convergence criterion of 0.005. The output filename will then be `image.RECON-3-2-4-1-5-0.005.fits` (i.e. it uses the six parameters relevant for the *à trous* reconstruction as listed in Appendix B). The new FITS file will also have these parameters as header keywords. If a subsection of the input image has been used (see §3.1), the format of the output filename will be `image.sub.RECON-3-2-4-1-5-0.005.fits`, and the subsection that has been used is also stored in the FITS header.

Likewise, the residual image, defined as the difference between the input and reconstructed arrays, can also be saved in the same manner by setting `flagOutputResid = true`. Its filename will be the same as above, with `RESID` replacing `RECON`.

If a reconstructed image has been saved, it can be read in and used instead of redoing the reconstruction. To do so, the user should set the parameter `flagReconExists = true`. The user can indicate the name of the reconstructed FITS file using the `reconFile` parameter, or, if this is not specified, *Duchamp* searches for the file with the name as defined above. If the file is not found, the reconstruction is performed as normal. Note that to do this, the user needs to set `flagAtrous = true` (obviously, if this is `false`, the reconstruction is not needed).

To save the smoothed array, set `flagOutputSmooth = true`. As for the reconstructed/residual arrays, the name of the file can given by the parameter `fileOutputSmooth`, but this can be ignored and *Duchamp* will present a name that indicates the both the type and the details of the smoothing method used. It will be either `image.SMOOTH-1D-a.fits`, where a is replaced by the Hanning width used, or `image.SMOOTH-2D-a-b-c-X-f.fits`, where a,b,c are the Gaussian kernel parameters, X represents the edge method (E=equal, T=truncate, S=scale), and f is $-\log_{10}(\text{cutoff})$. Similarly to the reconstruction case, a saved file can be read in by setting `flagSmoothExists = true` and either specifying a file to be read with the `smoothFile` parameter or relying on *Duchamp* to find the file with the name as given above.

3.7 Searching the image

3.7.1 Representation of detected objects

The principle aim of *Duchamp* is to provide a catalogue of sources located in the image. While running, *Duchamp* needs to maintain for each source several data structures that will contribute to the memory footprint: a record of which pixels contribute to the source; a set of measured parameters that will go into the catalogue; and a separate two-dimensional map showing the spatial location of detected pixels (carrying this around makes the computation of detection maps easier – see §5.3.2).

To keep track of the set of detected pixels, *Duchamp* employs specialised techniques that keep the memory usage manageable. A naive method could be to store each single pixel, but this results in a lot of redundant information being stored in memory.

To reduce the storage requirements, the run-length encoding method is used for storing the spatial information. In this fashion, an object in 2D is stored as a series of "runs", encoded by a row number (the y-value), the starting column (the minimum x-value) and the run length (ℓ_x: the number of contiguous pixels in that row connected to the starting pixel). A single set of (y, x, ℓ_x) values is called a "scan". A two-dimensional image is therefore made up of a set of scans.

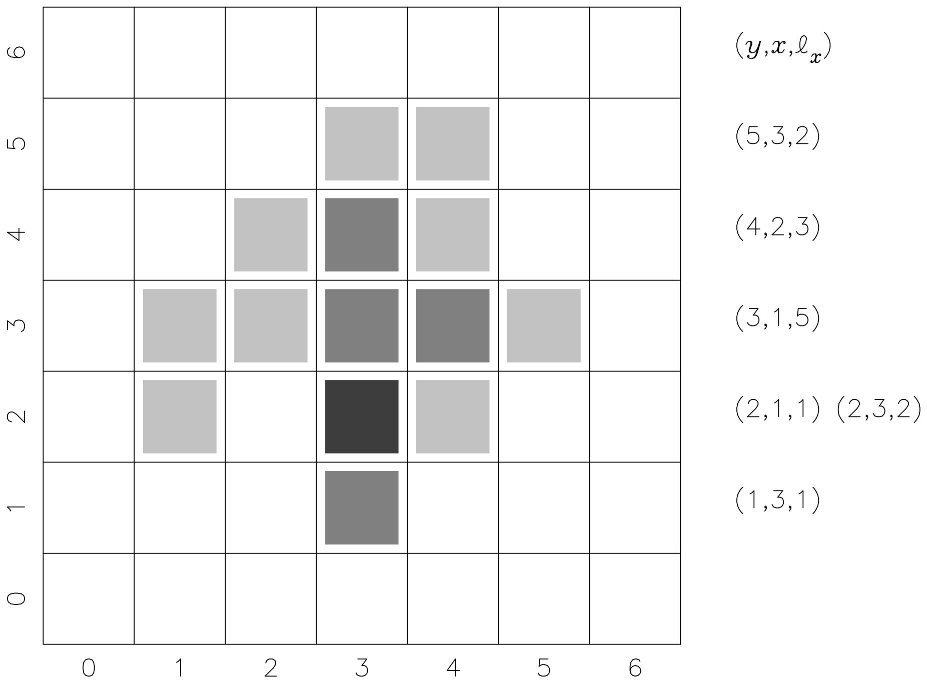


Figure 1: An example of the run-length encoding method of storing pixel information. The scans used to encode the image are listed alongside the relevant row. The pixels are colour-coded by nominal pixel values, but note that the pixel values themselves do not form part of the encoding and are not kept as part of the object class.

An example can be seen in Fig. 1. Note that the object shown has fourteen pixels, and so would require 28 integers to record the positions of all pixels. The run-length encoding uses just 18 integers to record the same information. The longer the runs are in each scan, the greater the saving of storage over the naive method.

A 3D object is stored as a set of channel maps, with a channel map being a 2D plane with constant z-value. Each channel map is itself a set of scans showing the (x, y) position of the pixels. The additional detection map is stored as a separate channel map, also made up of scans.

Note that these pixel map representations do not carry the flux information with them. They store just the pixel locations and need to be combined with an array of flux values to provide parameters such as integrated flux. The advantage of this approach is that the pixel locations can be easily applied to different flux arrays as the need permits (for instance, defining them using the reconstructed array, yet evaluating parameters on the original array).

This scan-based run-length encoding is how the individual detections are stored in the binary catalogue described in §5.5.1.

3.7.2 Technique

The basic idea behind detection in *Duchamp* is to locate sets of contiguous voxels that lie above some threshold. No size or shape requirement is imposed upon the detections, and no fitting (for instance, fitting Gaussian profiles) is done on the sources. All *Duchamp* does is find connected groups of bright voxels and report their locations and basic parameters.

One threshold is calculated for the entire cube, enabling calculation of signal-to-noise ratios for each source (see §5 for details). The user can manually specify a value (using the parameter `threshold`) for the threshold, which will override the calculated value. Note that this option overrides any settings of `snrCut` or FDR options (see below).

The cube can be searched in one of two ways, governed by the input parameter `searchType`. If `searchType=spatial`, the cube is searched one channel map at a time, using the 2-dimensional raster-scanning algorithm of Lutz (1980) that connects groups of neighbouring pixels. Such an algorithm cannot be applied directly to a 3-dimensional case, as it requires that objects are completely nested in a row (when scanning along a row, if an object finishes and other starts, you won't get back to the first until the second is completely finished for the row). Three-dimensional data does not have this property, hence the need to treat the data on a 2-dimensional basis at most.

Alternatively, if `searchType=spectral`, the searching is done in one dimension on each individual spatial pixel's spectrum. This is a simpler search, but there are potentially many more of them.

Although there are parameters that govern the minimum number of pixels in a spatial, spectral and total sense that an object must have (`minPix`, `minChannels` and `minVoxels` respectively), these criteria are not applied at this point - see §3.8.4 for details.

Finally, the search only looks for positive features. If one is interested instead in negative features (such as absorption lines), set the parameter `flagNegative = true`. This will invert the cube (i.e. multiply all pixels by -1) prior to the search, and then re-invert the cube (and the fluxes of any detections) after searching is complete. If the reconstructed or smoothed array has been read in from disk, this will also be inverted at the same time. All outputs are done in the same manner as normal, so that fluxes of detections will be negative.

3.7.3 Calculating statistics

A crucial part of the detection process (as well as the wavelet reconstruction: §3.4) is estimating the statistics that define the detection threshold. To determine a threshold, we need to estimate from the data two parameters: the middle of the noise distribution (the "noise level"), and the width of the distribution (the "noise spread"). The noise level is estimated by either the mean or the median, and the noise spread by the rms (or the standard deviation) or the median absolute deviation from the median (MADFM). The median and MADFM are robust statistics, in that they are not biased by the presence of a few pixels much brighter than the noise.

All four statistics are calculated automatically, but the choice of parameters that will be used is governed by the input parameter `flagRobustStats`. This has the default value `true`, meaning the underlying mean of the noise distribution is estimated by the median, and the underlying standard deviation is estimated by the MADFM. In the latter case, the value is corrected, under the assumption that the underlying distribution is Normal (Gaussian), by dividing by 0.6744888 – see Appendix I for details. If `flagRobustStats=false`, the mean and rms are used instead.

The choice of pixels to be used depend on the analysis method. If the wavelet reconstruction has been done, the residuals (defined in the sense of original − reconstruction) are used to estimate the noise spread of the cube, since the reconstruction should pick out all significant structure. The noise level (the middle of the distribution) is taken from the original array.

If smoothing of the cube has been done instead, all noise parameters are measured from the smoothed array, and detections are made with these parameters. When the signal-to-noise level is quoted for each detection (see §5), the noise parameters of the original array are used, since the smoothing process correlates neighbouring pixels, reducing the noise level.

If neither reconstruction nor smoothing has been done, then the statistics are calculated from the original, input array.

The parameters that are estimated should be representative of the noise in the cube. For the case of small objects embedded in many noise pixels (e.g. the case of HI surveys), using the full cube will provide good estimators. It is possible, however, to use only a subsection of the cube by setting the parameter `flagStatSec = true` and providing the desired subsection to the `StatSec` parameter. This subsection works in exactly the same way as the pixel subsection discussed in §3.1. The `StatSec` will be trimmed if necessary so that it lies wholly within the image subsection being used (i.e. that given by the `subsection` parameter - this governs what pixels are read in and so are able to be used in the calculations).

Note that `StatSec` applies only to the statistics used to determine the threshold. It does not affect the calculation of statistics in the case of the wavelet reconstruction. Note also that pixels identified as BLANK or as flagged via the `flaggedChannels` parameter are ignored in the statistics calculations.

3.7.4 Determining the threshold

Once the statistics have been calculated, the threshold is determined in one of two ways. The first way is a simple sigma-clipping, where a threshold is set at a fixed number n of standard deviations above the mean, and pixels above this threshold are flagged as detected. The value of n is set with the parameter `snrCut`. The “mean” and “standard deviation” here are estimated according to `flagRobustStats`, as discussed in §3.7.3. In this first case only, if the user specifies a threshold, using the `threshold` parameter, the sigma-clipped value is ignored.

The second method uses the False Discovery Rate (FDR) technique (Hopkins et al. 2002; Miller et al. 2001), whose basis we briefly detail here. The false discovery rate (given by the number of false detections divided by the total number of detections) is fixed at a certain value α (e.g. $\alpha = 0.05$ implies 5% of detections are false positives). In practice, an α value is chosen, and the ensemble average FDR (i.e. $\langle FDR \rangle$) when the method is used will be less than α. One calculates p – the probability, assuming the null hypothesis is true, of obtaining a test statistic as extreme as the pixel value (the observed test statistic) – for each pixel, and sorts them in increasing order. One then calculates d where

$$d = \max_j \left\{ j : P_j < \frac{j\alpha}{c_N N} \right\},$$

and then rejects all hypotheses whose p-values are less than or equal to P_d. (So a $P_i < P_d$ will be rejected even if $P_i \geq j\alpha/c_N N$.) Note that "reject hypothesis" here means "accept the pixel as an object pixel" (i.e. we are rejecting the null hypothesis that the pixel belongs to the background).

The c_N value here is a normalisation constant that depends on the correlated nature of the pixel values. If all the pixels are uncorrelated, then $c_N = 1$. If N pixels are correlated, then their tests will be dependent on each other, and so $c_N = \sum_{i=1}^{N} i^{-1}$. Hopkins et al. (2002) consider real radio data, where the pixels are correlated over the beam. For the calculations done in *Duchamp*, $N = B \times C$, where B is the beam area in pixels, calculated from the FITS header (if the correct keywords – BMAJ, BMIN – are not present, the size of the beam is taken from the input parameters - see discussion in §5.2.1, and if these parameters are not given, $B = 1$), and C is the number of neighbouring channels that can be considered to be correlated.

The use of the FDR method is governed by the `flagFDR` flag, which is `false` by default. To set the relevant parameters, use `alphaFDR` to set the α value, and `FDRnumCorChan` to set the C value discussed above. These have default values of 0.01 and 2 respectively.

The theory behind the FDR method implies a direct connection between the choice of α and the fraction of detections that will be false positives. These detections, however, are individual pixels, which undergo a process of merging and rejection (§3.8), and so the fraction of the final list of detected objects that are false positives will be much smaller than α. See the discussion in §6.

3.8 Merging, growing and rejecting detected objects

3.8.1 Merging

The searches described above are either 1- or 2-dimensional only. They do not know anything about the third dimension that is likely to be present. To build up 3D sources, merging of detections must be done. This is done via an algorithm that matches objects judged to be "close", according to one of two criteria.

One criterion is to define two thresholds – one spatial and one in velocity – and say that two objects should be merged if there is at least one pair of pixels that lie within these threshold distances of each other. These thresholds are specified by the parameters `threshSpatial` and `threshVelocity` (in units of pixels and channels respectively).

Alternatively, the spatial requirement can be changed to say that there must be a pair of pixels that are *adjacent* – a stricter, but perhaps more realistic requirement, particularly when the spatial pixels have a large angular size (as is the case for HI surveys). This method can be selected by setting the parameter `flagAdjacent=true` in the parameter file. The velocity thresholding is always done with the `threshVelocity` test.

3.8.2 Stages of merging

This merging can be done in two stages. The default behaviour is for each new detection to be compared with those sources already detected, and for it to be merged with the first one judged to be close. No other examination of the list is done at this point.

This step can be turned off by setting `flagTwoStageMerging=false`, so that new detections are simply added to the end of the list, leaving all merging to be done in the second stage.

The second, main stage of merging is more thorough, Once the searching is completed, the list is iterated through, looking at each pair of objects, and merging appropriately. The merged objects are then included in the examination, to see if a merged pair is suitably close to a third.

3.8.3 Growing

Once the detections have been merged, they may be "grown" (this is essentially the process known elsewhere as "floodfill"). This is a process of increasing the size of the detection by adding nearby pixels (according to the `threshSpatial` and `threshVelocity` parameters) that are above some secondary threshold and not already part of a detected object. This threshold should be lower than the one used for the initial detection, but above the noise level, so that faint pixels are only detected when they are close to a bright pixel. This threshold is specified via one of two input parameters. It can be given in terms of the noise statistics via `growthCut` (which has a default value of 3σ), or it can be directly given via `growthThreshold`. Note that if you have given the detection threshold with the `threshold` parameter, the growth threshold **must** be given with `growthThreshold`. If `growthThreshold` is not provided in this situation, the growing will not be done.

The use of the growth algorithm is controlled by the `flagGrowth` parameter – the default value of which is `false`. If the detections are grown, they are sent through the merging algorithm a second time, to pick up any detections that should be merged at the new lower threshold (i.e. they have grown into each other).

3.8.4 Rejecting

Finally, to be accepted, the detections must satisfy minimum (and, optionally, maximum) size criteria, relating to the number of channels, spatial pixels and voxels occupied by the object. These criteria are set using the `minChannels`, `minPix` and `minVoxels` parameters respectively. The channel requirement means a source must have at least one set of `minChannels` consecutive channels to be accepted. The spatial pixels (`minPix`) requirement refers to distinct spatial pixels (which are possibly in different channels), while the voxels requirement refers to the total number of voxels detected. If the `minVoxels` parameter is not provided, it defaults to `minPix`+`minChannels`-1.

It is possible to also place upper limits on the size of detections in a similar fashion. The corresponding parameters are `maxPix`, `maxChannels` and `maxVoxels`. These will all default to a value of -1 – if they are negative the check is **not** made. If they are provided by the user, an object will be rejected if one of the metrics exceeds the limit (in each case, the value is the highest allowed value for an object to be accepted). The channel requirement differs slightly from the minimum check – the limit applies to the total number of channels in the object.

It is possible to do this rejection stage before the main merging and growing stage. This could be done to remove narrow (hopefully spurious) sources from the list before growing them, to reduce the number of false positives in the final list. This mode can be selected by setting the input parameter `flagRejectBeforeMerge=true` – caution is urged if you use this in conjunction with `flagTwoStageMerging=false`, as you can throw away parts of objects that you may otherwise wish to keep.

4 Source Parameterisation

Once sources have been located, numerous measurements are made so that they can be placed in a catalogue. This section details each of the source parameters, explaining what they are and how they are calculated. Each parameter is referred to by the heading of the relevant column(s) in the output source list (see §5).

4.1 Object ID, `Obj#`

The ID of the detection is an integer, simply the sequential count for the list. The default is ordering by increasing spectral coordinate, or channel number, if the WCS is not good enough to determine the spectral world coordinate, but this can be changed by the `sortingParam` input parameter. See Sec 5.2.1 for details.

4.2 Object Name, `Name`

This is the "IAU"-format name of the detection, derived from the WCS position if available. For instance, a source that is centred on the RA,Dec position $12^h53^m45^s$, -36°24′12″ will be given the name J125345−362412, if the epoch is J2000, or the name B125345−362412 if it is B1950. The precision of the RA and Dec values is determined by the pixel size, such that sufficient precision is used to uniquely define a position. The RA value will have one figure greater precision than Dec.

An alternative form is used for Galactic coordinates: a source centred on the position $(l,b) = (323.1245, 5.4567)$ will be called G323.124+05.457.

If the WCS is not valid (i.e. is not present or does not have all the necessary information), the name will instead be of the form "ObjXXX", where XXX is replaced with the objectID, padded sufficiently with zeros.

4.3 Pixel locations

There are three ways in which the pixel location of the detection is calculated:

- Peak: the pixel value in which the detection has its peak flux. Appears in the results file under columns `X_peak, Y_peak, Z_peak`.
- Average: the average over all detected pixels. Specifically, $x_{\rm av} = \sum x_i/N$ and similarly for $y_{\rm av}$ and $z_{\rm av}$. Appears in the results file under columns `X_av, Y_av, Z_av`.
- Centroid: the flux-weighted average over all detected pixels. Specifically, $x_{\rm cent} = \sum F_i x_i / \sum F_i$ and similarly for $y_{\rm cent}$ and $z_{\rm cent}$. Appears in the results file under columns `X_cent, Y_cent, Z_cent`.

All three alternatives are calculated, and written to the results file, but the choice of the `pixelCentre` input parameter will determine which option is used for the reference values `X, Y, Z`.

4.4 Spatial world location, RA/GLON, DEC/GLAT

These are the conversion of the X and Y pixel positions to world coordinates (that is, the pixel position determined by pixelCentre). These will typically be Right Ascension and Declination, or Galactic Longitude and Galactic Latitude, but the actual names used in the output file will be taken from the WCS specification.

If there is no useful WCS, these are not calculated.

4.5 Spectral world location, VEL

The conversion of the Z pixel position to the spectral world coordinates. This is dictated by the WCS of the FITS file plus the input parameter spectralType. The name of the output column will come from the CTYPE of the spectral axis (or spectralType – see §3.2), specifically , the 4-character S-type code) (i.e. not necessarily "VEL")

The spectral units can be specified by the user, using the input parameter spectralUnits (enter it as a single string with no spaces). The default value comes from the FITS header.

4.6 Spatial size, MAJ, MIN, PA

The spatial size of the detection is measured from the moment-0 map (in the case of three-dimensional data) or the two-dimensional image, and is parameterised by the FWHM of the major and minor axes, plus the position angle of the major axis.

The position angle will be measured in the usual astronomical sense, in degrees East of North. The major and minor axes will be specified in angular units (assuming the WCS allows this), with the units chosen such that the numbers are not too small.

The method for calculating these parameters is to form the moment-0 map (if necessary), select all pixels greater than half the maximum [6], then calculate the parameters a (major FWHM), b (minor FWHM) and θ (position angle) according to

$$\begin{aligned}
\frac{1}{2}a^2 &= S_{xx} + S_{yy} + \sqrt{(S_{xx} - S_{yy})^2 + 4(S_{xy})^2} \\
\frac{1}{2}b^2 &= S_{xx} + S_{yy} - \sqrt{(S_{xx} - S_{yy})^2 + 4(S_{xy})^2} \\
\tan 2\theta &= \frac{2S_{xy}}{S_{xx} - S_{yy}}
\end{aligned}$$

where the sums S_{xx}, S_{yy} and S_{xy} are calculated in one of two ways. First, the

[6] In the event of a negative search (see §3.7.2), the moment map is inverted prior to this selection.

algorithm tries to weight each pixel by its flux:

$$\begin{aligned} S_{xx} &= \sum F_i x_i^2 / \sum F_i \\ S_{yy} &= \sum F_i y_i^2 / \sum F_i \\ S_{xy} &= \sum F_i x_i y_i / \sum F_i \end{aligned}$$

Mostly, this will work. But there can be situations where the calculated value of b^2 is negative (that is, $S_{xx} + S_{yy} < \sqrt{(S_{xx} - S_{yy})^2 + 4S_{xy}^2}$, or $S_{xx}S_{yy} < S_{xy}^2$). These situations are often where the moment-0 map is very disordered with no clear primary axis.

In this case, the calculation of the sums is redone without the flux weighting ($S_{xx} = \sum x_i^2$ etc), and the shape parameters recalculated. A **W** flag will be recorded for the detection to indicate that the weighting failed: see §4.14 below.

It is possible that this second calculation will fail, particularly for sources with only a small number of spatial pixels. In this case, we revert to using the method of principle axes.

We first calculate the angle of minimum moment and assign this as the position angle. This is defined by calculating the net moments:

$$\begin{aligned} M_{xx} &= \sum x_i^2 - \frac{1}{N}\left(\sum x_i\right)^2 \\ M_{yy} &= \sum y_i^2 - \frac{1}{N}\left(\sum y_i\right)^2 \\ M_{xy} &= \sum x_i y_i - \frac{1}{N}\sum x_i \sum y_i \end{aligned}$$

then

$$\tan\theta = \frac{(M_{xx} - M_{yy} + \sqrt{(M_{xx} - M_{yy})^2 + 4M_{xy}})}{2M_{xy}}.$$

To find the sizes of the principle axes (the major axis aligning with the angle just calculated, and the minor being perpendicular to it), we calculate the projection along these two axes of each pixel above half the peak in the moment-0 map, and take the range between the maximum and minimum, requiring it to be at least one pixel. Note this is not sensitive to the flux distribution. If this calculation is used, a **w** flag will be recorded for the detection: see §4.14 below.

4.7 Spatial extent, `w_RA/w_GLON`, `w_DEC`

The extent of the detection in each of the spatial directions is also calculated. This is simply the angular width of the detection (in arcmin), converting the minimum and maximum values of x (usually RA) and y (Dec) to the world coordinates.

4.8 Spectral width, `w_50`, `w_20`, `w_VEL`

Several measures of the spectral extent of a detection are provided. The simplest is the full spectral width, calculated as for the spatial extents above. This is referred to as `w_VEL`, but need not be velocity. It is obtained by taking the difference in world coordinates of the minimum and maximum values of z. The units are as described in §4.5.

Two other measures of the spectral width are provided, `w_50` and `w_20`, being the width at 50% and 20% of the peak flux. These are measured on the integrated spectrum (i.e. the spectra of all detected spatial pixels summed together), and are defined by starting at the outer spectral extent of the object (the highest and lowest spectral values) and moving in or out until the required flux threshold is reached. The widths are then just the difference between the two values obtained. If the detection threshold is greater than 20% or 50% of the peak, then these values will be the same as `w_VEL`.

4.9 Source flux, `F_tot`, `F_int`, `F_peak`

There are two measurements of the total flux of the detection. The simplest, `F_tot`, is just the sum of all detected pixels in the image: $F_{\mathrm{tot}} = \sum F_i$. The alternative, `F_int`, is the flux integrated over the detected pixels, taking into account the spectral range. For the case of velocity, the expression is $F_{\mathrm{int}} = \sum F_i \Delta v_i$, where Δv_i is the velocity width of the channel containing pixel i. The actual units of the spectral range are as described in §4.5.

When the cube brightness units are quoted per beam (e.g. Jy/beam), then the integrated flux `F_int` includes a correction for this. This involves dividing by the integral over the beam. This is calculated using the BMAJ, BMIN & BPA header keywords from the FITS file. BMAJ and BMIN are assumed to be the full-width at half maximum (FWHM) in the major and minor axis directions of a Gaussian beam. The integral is calculated as follows: the functional form of a 2D elliptical Gaussian can be written as $\exp(-((x^2/2\sigma_x^2) + (y^2/2\sigma_y^2)))$, and the FWHM in a given direction is then $f = 2\sqrt{2\ln 2}\sigma$. Then, $F_{\mathrm{int}} = C \sum F_i \Delta v_i$, where

$$C = \int \exp\left(-\left(\frac{x^2}{2\sigma_x^2} + \frac{y^2}{2\sigma_y^2}\right)\right) = 2\pi\sigma_x\sigma_y = \frac{\pi f_x f_y}{4\ln 2}$$

(with f_x = BMAJ and f_y = BMIN) provides the correction factor to go from units of Jy/beam to Jy.

If the FITS file does not have the beam information, the user can either:

1. Specify the FWHM of the beam in pixels (assuming a circular beam) via the parameter `beamFWHM`.
2. Specify the area of the beam, again in pixels, via the parameter `beamArea`[7]. This should be the value given by the equation above.

If both are given, `beamFWHM` takes precendence. If neither are given, and there is no beam information in the header, then no correction to the integrated flux is made (and so it will stay in units of Jy/beam or equivalent).

Note that these parameters are measured using *only the detected pixels*. The summing of the flux will not include voxels that fall below the detection (or growth) threshold – which is in accord with the definition of the threshold as dividing source and non-source voxels. If the threshold is not low enough, this

[7] Note that this is equivalent to the old parameter `beamSize`, which was highlighted as being ambiguous.

will bias the measurement of the fluxes. This applies to all parameters (with the exception of the w_50 and w_20 widths, which are measured from the integrated spectrum, including channels not necessarily forming part of the detection).

Finally, the peak flux F_peak is simply the maximum value of the flux over all the detected pixels.

4.10 Error on total/integrated flux, eF_tot, eF_int

Both F_tot and F_int can also have their associated error calculated (eF_tot and eF_int respectively). This is the random error due to the noise in the image, and is simply the sum in quadrature of the noise on each of the voxels, and, in the case of F_int multiplied by the spectral width and corrected for the beam if necessary. Since we assume a constant noise level in the image ($\sigma_i = \sigma\ \forall i$), we have:

$$\begin{aligned} eF_{\mathrm{int}} &= \sqrt{\sum \sigma_i^2} \\ &= \sigma\sqrt{N} \\ eF_{\mathrm{tot}} &= \sqrt{\sum C^2 \sigma_i^2 \Delta v_i^2} \\ &= C\sigma\sqrt{N}\Delta v \text{ (for the case of } \Delta v_i = \Delta v\text{)} \end{aligned}$$

In the case that a flux threshold is provided, these quantities are not calculated, since we don't measure the image noise statistics. Likewise, when the array is smoothed we measure the noise only in the smoothed image, and this value is not applicable to the flux measured from the original image, so the errors are not reported.

4.11 Peak signal-to-noise, S/Nmax

This parameter converts the peak flux to a signal-to-noise value, based on the measured noise level in the image. As for the error quantities above, if no noise is measured (i.e. a flux threshold is provided by the user), then this is not calculated.

When the array is pre-processed (via smoothing or wavelet reconstruction), we take the peak flux here to be the peak in the smoothed or reconstructed array. This is because this is where the detection is made, and so the S/Nmax value can be directly compared to the requested signal-to-noise threshold. Note that the peak flux discussed in §4.9 is always measured from the original image array.

4.12 Pixel ranges, X1, X2, Y1, Y2, Z1, Z2

These quantities give the range of pixel values spanned by the detection in each of the three axes. X1, Y1, Z1 give the minimum pixel in each direction, while X2, Y2, Z2 give the maximum pixel.

4.13 Size, Nvoxel, Nchan, Nspatpix

The number of detected pixels that make up the detection. These quantities show the total number of voxels in the detection (Nvoxel), the number of distinct spectral channels (`Nchan`) and the number of distinct spatial pixels (`Nspatpix`).

Note that `Nchan` here is different to the quantity tested by the input parameter `minChannels` (see §3.8.4) – this looks at the maximum number of *adjacent* channels, not the total.

4.14 Warning Flags, `Flag`

The detection can have warning flags recorded, to highlight potential issues:

- **E** – The detection is next to the spatial edge of the image, meaning either the limit of the pixels, or the limit of the non-BLANK pixel region.
- **S** – The detection lies at the edge of the spectral region.
- **F** – The detection is adjacent to, or overlaps one or more flagged channels (see §3.3.5).
- **N** – The total flux, summed over all the (non-BLANK) pixels in the smallest box that completely encloses the detection, is negative. Note that this sum is likely to include non-detected pixels. It is of use in pointing out detections that lie next to strongly negative pixels, such as might arise due to interference – the detected pixels might then also be due to the interference, so caution is advised.
- **W** – The weighting of fluxes in the shape calculation (Sec 4.6) failed, so the unweighted calculation was used. This likely indicates some very disordered shape for the moment-0 map.
- **w** – The unweighted calculation also failed, most likely because there are too few pixels. In this case, we use the alternate method of principle axes to estimate the shape.

In the absence of any of these flags, a `-` will be recorded.

5 Outputs

5.1 During execution

Duchamp provides the user with feedback whilst it is running, to keep the user informed on the progress of the analysis. Most of this consists of self-explanatory messages about the particular stage the program is up to. The relevant parameters are printed to the screen at the start (once the file has been successfully read in), so the user is able to make a quick check that the setup is correct (see Appendix C for an example).

The extent of memory allocation made at the start is indicated. This will include the arrays needed for the pixel array, the reconstruction or smoothed array, and the 2D detection map, but *not* additional space needed for working within individual algorithms, nor storage needed for the detected objects.

Duchamp will report the amount of memory that is allocated when the image is read in. This includes the storage for the array as well as additional storage for the reconstructed/smoothed array and/or the baseline arrays (if these are needed).

If the cube is being trimmed (§3.3), the resulting dimensions are printed to indicate how much has been trimmed. If a reconstruction is being done, a continually updating message shows either the current iteration and scale, compared to the maximum scale (when `reconDim = 3`), or a progress bar showing the amount of the cube that has been reconstructed (for smaller values of `reconDim`).

During the searching algorithms, the progress through the search is shown. When completed, the number of objects found is reported (this is the total number found, before any merging is done).

In the merging process (where multiple detections of the same object are combined – see §3.8), two stages of output occur. The first is when each object in the list is compared with all others. The output shows two numbers: the first being how far through the list the current object is, and the second being the length of the list. As the algorithm proceeds, the first number should increase and the second should decrease (as objects are combined). When the numbers meet, the whole list has been compared. If the objects are being grown, a similar output is shown, indicating the progress through the list. In the rejection stage, in which objects not meeting the minimum pixels/channels requirements are removed, the total number of objects remaining in the list is shown, which should steadily decrease with each rejection until all have been examined. Note that these steps can be very quick for small numbers of detections.

Since this continual printing to screen has some overhead of time and CPU involved, the user can elect to not print this information by setting the parameter `verbose = false`. In this case, the user is still informed as to the steps being undertaken, but the details of the progress are not shown.

There may also be Warning or Error messages printed to screen. The Warning messages occur when something happens that is unexpected (for instance, a desired keyword is not present in the FITS header), but not detrimental to the execution. An Error message is something more serious, and indicates some part of the program was not able to complete its task. This is not necessary fatal, but it may mean the full functionality requested will not be achieved. The message will also indicate which function or subroutine generated it – this is primarily a tool for debugging, but can be useful in determining what went wrong.

5.2 Text-based output files

5.2.1 Table of results

Finally, we get to the results – the reason for running *Duchamp* in the first place. Once the detection list is finalised and parameterised according to §4, it is sorted according to the value of the `sortingParam`. This can take the value "xvalue", "yvalue", "zvalue", "ra", "dec", "vel", "w50", "iflux" (for integrated flux), or "pflux" (for peak flux), or "snr". The default value is "vel" (which means the spectral WCS value – this could be frequency or wavelength depending on the nature of the FITS file). If no good WCS exists, the mean pixel position equivalent is used ("ra" is replaced by "xvalue", "dec" by "yvalue", "vel" and "w50" by "zvalue"). The sense of the sorting will be increasing value with position in the list. To sort in the opposite sense, prepend the parameter name with a '-' (e.g. "-vel" instead of "vel"). The object ID number (§4.1) is determined by the order of the list *after* this sorting, so sorting by a different parameter will result in a different object ID for the same object.

The results are then printed to the screen and to the output file, given by the `OutFile` parameter. The output file will contain all calculated parameters, as described in §4. The results list printed to the screen, however, will leave out certain columns:

- The spatial extent columns `w_RA & w_DEC`.
- The `w_20` and `w_VEL` spectral width columns.
- The total flux `F_tot` (unless there is no good WCS, in which case it is printed instead of `F_int`), and the errors on the total and integrated fluxes `eF_tot, eF_int`.
- The explicit columns for the average, centroid and peak pixel locations. The only pixel location columns printed are `X, Y, Z`, which are determined via the `pixelCentre` input parameter.
- *If the WCS is no good*, the world-coordinate columns `RA, DEC, VEL, F_int` will not be printed either.

The output file consists of two sections. The first section contains the metadata for the search. First, a list of the parameters are printed to the output file, for future reference. Next, the detection threshold that was used is given, so comparison can be made with other searches. The statistics estimating the noise parameters are given (see §3.7.3). Thirdly, the number of detections are reported.

All this information, known as the "header", can either be written to the start of the output file (denoted by the parameter `OutFile`), or written to a separate file from the list of detections. This second option is activated by the parameter `flagSeparateHeader`, and the information is written to the file given by `HeaderFile`.

The second part of the file, however, contains the most interesting part – the list of detected objects. This is written as an ASCII table, properly spaced so that it is readable. An example is shown in Appendix D.

The user can specify the precision used to display the flux, spectral location/width and S/Nmax values, by using the input parameters `precFlux`, `precVel` and `precSNR` respectively. These values apply to the tables written to the screen and to the output file, as well as for the VOTable (see below).

5.2.2 VOTable catalogue

Three additional results files can also be requested. One option is a VOTable-format XML file, containing just the RA, Dec, spectral location and the corresponding widths of the detections, as well as the fluxes. The user should set `flagVOT = true`, and put the desired filename in the parameter `votFile` – note that the default is for it not to be produced. An example of VOTable output can be found in Appendix E. This file should be compatible with all Virtual Observatory tools (such as Aladin[8] or TOPCAT[9]).

5.2.3 Annotation and region files

A second option are annotation files for use with several visualisation tools, including the Karma toolkit (in particular, with `kvis`), SAOImage DS9, and `casaviewer` (and `casapy` itself).

There are three options on how objects are represented, governed by the `annotationType` parameter. These are:

- `borders` – a border is drawn around the spatial pixels of the object, in a manner similar to that seen in Fig. 2. Note that Karma/`kvis` does not always do this perfectly, particularly as you change the zoom, so the lines may not be directly lined up with pixel borders.

- `circles` – draws a circle at the position of each detection, scaled by the spatial size of the detection.

- `ellipses` – draws an ellipse of size given by the `MAJ, MIN, PA` source parameters (§4.6).

In each case, the object is numbered according to the object ID (§4.1. To make use of this option, the user should set `flagKarma`, `flagDS9` or `flagCasa` to `true`, and put the desired filename in the parameter `karmaFile`, `ds9File` or `casaFile` – again, the default is for these not to be produced. Examples of these annotation files are in Appendices F,G,H.

5.2.4 Spectral text file

The final optional results file produced is a simple text file that contains the spectra for each detected object. The format of the file is as follows: the first column has the spectral coordinate, over the full range of values; the remaining columns represent the flux values for each object at the corresponding spectral position. The flux value used is the same as that plotted in the spectral plot

[8] http://aladin.u-strasbg.fr/

[9] http://www.star.bristol.ac.uk/ mbt/topcat/

detailed below, and governed by the `spectralMethod` parameter. An example of what a spectral text file might look like is given below:

```
1405.00219727   0.01323344   0.23648241   0.04202826   -0.00506790
1405.06469727   0.01302835   0.27393046   0.04686056   -0.00471084
1405.12719727   0.01583361   0.27760920   0.04114933   -0.01168737
1405.18969727   0.01271889   0.31489247   0.03307962   -0.00300790
1405.25219727   0.01597644   0.30401203   0.05356426   -0.00551653
1405.31469727   0.00773827   0.30031312   0.04074831   -0.00570147
1405.37719727   0.00738304   0.27921870   0.05272378   -0.00504959
1405.43969727   0.01353923   0.26132512   0.03667958   -0.00151006
1405.50219727   0.01119724   0.28987029   0.03497849   -0.00645589
1405.56469727   0.00813379   0.29839963   0.04711142   0.00536576
1405.62719727   0.00774377   0.26530230   0.04620502   0.00724631
1405.68969727   0.00576067   0.23215000   0.04995513   0.00290841
1405.75219727   0.00452834   0.16484940   0.04261605   -0.00612812
1405.81469727   0.01406293   0.15989439   0.03817926   -0.00758385
1405.87719727   0.01116611   0.11890682   0.05499069   -0.00626362
1405.93969727   0.00687582   0.10620256   0.04743370   0.00055177
      ⋮               ⋮            ⋮            ⋮            ⋮
```

5.2.5 Log file

In addition to these three files, a log file can also be produced. As the program is running, it also (optionally) records the detections made in each individual spectrum or channel (see §3.7 for details on this process). This is recorded in the file given by the parameter `LogFile`. This file does not include the columns `Name, RA, DEC, w_RA, w_DEC, VEL, w_VEL`. This file is designed primarily for diagnostic purposes: e.g. to see if a given set of pixels is detected in, say, one channel image, but does not survive the merging process. The list of pixels (and their fluxes) in the final detection list are also printed to this file, again for diagnostic purposes. The file also records the execution time, as well as the command-line statement used to run *Duchamp*. The creation of this log file can be prevented by setting `flagLog = false` (which is the default).

5.3 Graphical output

5.3.1 Spectral plots

As well as the output data file, a postscript file (with the filename given by the `spectralFile` parameter) is created that shows the spectrum for each detection, together with a small cutout image (the 0th moment) and basic information about the detection (note that any flags are printed after the name of the detection, in the format `[E]`). If the cube was reconstructed, the spectrum from the reconstruction is shown in red, over the top of the original spectrum. If the spectral baseline was removed prior to source detection, this is shown in yellow. The spectral extent of the detected object is indicated by two solid blue lines, and the regions covered by the flagged channels are shown by green hashed boxes. An example detection can be seen in Fig. 2.

The spectrum that is plotted is governed by the `spectralMethod` parameter. It can be either `peak` (the default), where the spectrum is from the spatial pixel containing the detection's peak flux; or `sum`, where the spectrum is summed over all spatial pixels, and then corrected for the beam size. If the `peak` method is used, the detection threshold (and growth threshold, if used) are indicated

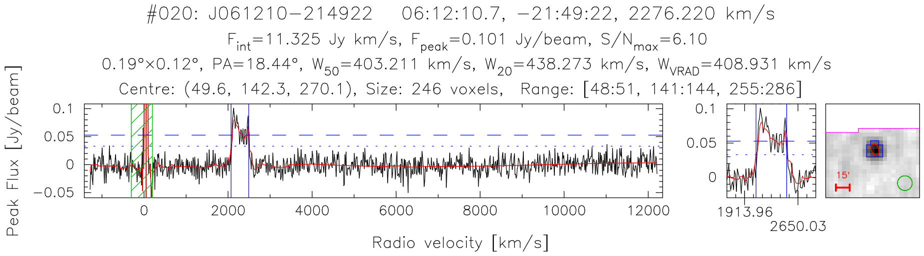


Figure 2: An example of the spectral output. Note several of the features discussed in the text: the red solid lines indicating the reconstructed spectrum; the blue dashed and dotted horizontal lines indicating the detection and growth thresholds respectively; the vertical blue solid lines indicating the spectral extent of the detection; the green hashed area indicating the flagged channels that are ignored by the searching algorithm; the blue border showing its spatial extent on the 0th moment map; the ellipses indicating the size of the object and the beam; and the 15 arcmin-long scale bar.

by dashed (and dotted) lines. When the spectral baseline has been removed, the thresholds will be a constant level above this (and so reflect its variability). Otherwise, the thresholds will be horizontal lines. The thresholds cannot be plotted on the integrated spectrum. The spectral extent of the detection is indicated with blue lines, and a zoom is shown in a separate window.

The cutout image shows a red ellipse indicating the spatial size of the detection (using `MAJ, MIN, PA` - §4.6). Also drawn in green in the corner of the image is an ellipse indicating the beam size (assuming the beam is defined).

The cutout image can optionally include a border around the spatial pixels that are in the detection (turned on and off by the `drawBorders` parameter – the default is `true`). It includes a scale bar in the bottom left corner to indicate size – its length is indicated next to it (the choice of length depends on the size of the image).

There may also be one or two extra lines on the image. A yellow line shows the limits of the cube's spatial region: when this is shown, the detected object will lie close to the edge, and the image box will extend outside the region covered by the data. A purple line, however, shows the dividing line between BLANK and non-BLANK pixels. The BLANK pixels will always be shown in black. The first type of line is always drawn, while the second is governed by the parameter `drawBlankEdges` (whose default is `true`), and obviously whether there are any BLANK pixel present.

Note that the creation of the spectral plots can be prevented by setting `flagPlotSpectra = false`.

When the input image is two-dimensional, with no spectral dimension, this spectral plot would not make much sense. Instead, *Duchamp* creates a similar postscript file that simply includes the text headers as well as the 0th-moment map of the detection. As for the normal spectral case, this file will be written to the filename given by the `spectralFile` parameter.

When the input image is one-dimensional, the spectral plot is identical save for the absence of the cutout image.

In addition to the spectral plot, it is possible to produce plots for each spectrum individually. Set `flagPlotIndividualSpectra=true`, and a postscript plot will be produced for each object. If the normal spectral output file (deter-

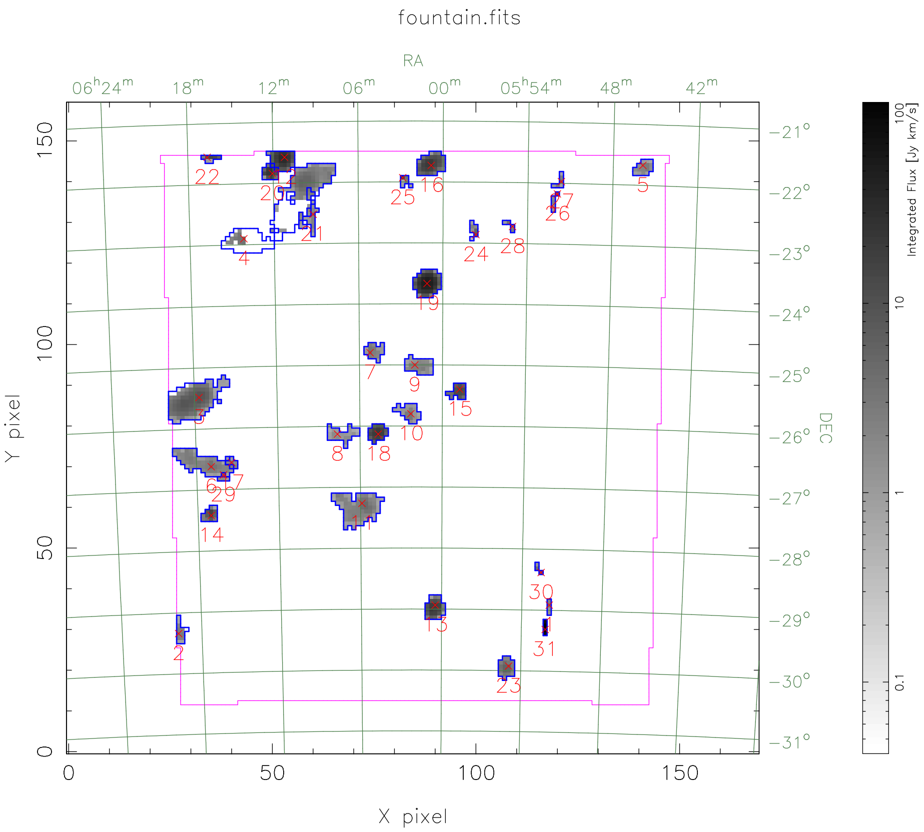


Figure 3: An example of the moment map created by *Duchamp*. The full extent of the cube is covered, and the 0th moment of each object is shown (integrated individually over all the detected channels). The purple line indicates the limit of the non-BLANK pixels.

mined by the `spectralFile` input parameter) is called `duchamp-Spectra.ps`, then the individual files will be called `duchamp-Spectra-01.ps` etc.

5.3.2 Spatial maps

Additionally, two types of spatial images are optionally produced: a combined 0th-moment map of the cube, combining just the detected channels in each object, showing the integrated flux in grey-scale; and a "detection image", a grey-scale image where the pixel values are the number of channels in which that spatial pixel is detected. These detections include pixels that are subsequently discarded (due to the minimum- or maximum-size criteria). In both cases, if `drawBorders = true`, a border is drawn around the spatial extent of each detection, and if `drawBlankEdges = true`, the purple line dividing BLANK and non-BLANK pixels (as described above) is drawn. An example moment map is shown in Fig. 3. The production or otherwise of these images is governed by the `flagMaps` parameter.

The moment map is also displayed in a PGPlot XWindow (with the `/xs` display option). This feature can be turned off by setting `flagXOutput = false`

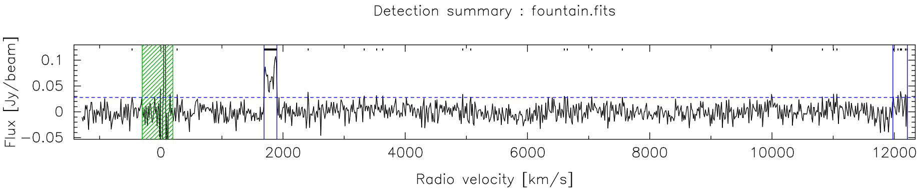


Figure 4: An example of the one-dimensional detection spectrum plot, indicating detected sources and detected pixels, including those subsquently discarded due to the minimum-size criteria. The detection threshold is low to show the effect of detecting lots of single-pixel channels, which are then discarded, leaving just the two detections delimited by the blue lines.

– this might be useful if running *Duchamp* on a terminal with no window display capability, or if you have set up a script to run it in a batch mode.

If the input image is one-dimensional, such a spatial map is not possible. Instead, the detection map becomes a detection spectrum. This shows the full spectral range, indicating (as for the spectral plots above) the detection and growth thresholds, as well as the flagged channels and every detection that appears in the final catalogue. It also indicates all pixels that were detected, including those subsequently discarded, by thick black lines above the spectrum. An example can be see in Fig. 4. Again, this plot is also displayed in a PGPlot XWindow.

The purpose of these images is to provide a visual guide to where the detections have been made, and, particularly in the case of the moment map, to provide an indication of the strength of the source. In both cases, the detections are numbered (in the same sense as the output list and as the spectral plots), and the spatial borders are marked out as for the cutout images in the spectra file. Both these images are saved as postscript files (given by the parameters `momentMap` and `detectionMap` respectively), with the latter also displayed in a PGPLOT window (regardless of the state of `flagMaps`).

5.4 FITS output

5.4.1 Moment map

The moment map described above can also be written to a FITS file, so that it can be examined more closely, and have annotation files overlaid. This works in the same way as for the mask image. To create the FITS file, set the input parameter `flagOutputMomentMap=true`. The file will be named according to the `fileOutputMomentMap` parameter, or, if this is not given, `image.MOM0.fits` (where the input image is called `image.fits`).

5.4.2 Mask images

It is also possible to write the mask array to a FITS file, for use in other forms of post-processing. This array is designed to indicate the location of detected objects. The value of the detected pixels is determined by the input parameter `flagMaskWithObjectNum`: if `true`, the value of the pixels is given by the corresponding object ID number; if `false`, they take the value 1 for all objects. Pixels not in a detected object have the value 0. To create this FITS file, set the input parameter `flagOutputMask=true`. The file will be named according to the `fileOutputMask` parameter, or, if this is not given, `image.MASK.fits` (where the input image is called `image.fits`).

A spatial mask, or moment-0 mask, can also be written. This is simply a two-dimensional image that shows which spatial pixels are detected in one or more channels. Unlike the full mask file above, detected pixels can only be recorded as 1 (as a given spatial pixel may appear in multiple objects) – that is, the parameter `flagMaskWithObjectNum` does not affect the moment-0 mask. To create this FITS file, set the input parameter `flagOutputMomentMask=true`. The file will be named according to the `fileOutputMomentMask` parameter, or, if this is not given, `image.MOMOMASK.fits` (where the input image is called `image.fits`).

5.4.3 Smoothed or Reconstructed image

As discussed in §3.6, the reconstructed array, its residual, or the smoothed array can be saved to a FITS file. This allows examination of them offline, as well as their re-use by *Duchamp* to save the expense of re-calculating. This behaviour is controlled by `flagOutputRecon`, `flagOutputResid` and `flagOutputSmooth`. Consult §3.6 for further details.

5.4.4 Baseline image

As mentioned in §3.3.3, the spectral baseline values can be saved to a FITS file, allowing examination of them offline. There is no scope at present for reloading previously-calculated baselines (although the overheads in calculating these are not too prohibitive). Saving to a FITS file is controlled by the input parameters `flagOutputBaseline` and `fileOutputBaseline`. If `fileOutputBaseline` is not provided, the file will be named `image.BASE.fits` (for an input image called `image.fits`).

5.5 Re-examining previous *Duchamp* results

5.5.1 Binary Catalogue

It is often the case that the bulk of the work in a *Duchamp* run is in the searching for sources. If you are interested in re-doing some of the spectral plots, or re-parameterising with different `spectralType` settings, then having to re-run the searching can be a bit off-putting.

A solution to this problem exists in the ability to save a binary catalogue, containing the information on the individual pixels detected in each object. This is sufficient to recreate a set of detections and re-do the parameterisation. To enable this mode, set `flagWriteBinaryCatalogue=true`, and provide a filename with `binaryCatalogue` (or use the default of `duchamp-Catalogue.dpc`). The following will be written to the catalogue:

- Version of *Duchamp*. If it is not the same version, a warning is raised.
- Current date and time.

- The parameter set. Only the parameters affecting the pre-processing and searching are stored. Those related to, say, graphical output are not.
- The measured statistics.
- The pixels of each detected object, written using the run-length encoding described in §3.7.1.

These are written in binary format to conserve disk space, and are sufficient to recreate the state of *Duchamp* after the searching has taken place.

To re-use this catalogue, set the flag `usePrevious=true` and provide the binary catalogue filename via `binaryCatalogue`. The catalogue will be loaded, and (provided it loads correctly) the preprocessing and searching steps will be skipped. The post-processing (i.e. plotting and catalogue output) steps will occur as normal, using the settings provided in the input parameter file.

Note that while at this stage this is the only use for the binary catalogues, it is anticipated that other functionality will be provided in future - for instance, to allow conversion into mask images. The binary catalogues are seen as a compact way of storing the results of a *Duchamp* run.

5.5.2 Selection of objects

When re-running *Duchamp* on a previously-generated catalogue, it is possible to produce the plots for only a selection of objects. Use the `objectList` parameter to specify a set of objects, listing individual object numbers or ranges, for example "1,3-6,9,11" means objects 1,3,4,5,6,9,11. The output plots will be appropriately modified: the spectral plots will only show these objects; the moment map plot will only show the contribution from these objects; the detection map will show the outlines of only these objects, although all detected pixels are still shown in greyscale.

Note that the object numbers here are valid for the catalogue as sorted according to the `sortingParam` specification in the parameter file. If you change this, the order of the catalogue may change and the specific objects selected by `objectList` will differ.

This option is designed for the case of re-using a catalogue, but can be used for a blind search as well. Of course, you may not know what numbers the sources will turn out to be.

6 Notes and hints on the use of *Duchamp*

In using *Duchamp*, the user has to make a number of decisions about the way the program runs. This section is designed to give the user some idea about what to choose.

6.1 Memory usage

A lot of attention has been paid to the memory usage in *Duchamp*, recognising that data cubes are going to be increasing in size with new generation correlators and wider fields of view. However, users with large cubes should be aware of the likely usage for different modes of operation and plan their *Duchamp* execution carefully.

At the start of the program, memory is allocated sufficient for:

- The entire pixel array (as requested, subject to any subsection).
- The spatial extent, which holds the map of detected pixels (for output into the detection map).
- If smoothing or reconstruction has been selected, another array of the same size as the pixel array. This will hold the smoothed/reconstructed array (the original needs to be kept to do the correct parameterisation of detected sources).
- If baseline-subtraction has been selected, a further array of the same size as the pixel array. This holds the baseline values, which need to be added back in prior to parameterisation.

All of these will be float type, except for the detection map, which is short.

There will, of course, be additional allocation during the course of the program. The detection list will progressively grow, with each detection having a memory footprint as described in §3.7.1. But perhaps more important and with a larger impact will be the temporary space allocated for various algorithms.

The largest of these will be the wavelet reconstruction. This will require an additional allocation of twice the size of the array being reconstructed, one for the coefficients and one for the wavelets - each scale will overwrite the previous one. So, for the 1D case, this means an additional allocation of twice the spectral dimension (since we only reconstruct one spectrum at a time), but the 3D case will require an additional allocation of twice the cube size (this means there needs to be available at least four times the size of the input cube for 3D reconstruction, plus the additional overheads of detections and so forth).

The smoothing has less of an impact, since it only operates on the lower dimensions, but it will make an additional allocation of twice the relevant size (spectral dimension for spectral smoothing, or spatial image size for the spatial Gaussian smoothing).

The other large allocation of temporary space will be for calculating robust statistics. The median-based calculations require at least partial sorting of the data, and so cannot be done on the original image cube. This is done for the entire cube and so the temporary memory increase can be large.

6.2 Timing considerations

Another intersting question from a user's perspective is how long you can expect *Duchamp* to take. This is a difficult question to answer in general, as different users will have different sized data sets, as well as machines with different capabilities (in terms of the CPU speed and I/O & memory bandwidths). Additionally, the time required will depend slightly on the number of sources found and their size (very large sources can take a while to fully parameterise).

Having said that, in Whiting (2012) a brief analysis was done looking at different modes of execution applied to a single HIPASS cube (#201) using a MacBook Pro (2.66GHz, 8MB RAM). Two sets of thresholds were used, either 10^8 Jy beam^{-1} (no sources will be found, so that the time taken is dominated by preprocessing), or 35 mJy beam^{-1} (or $\sim 2.58\sigma$, chosen so that the time taken will include that required to process sources). The basic searches, with no pre-processing done, took less than a second for the high-threshold search, but between 1 and 3 min for the low-threshold case – the numbers of sources detected ranged from 3000 (rejecting sources with less than 3 channels and 2 spatial pixels) to 42000 (keeping all sources).

When smoothing, the raw time for the spectral smoothing was only a few seconds, with a small dependence on the width of the smoothing filter. And because the number of spurious sources is markedly decreased (the final catalogues ranged from 17 to 174 sources, depending on the width of the smoothing), searching with the low threshold did not add much more than a second to the time. The spatial smoothing was more computationally intensive, taking about 4 minutes to complete the high-threshold search.

The wavelet reconstruction time primarily depended on the dimensionality of the reconstruction, with the 1D taking 20 s, the 2D taking 30 - 40 s and the 3D taking 2 - 4 min. The spread in times for a given dimensionality was caused by different reconstruction thresholds, with lower thresholds taking longer (since more pixels are above the threshold and so need to be added to the final spectrum). In all cases the reconstruction time dominated the total time for the low-threshold search, since the number of sources found was again smaller than the basic searches.

6.3 Why do preprocessing?

The preprocessing options provided by *Duchamp*, particularly the ability to smooth or reconstruct via multi-resolution wavelet decomposition, provide an opportunity to beat the effects of the random noise that will be present in the data. This noise will ultimately limit ones ability to detect objects and form a complete and reliable catalogue. Two effects are important here. First, the noise reduces the completeness of the final catalogue by suppressing the flux of real sources such that they fall below the detection threshold. Secondly, the noise provides false positive detections through noise peaks that fall above the threshold, thereby reducing the reliability of the catalogue.

Whiting (2012) examined the effect on completeness and reliability for the reconstruction and smoothing (1D cases only) when applied to a simple simulated dataset. Both had the effect of reducing the number of spurious sources, which means the searches can be done to fainter thresholds. This led to completeness levels of about one flux unit (equal to one standard-deviation of the noise) fainter than searches without pre-processing, with $> 95\%$ reliability. The smoothing did slightly better, with the completeness level nearly half a flux unit fainter than the reconstruction, although this was helped by the sources in the simulation all having the same spectral size.

6.4 Reconstruction considerations

The *à trous* wavelet reconstruction approach is designed to remove a large amount of random noise while preserving as much structure as possible on the full range of spatial and/or spectral scales present in the data. While it is relatively more expensive in terms of memory and CPU usage (see previous sections), its effect on, in particular, the reliability of the final catalogue makes it worth investigating.

There are, however, a number of subtleties to it that need to be considered by potential users. Whiting (2012) shows a set of examples of reconstruction applied to simulated and real data. The real data, in this case a HIPASS cube, shows differences in the quality of the reconstructed spectrum depending on the dimensionality of the reconstruction. The two-dimensional reconstruction (where the cube is reconstructed one channel map at a time) shows much larger channel-to-channel noise, with a number of narrow peaks surviving the reconstruction process. The problem here is that there are spatial correlations between pixels due to the beam, which allow beam-sized noise fluctuations to rise above the threshold more frequently in one-dimension. The other effect is that when compared to a spectrum from the 1D reconstruction, each channel is independently reconstructed, and does not depend on its neighbouring channels. This is also why the 3D reconstruction (which also suffers from the beam effects) has improved noise in the output spectrum, since the information on neighbouring channels is taken into account.

Caution is also advised when looking at subsections of a cube. Due to the multi-scale nature of the algorithm, the wavelet coefficients at a given pixel are influenced by pixels at very large separations, particularly given that edges are dealt with by assuming reflection (so the whole array is visible to all pixels). Also, if one decreases the dimensions of the array being reconstructed, there may be fewer scales used in the reconstruction. These points mean that the reconstruction of a subsection of a cube will differ from the same subsection of the reconstructed cube. The difference may be small (depending on the relative size difference and the amount of structure at large scales), but there will be differences at some level.

Note also that BLANK pixels have no effect on the reconstruction: they remain as BLANK in the output, and do not contribute to the discrete convolution when they otherwise would. Flagging channels with the `flaggedChannels` parameter, however, has no effect on the reconstruction – this are applied after the preprocessing, either in the searching or the rejection stage.

6.5 Smoothing considerations

The smoothing approach differs from the wavelet reconstruction in that it has a single scale associated with it. The user has two choices to make - which dimension to smooth in (spatially or spectrally), and what size kernel to smooth with. Whiting (2012) show examples of how different smoothing widths (in one-dimension in this case) can highlight sources of different sizes. If one has some *a priori* idea of the typical size scale of objects one wishes to detect, then choosing a single smoothing scale can be quite beneficial.

Note also that beam effects can be important here too, when smoothing spatial data on scales close to that of the beam. This can enhance beam-sized noise fluctuations and potentially introduce spurious sources. As always, examining the smoothed array (after saving via `flagOutputSmooth`) is a good idea.

6.6 Threshold method

When it comes to searching, the FDR method produces more reliable results than simple sigma-clipping, particularly in the absence of reconstruction. However, it does not work in exactly the way one would expect for a given value of `alpha`. For instance, setting fairly liberal values of `alpha` (say, 0.1) will often lead to a much smaller fraction of false detections (i.e. much less than 10%). This is the effect of the merging algorithms, that combine the sources after the detection stage, and reject detections not meeting the minimum pixel or channel requirements. It is thus better to aim for larger `alpha` values than those derived from a straight conversion of the desired false detection rate.

If the FDR method is not used, caution is required when choosing the S/N cutoff. Typical cubes have very large numbers of pixels, so even an apparently large cutoff will still result in a not-insignificant number of detections simply due to random fluctuations of the noise background. For instance, a 4σ threshold on a cube of Gaussian noise of size $100 \times 100 \times 1024$ will result in ~ 340 single-pixel detections. This is where the minimum channel and pixel requirements are important in rejecting spurious detections.

7 Future developments

Here are lists of planned improvements and a wish-list of features that would be nice to include (but are not planned in the immediate future). Let me know if there are items not on these lists, or items on the list you would like prioritised.

Planned developments:

- Parallelisation of the code, to improve speed particularly on multi-core machines.
- Better determination of the noise characteristics of spectral-line cubes, including understanding how the noise is generated and developing a model for it.
- Include more source analysis. Examples could be: shape information; measurements of HI mass; more variety of measurements of velocity width and profile.
- Improved ability to reject interference, possibly on the spectral shape of features.
- Ability to separate (de-blend) distinct sources that have been merged.

Wish-list:

- Incorporation of Swinburne's S2PLOT [10] code for improved visualisation.
- Link to lists of possible counterparts (e.g. via NED/SIMBAD/other VO tools?).
- On-line web service interface, so a user can upload a cube and get back a source-list.
- Embed *Duchamp* in a GUI, to move away from the text-based interaction.

[10] http://astronomy.swin.edu.au/s2plot/

8 Acknowledgements

Thanks are due to the many people who have provided assistance and advice during the development and testing of *Duchamp*, particularly Ivy Wong, Kathrin Wolfinger, Tobias Westmeier, Sara Shakouri, Mary Putman, Cormac Purcell, Attila Popping, Tara Murphy, Enno Middelberg, Korinne McDonnell, Malte Marquarding, Philip Lah, Russell Jurek, Simon Guest, Jose Francisco Gomez, BiQing For, Luca Cortese, Mark Calabretta, David Barnes and Robbie Auld. Additionally, Emil Lenc and Anita Richards both provided valuable comments on the journal paper that have helped the descriptions therein. I'd like to thank the members of the ASKAP Working Group 2 (Source Finding) for their interest and feedback - *Duchamp* has been considered as part of the development of ASKAP Survey Science Projects, and this has driven further development of the core *Duchamp* software.

The graphics in *Duchamp* are created using the PGPLOT[11] library, FITS file access is controlled through the CFITSIO[12] library (Pence 1999), while the world-coordinate transformations are performed using the WCSLIB[13] library (described in Calabretta and Greisen (2002)).

The bulk of this work was conducted as part of a CSIRO Emerging Science Postdoctoral Fellowship, and *Duchamp* continues to be maintained both as a standalone package and as part of the software development effort for the ASKAP telescope. This work was supported by the NCI National Facility at the ANU.

[11] http://www.astro.caltech.edu/∼tjp/pgplot/

[12] http://heasarc.gsfc.nasa.gov/docs/software/fitsio/fitsio.html

[13] http://www.atnf.csiro.au/people/mcalabre/WCS/index.html

References

M.R. Calabretta and E.W. Greisen. “Representations of celestial coordinates in FITS”. *A&A*, 395:1077–1122, December 2002. .

E.W. Greisen and M.R. Calabretta. “Representations of world coordinates in FITS”. *A&A*, 395:1061–1075, December 2002.

E.W. Greisen, M.R. Calabretta, F.G. Valdes, and S.L. Allen. Representations of spectral coordinates in FITS. *A&A*, 446:747–771, 2006.

R.J. Hanisch, A. Farris, E.W. Greisen, W.D. Pence, B.M. Schlesinger, P.J. Teuben, R.W. Thompson, and A. Warnock. “Definition of the Flexible Image Transport System (FITS)”. *A&A*, 376:359–380, September 2001.

A.M. Hopkins, C.J. Miller, A.J. Connolly, C. Genovese, R.C. Nichol, and L. Wasserman. “A New Source Detection Algorithm Using the False-Discovery Rate”. *AJ*, 123:1086–1094, February 2002.

R.K. Lutz. “An algorithm for the real time analysis of digitised images”. *The Computer Journal*, 23:262–269, 1980.

M.J. Meyer et al. “The HIPASS catalogue - I. Data presentation”. *MNRAS*, 350:1195–1209, June 2004.

C.J. Miller, C. Genovese, R.C. Nichol, L. Wasserman, A. Connolly, D. Reichart, A. Hopkins, J. Schneider, and A. Moore. “Controlling the False-Discovery Rate in Astrophysical Data Analysis”. *AJ*, 122:3492–3505, December 2001.

R.F. Minchin. “Finding the Bivariate Brightness Distribution of Galaxies from an HI Selected Sample”. *PASA*, 16:12–17, 1999.

W. Pence. CFITSIO, v2.0: A New Full-Featured Data Interface. In D. M. Mehringer, R. L. Plante, & D. A. Roberts, editor, *Astronomical Data Analysis Software and Systems VIII*, volume 172 of *Astronomical Society of the Pacific Conference Series*, page 487, 1999.

J.-L. Starck and F. Murtagh. *“Astronomical Image and Data Analysis”*. Springer, 2002.

M. T. Whiting. DUCHAMP: a 3D source finder for spectral-line data. *MNRAS*, 421:3242–3256, 2012.

A Obtaining and installing *Duchamp*

A.1 Installing

The *Duchamp* web page can be found at the following location:
http://www.atnf.csiro.au/people/Matthew.Whiting/Duchamp
Here you can find a gzipped tar archive of the source code that can be downloaded and extracted, as well as this User's Guide in postscript and hyperlinked PDF formats.

To build *Duchamp*, you will need three main external libraries: PGPLOT, CFITSIO (this needs to be version 2.5 or greater – version 3+ is better) and WCSLIB. If these are not present on your system, you can download them from the following locations:

- PGPLOT: http://www.astro.caltech.edu/ tjp/pgplot/
- CFITSIO: http://heasarc.gsfc.nasa.gov/docs/software/fitsio/fitsio.html
- WCSLIB: http://www.atnf.csiro.au/people/Mark.Calabretta/WCS/index.html

A.1.1 Basic installation

Duchamp can be built on Unix/Linux systems by typing (assuming that the prompt your terminal provides is a `>` – don't type this character!):

```
> ./configure
> make
> make lib (optional -- to create libraries for development purposes)
> make clean (optional -- to remove the object files)
> make install
```

This default setup will search in standard locations for the necessary libraries, and install the executable (`Duchamp-1.6.1`) in `/usr/local/bin`, along with a `Duchamp` symbolic link (a copy will also be in the current directory). The full set of header files will be installed in `/usr/local/include/duchamp` and subdirectories thereof.

If you have made the libraries, both static (`libduchamp.1.6.1.a`) and shared (`libduchamp.1.6.1.so` or `libduchamp.1.6.1.dylib` depending on your system) libraries will be created and installed in `/usr/local/lib`. Symbolic links will also be created that don't have the version number.

If you want these to go somewhere else, e.g. if you don't have write-access to that directory, or you need to tweak the location of the libraries, see the next section. Otherwise, jump to the testing section.

A.1.2 Tweaking the installation process

The `configure` script allows the user to tailor the installation according to the particular requirements of their system.

To install *Duchamp* in a directory other than `/usr/local/bin`, use the `--prefix` option with configure, specifying the directory above the `bin/` directory e.g.

```
> ./configure --prefix=/home/mduchamp
```

and then run `make`, (`make lib` if you like), and `make install` as stated above. This will put the binary in the directory `/home/mduchamp/bin`. The library, if made, will be put in `/home/mduchamp/lib` and the header files will be put in `/home/mduchamp/include/duchamp` and subdirectories.

If the above-mentioned libraries have been installed in non-standard locations, or you have more than one version installed on your system, you can specify specific locations by using the `configure` flags `--with-cfitsio=<dir>`, `--with-wcslib=<dir>` or `--with-pgplot=<dir>`. For example:

```
> ./configure --with-wcslib=/home/mduchamp/wcslib-4.2
```

Duchamp can be compiled without PGPLOT if it is not installed on your system – the searching and text-based output remains the same, but you will not have any graphical output. To manually specify this option, you can either give `--without-pgplot` or `--with-pgplot=no` as arguments to `configure`:

```
> ./configure --without-pgplot
```

(Note that CFITSIO and WCSLIB are essential, however, so flags such as `--without-wcslib` or `--without-cfitsio` will not work.). Even if you do not give the `--without-pgplot` option, and the PGPLOT library is not found, *Duchamp* will still compile (albeit without graphical capabilities).

An additional option that is useful is the ability to specify which compiler to use. This is very important for the Fortran compiler (used for linking due to the use of PGPLOT), particularly on Mac OS X, where `gfortran` is often used instead of `gcc`. To specify a particular Fortran compiler, use the `F77` flag:

```
> ./configure F77=gfortran
```

Of course, all desired flags should be combined in one `configure` call. For a full list of the options with `configure`, run:

```
> ./configure --help
```

Once `configure` has run correctly, simply run `make` and `make install` to build *Duchamp* and put it in the correct place (either `/usr/local/bin` or the location given by the `--prefix` option discussed above).

A.1.3 Making sure it all works

Running make will create the executable `Duchamp-1.6.1`. You can verify that it is running correctly by running the verification shell script:

```
> ./VerifyDuchamp.sh
```

This will use a dummy FITS image in the `verification/` directory – this image has some Gaussian random noise, with five Gaussian sources present, plus a dummy WCS. The script runs *Duchamp* on this image with nine different sets of inputs, and compares to known results, looking for differences and reporting any. There should be none reported if everything is working correctly.

The script performs basic checks on the output files (results, log, VOTable, and annotation files), but ignores most of the actual values of source parameters (to avoid picking up just differences due to precision errors). For complete checks of the files, run

```
> ./VerifyDuchamp.sh -f
```

Be warned that on some systems this could provide a large number of apparent errors which may only be due to precision differences.

If everything worked, you can then install *Duchamp* on your system via:

```
> make install
```

(this may need to be run as sudo depending on your system setup and your prefix directory).

A.2 Troubleshooting the installation process

This section deals with a few common problems encountered in building *Duchamp*, along with suggested fixes. As always, if you come across particularly intractable problems, you are welcome to submit a bug report – see below for details.

A.2.1 Unrecognised libraries

It may be that even after explicitly giving the location of particular libraries, they are still not being found properly by `configure`. This can be a particular problem when those libraries are installed in a non-standard location (for instance, if you do not have root permissions on the machine you are using and have installed them yourself in a local directory).

One thing to be aware of is that the paths used for linking libraries include your new library. You can set this using the environment variables `LD_LIBRARY_PATH` or, on a Mac, `DYLD_LIBRARY_PATH`. For instance, if you've installed `wcslib` in `/home/mduchamp/mylibs/wcslib` then you can update the path via

```
> export LD_LIBRARY_PATH="${LD_LIBRARY_PATH}:/home/mduchamp/mylibs/wcslib"
```

for a `bash`-like shell, or

```
> setenv LD_LIBRARY_PATH "${LD_LIBRARY_PATH}:/home/mduchamp/mylibs/wcslib"
```

for a `csh`-like shell

and similarly for `DYLD_LIBRARY_PATH`.

A.2.2 Non-standard system library locations

It may be that one of the system libraries is not being found correctly. This can be a problem with the `gfortran` libary. If `configure` cannot find it, then it will likely leave out one or more libraries from the linking command (such as `-lcpgplot` or `-lpgsbox`), resulting in *Duchamp* not building correctly. You may also get error messages at the linking stage (the final stage of running `make`) that complain of missing symbols.

To force `configure` to use a non-standard location, you can use the `LIBS` option. For instance, say your system has `libgfortran` in some non-standard location like `/some/other/path/libgfortran.so`, then you can run `configure` like so:

```
> ./configure LIBS="-L/some/other/path -lgfortran"
```

This will allow `configure` to test for other libraries using that location for `libgfortran`, and put that library location into the Makefile.

Another potential problem are the X11 libraries. These are used for the `pgplot` display during run-time. The `configure` script has a specific function that searches for them, but sometimes it fails to locate them if they are in a non-standard place on your system. This can be a problem as `configure` will then fail to test properly for the presence of `pgplot`. If this is the case, use the `--x-includes` and `--x-libraries` flags to give the relevant directories.

A.2.3 Bad command-line options

While the configure script tries to get everything right, it can exhibit some quirks. For instance, in getting the X11 library configuration right, it will sometimes provide a `-R/path/to/X11` argument for the linking string. This is accepted by `ld`, but not by all versions of `gfortran`. This can cause the final linking step to fail (and for the `-lpgplot` argument to be left off).

To fix this, a shell script is provided to quickly patch the Makefile if necessary. If you run make and it fails, due to this error:

```
gfortran: error: unrecognized command line option -R
```

or similar, run

```
> ./fixMakefile.sh
```

and try again. If it still fails, you may have to manually edit the Makefile. Please log a bug report to let me know!

A.3 Running *Duchamp*

You can now run *Duchamp* on your own data. This can be done in one of two ways. The first is:

```
> Duchamp -f [FITS file]
```

where `[FITS file]` is the file you wish to search. This method simply uses the default values of all parameters. The flux threshold can be specified using the `-t [THRESHOLD]` option:

```
> Duchamp -f [FITS file] -t [THRESHOLD]
```

The second method allows some determination of the parameter values by the user. Type:

```
> Duchamp -p [parameter file]
```

where `[parameterFile]` is a file with the input parameters, including the name of the cube you want to search. The `-t` flag can also be specified - its threshold value will override anything given in the parameter file.

There are two example input files included with the distribution. The smaller one, `InputExample`, shows the typical parameters one might want to set. The large one, `InputComplete`, lists all possible parameters that can be entered, along with their default values, and a brief description of them. To get going quickly, just replace the `"your-file-here"` in the `InputExample` file with your image name, and type

```
> Duchamp -p InputExample
```

To disable the use of X-window plotting (in displaying the map of detections), one can either set the parameter `flagXOutput = false` or use the `-x` command-line option:

```
> Duchamp -x -p [parameter file] , or
> Duchamp -x -f [FITS file]
```

Note that the postscript outputs will still be produced (if required) – this just affects the runtime display.

The following appendices provide details on the individual parameters, and show examples of the output files that *Duchamp* produces.

A.4 Feedback

It may happen that you discover bugs or problems with *Duchamp*, or you have suggestions for improvements or additional features to be included in future releases. You can submit a "ticket" (a trackable bug report) at the *Duchamp* Trac wiki at the following location:
http://svn.atnf.csiro.au/trac/duchamp/newticket
(there is a link to this page from the *Duchamp* website).

There is also an email exploder, duchamp-user**[at]**atnf.csiro.au, that users can subscribe to keep up to date with changes, updates, and other news about *Duchamp*. To subscribe, send an email (from the account you wish to subscribe to the list) to duchamp-user-request**[at]**atnf.csiro.au with the single word "subscribe" in the body of the message. To be removed from this list, send a message with "unsubscribe" in its body to the same address.

A.5 Beta Versions

On the *Duchamp* website there may be a beta version listed in the downloads section. As *Duchamp* is still under development, there will be times when there has been new functionality added to the code, but the time has not yet come to release a new minor (or indeed major) version.

Sometimes I will post the updated version of the code on the website as a "beta" version, particularly if I'm interested in people testing it. It will not have been tested as rigorously as the proper releases, but it will certainly work in the basic cases that I use to test it during development. So feel free to give it a try – the `CHANGES` file will usually detail what is different to the last numbered release.

B Available parameters

The full list of parameters that can be listed in the input file are given here. Following each parameter name are listed three things: the default parameter taken when not provided in the input file; the data type that should be provided; and some indication (where applicable) of the range of values required or the type of information expected. Since the order of the parameters in the input file does not matter, they are grouped here in logical sections.

Input related

`ImageFile [no default | string | filename]`:
The filename of the data cube to be analysed.

`flagSubsection [false | bool | true/false/1/0]`:
A flag to indicate whether one wants a subsection of the requested image.

`Subsection [full field | string | Subsection string]`:
The requested subsection – see §3.1 for details on the subsection format. If the full range of a dimension is required, use a `*` (thus the default is the full cube).

`flagReconExists [false | bool | true/false/1/0]`:
A flag to indicate whether the reconstructed array has been saved by a previous run of *Duchamp*. If set true, the reconstructed array will be read from the file given by `reconFile`, rather than calculated directly.

`reconFile [no default | string | filename]`:
The FITS file that contains the reconstructed array. If `flagReconExists` is true and this parameter is not defined, the default file that is looked for will be determined by the *à trous* parameters (see §3.4).

`flagSmoothExists [false | bool | true/false/1/0]`:
A flag to indicate whether the Hanning-smoothed array has been saved by a previous run of *Duchamp*. If set true, the smoothed array will be read from the file given by `smoothFile`, rather than calculated directly.

`smoothFile [no default | string | filename]`:
The FITS file that has a previously smoothed array. If `flagSmoothExists` is true and this parameter is not defined, the default file that is looked for will be determined by the smoothing parameters (see §3.5).

`usePrevious [false | bool | true/false/1/0]`:
A flag to indicate that *Duchamp* should read the list of objects from a previously-created log file, rather than doing the searching itself. The set of outputs will be created according to the flags in the following section.

`objectList [no default | string | comma-separated list]`:
When `usePrevious=true`, this list is used to output individual spectral plots, as well as a postscript file for all

spectral plots as given by `SpectraFile`. The filenames of the plots will be the same as `SpectraFile`, but with -XX at the end, where XX is the object number (e.g. `duchamp-Spectra-07.ps`). The format of the parameter value should be a string listing individual objects or object ranges: e.g. 1,2,4-7,8,14.

Output related

`OutFile [duchamp-Results.txt | string | filename]:`
The file containing the final list of detections. This also records the list of input parameters.

`flagSeparateHeader [false | bool | true/false/1/0]:`
A flag to indicate that the header information that would normally be printed at the start of the results file (containing information on the parameters, image statistics and number of detections) should instead be written to a separate file.

`HeaderFile [duchamp-Results.hdr | string | filename]:`
The file to which the header information should be written when `flagSeparateHeader=true`.

`flagWriteBinaryCatalogue [true | bool | true/false/1/0]:`
Whether to write a binary catalogue of the detections, for later re-use (see §5.5 for details).

`binaryCatalogue [duchamp-Catalogue.dpc | string | filename]:`
The filename for the binary catalogue.

`flagPlotSpectra [true | bool | true/false/1/0]:`
Whether to produce a postscript file containing spectra of all detected objects. If PGPlot has not been enabled, this parameter defaults to `false`.

`SpectraFile [duchamp-Spectra.ps | string | filename]:`
The postscript file that contains the resulting integrated spectra and images of the detections.

`flagPlotIndividualSpectra [false | bool | true/false/1/0]:`
Whether to produce individual spectral plots for listed sources.

`flagTextSpectra [false | bool | true/false/1/0]:`
A flag to say whether the spectra should be saved in text form in a single file. See below for a description.

`spectraTextFile [duchamp-Spectra.txt | string | filename]:`
The file containing the spectra of each object in ascii format. This file will have a column showing the spectral coordinates, and one column for each of the detected objects, showing the flux values as plotted in the graphical output of `spectraFile`.

`flagLog [false | bool | true/false/1/0]:`
A flag to indicate whether the details of intermediate detections should be logged.

`LogFile [duchamp-Logfile.txt | string | filename]:`
The file in which intermediate detections and the pixel content of the final list of detections are logged. These are detections that have not been merged. This is primarily for use in debugging and diagnostic purposes: normal use of the program will probably not require it.

`flagOutputMomentMap [false | bool | true/false/1/0]:`
A flag to say whether or not to save a FITS file containing the moment-0 map.

`fileOutputMomentMap [see text | string | filename]:`
The file to which the moment-0 array is written. If left blank (the default), the naming scheme detailed in §5.4.1 is used.

`flagOutputMomentMask [false | bool | true/false/1/0]:`
A flag to say whether or not to save a FITS file containing the moment-0 mask (a mask showing which spatial pixels are detected in one or more channels).

`fileOutputMomentMask [see text | string | filename]:`
The file to which the moment-0 mask is written. If left blank (the default), the naming scheme detailed in §5.4.2 is used.

`flagOutputMask [false | bool | true/false/1/0]:`
A flag to say whether or not to save a FITS file containing a mask array, with values 1 where there is a detected object and 0 elsewhere.

`fileOutputMask [see text | string | filename]:`
The file to which the mask array is written. If left blank (the default), the naming scheme detailed in §5.4.2 is used.

`flagMaskWithObjectNum [false | bool | true/false/1/0]:`
If this flag is true, the detected pixels in the mask image have the corresponding object ID as their value. If false, they have the value 1. All non-detected pixels have the value 0.

`flagOutputRecon [false | bool | true/false/1/0]:`
A flag to say whether or not to save the reconstructed cube as a FITS file.

`fileOutputRecon [see text | string | filename]:`
The file to which the reconstructed array is written. If left blank (the default), the naming scheme detailed in §3.6 is used.

`flagOutputResid [false | bool | true/false/1/0]:`
As for `flagOutputRecon`, but for the residual array – the difference between the original cube and the reconstructed cube.

`fileOutputResid [see text | string | filename]:`
The file to which the residual array is written. If left blank (the default), the naming scheme detailed in §3.6 is used.

`flagOutputSmooth [false | bool | true/false/1/0]:`
A flag to say whether or not to save the smoothed cube as a FITS file.

`fileOutputSmooth [see text | string | filename]:`
The file to which the smoothed array is written. If left blank (the default), the naming scheme detailed in §3.6 is used.

`flagOutputBaseline [false | bool | true/false/1/0]:`
A flag to say whether or not to save the cube of spectral baseline values as a FITS file.

`fileOutputBaseline [see text | string | filename]:`
The file to which the baseline values are written. If left blank (the default), the naming scheme detailed in §5.4.4 is used.

`flagVOT [false | bool | true/false/1/0]:`
A flag to say whether to create a VOTable file with the detection information. This will be an XML file in the Virtual Observatory VOTable format.

`votFile [duchamp-Results.xml | string | filename]:`
The VOTable file with the list of final detections. Some input parameters are also recorded.

`flagKarma [false | bool | true/false/1/0]:`
A flag to say whether to create a Karma annotation file corresponding to the information in `outfile`. This can be used as an overlay in Karma programs such as `kvis`.

`karmaFile [duchamp-Results.ann | string | filename]:`
The Karma annotation file showing the list of final detections.

`flagDS9 [false | bool | true/false/1/0]:`
A flag to say whether to create a DS9 region file corresponding to the information in `outfile`. This can be used as an overlay in SAOImage DS9 or casaviewer.

`ds9File [duchamp-Results.ann | string | filename]:`
The DS9 region file showing the list of final detections.

`flagCasa [false | bool | true/false/1/0]:`
A flag to say whether to create a CASA region file corresponding to the information in `outfile`. This can be used as an overlay in casaviewer (when this functionality is available) or import into casapy.

`casaFile [duchamp-Results.crf | string | filename]:`
The CASA region file showing the list of final detections.

`annotationType [borders | string | borders/circles/ellipses]:`
Which type of annotation plot to use. Specifying "borders" gives an outline around the detected spatial pixels, "circles" gives a circle centred on the centre of the object with radius large enough to encompass all spatial pixels, and "ellipses" gives an ellipse centred on the centre of the object of size given by the MAJ, MIN & PA values.

`flagMaps [true | bool | true/false/1/0]:`
A flag to say whether to save postscript files showing the 0th moment map of the whole cube (`momentMap`) and the detection image (`detectionMap`). If PGPlot has not been enabled, this parameter defaults to `false`.

`momentMap [duchamp-MomentMap.ps | string | filename]:`
A postscript file containing a map of the 0th moment of the detected sources, as well as pixel and WCS coordinates.

`detectionMap [duchamp-DetectionMap.ps | string | filename]:`
A postscript file with a map showing each of the detected objects, coloured in greyscale by the number of detected channels in each spatial pixel. Also shows pixel and WCS coordinates.

`flagXOutput [true | bool | true/false/1/0]:`
A flag to say whether to display a 0th moment map in a PGPlot X-window. This will be in addition to any that are saved to a file. This parameter can be overridden by the use of the `-x` command-line option, which disables the X-windows output. If PGPlot has not been enabled, this parameter defaults to `false`.

`newFluxUnits [no default | string | units string]:`
Flux units that the pixel values should be converted into. These should be directly compatible with the units in the FITS header, given by the BUNIT keyword.

`precFlux [3 | int |` > 0`]:`
The desired precision (i.e. number of decimal places) for flux values given in the output files and tables.

`precVel [3 | int |` > 0`]:`
The desired precision (i.e. number of decimal places) for velocity/frequency values given in the output files and tables.

`precSNR [2 | int |` > 0`]:`
The desired precision (i.e. number of decimal places) for the peak SNR value given in the output files and tables.

Modifying the cube

`flagTrim [false | bool | true/false/1/0]:`
A flag to say whether to trim BLANK pixels from the edges of the cube – these are typically pixels set to some particular value because they fall outside the imaged area, and trimming them can help speed up the execution.

`flaggedChannels [no default | string | comma-separated list]:`
Channels that are to be ignored by the source-finding. These should be specified by a comma-separated list of single values and ranges, such as 1,3,6-12,18. Channel numbers are zero-based, so that the first channel in the cube has value 0.

`flagBaseline [false | bool | true/false/1/0]`:
A flag to say whether to remove the baseline from each spectrum in the cube for the purposes of reconstruction and detection.

`baselineType [atrous | string | atrous/median]`:
The algorithm used to calculate the spectral baseline. Only `atrous` or `median` are accepted.

`baselineBoxWidth [51 | int | odd integer` > 0`]`:
The box width used by the `median` baseline algorithm. Needs to be odd - if even, it will be incremented by one.

Detection related

General detection

`searchType [spatial | string | spectral/spatial]`:
How the searches are done. Only "spatial" and "spectral" are accepted. A value of "spatial" means each 2D channel map is searched, whereas "spectral" means each 1D spectrum is searched.

`flagStatSec [false | bool | true/false/1/0]`:
A flag indicating whether the statistics should be calculated on a subsection of the cube, rather than the full cube. Note that this only applies to the statistics used to determine the threshold, and not for other statistical calculations (such as those in the reconstruction phase).

`StatSec [full field | string | Subsection string]`:
The subsection of the cube used for calculating statistics – see §3.1 for details on the subsection format. Only used if `flagStatSec=true`.

`flagRobustStats [true | bool | true/false/1/0]`:
A flag indicating whether to use the robust statistics (median and MADFM) to estimate the noise parameters of the cube, rather than the mean and rms. See §3.7.3 for details.

`flagNegative [false | bool | true/false/1/0]`:
A flag indicating that the features of interest are negative. The cube is inverted prior to searching.

`snrCut [5.  | float | any]`:
The threshold, in multiples of σ above the mean.

`threshold [no default | float | any]`:
The actual value of the threshold. Normally the threshold is calculated from the cube's statistics, but the user can manually specify a value to be used that overrides the calculated value. If this is not specified, the calculated value is used, but this value will take precedence over other means of calculating the threshold (i.e. via `snrCut` or the FDR method).

`flagGrowth [false | bool | true/false/1/0]:`
A flag indicating whether or not to grow the detected objects to a smaller threshold.

`growthCut [3.  | float | any]:`
The smaller threshold using in growing detections. In units of σ above the mean.

`growthThreshold [no default | float | any]:`
Alternatively, the threshold to which detections are grown can be specified in flux units (in the same manner as the `threshold` parameter). When the `threshold` parameter is given, this option **must** be used instead of `growthCut`.

`beamFWHM [0.  | float |` $> 0.$`]:`
The full-width at half maximum of the beam, in pixels. If the header keywords BMAJ and BMIN are present, then these will be used to calculate the beam area, and this parameter will be ignored. This will take precedence over `beamArea` (but is ignored if not specified).

`beamArea [0.  | float |` $> 0.$`]:`
The **area** of the beam in pixels (i.e. how many pixel does the beam cover?). As above, if the header keywords BMAJ and BMIN are present, then these will be used to calculate the beam area, and this parameter will be ignored.

À trous reconstruction

`flagATrous [false | bool | true/false/1/0]:`
A flag indicating whether or not to reconstruct the cube using the *à trous* wavelet reconstruction. See §3.4 for details.

`reconDim [1 | int | 1, 2 or 3]:`
The number of dimensions to use in the reconstruction. 1 means reconstruct each spectrum separately, 2 means each channel map is done separately, and 3 means do the whole cube in one go.

`scaleMin [1 | int |` > 0`]:`
The minimum wavelet scale to be used in the reconstruction. A value of 1 means "use all scales".

`scaleMax [0 | int | any]:`
The maximum wavelet scale to be used in the reconstruction. If the value is ≤ 0 then the maximum scale is calculated from the size of the input array. Similarly, if the value given is larger than this calculated value, the calculated value is used instead.

`snrRecon [4 | float |` > 0`]:`
The thresholding cutoff used in the reconstruction – only wavelet coefficients at least this many σ above the mean are included in the reconstruction.

`reconConvergence [0.005 | float |` $> 0.$`]:`
The convergence criterion used in the reconstruction. The

à trous algorithm iterates until the relative change in the standard deviation of the residuals is less than this amount.

`filterCode [1 | int | 1/2/3]`:

The code number of the filter to use in the reconstruction. The options are:

- **1:** B$_3$-spline filter: coefficients = $(\frac{1}{16}, \frac{1}{4}, \frac{3}{8}, \frac{1}{4}, \frac{1}{16})$
- **2:** Triangle filter: coefficients = $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$
- **3:** Haar wavelet: coefficients = $(0, \frac{1}{2}, \frac{1}{2})$

Smoothing the cube

`flagSmooth [false | bool | true/false/1/0]`:

A flag indicating whether to smooth the cube. See §3.5 for details.

`smoothType [spectral | string | spectral/spatial]`:

The smoothing method used: either "spectral" (with a 1D Hanning filter) or "spatial" (with a 2D Gaussian filter).

`hanningWidth [5 | int |` >0`]`:

The width of the Hanning smoothing kernel.

`kernMaj [3 | float |` $>0.$`]`:

The full-width-half-maximum (FWHM), in pixels, of the 2D Gaussian smoothing kernel's major axis.

`kernMin [3 | float |` $>0.$`]`:

The FWHM (in pixels) of the 2D Gaussian smoothing kernel's minor axis.

`kernPA [0 | float | any]`:

The position angle, in degrees, anticlockwise from vertical (i.e. usually East of North).

`smoothEdgeMethod [equal | string | equal/truncate/scale]`:

What method to use for dealing with pixels on the edge of the spatial image (i.e. within the width of the kernel). Can be one of `equal`, `truncate`, `scale`. See §3.5 for details.

`spatialSmoothCutoff [1.e-10 | float |` $>0.$`]`:

The cutoff value for determining the width of the smoothing kernel. See §3.5 for details.

FDR method

`flagFDR [false | bool | true/false/1/0]`:

A flag indicating whether or not to use the False Discovery Rate method in thresholding the pixels.

`alphaFDR [0.01 | float |` $0.-1.$`]`:

The α parameter used in the FDR analysis. The average number of false detections, as a fraction of the total number, will be less than α (see §3.7).

`FDRnumCorChan [2 | int |` >0`]`:

The number of neighbouring spectral channels that are assumed to be correlated. This is needed by the FDR algorithm to calculate the normalisation constant c_N (see §3.7).

Merging detections

`minPix` [`2` | `int` | > 0]:
The minimum number of spatial pixels for a single detection to be counted.

`minChannels` [`3` | `int` | > 0]:
At least one contiguous set of this many channels must be present in the detection for it to be accepted.

`minVoxels` [`minPix`+`minChannels`−`1` | `int` | > 0]:
The minimum size of the object, in terms of the total number of voxels, for it to be accepted. This will be *at least* minPix+minChannels−1, but can be set higher.

`maxPix` [−`1` | `int` | `any`]:
The maximum number of spatial pixels an object can have. No check is made if the value is negative.

`maxChannels` [`-1` | `int` | `any`]:
The maximum number of channels an object can have. No check is made if the value is negative.

`maxVoxels` [−`1` | `int` | `any`]:
The maximum size of the object, in terms of the total number of voxels, for it to be accepted. No check is made if the value is negative.

`flagRejectBeforeMerge` [`false` | `bool` | `true/false/1/0`]:
A flag indicating whether to reject sources that fail to meet the `minPix` or `minChannels` criteria **before** the merging stage. Default behaviour is to do the rejection last.

`flagTwoStageMerging` [`true` | `bool` | `true/false/1/0`]:
A flag indicating whether to do an initial merge of newly-detected sources into the source list as they are found. If `false`, new sources are simply added to the end of the list for later merging.

`flagAdjacent` [`true` | `bool` | `true/false/1/0`]:
A flag indicating whether to use the "adjacent pixel" criterion to decide whether to merge objects. If not, the next two parameters are used to determine whether objects are within the necessary thresholds.

`threshSpatial` [3. | `float` | ≥ 0]:
The maximum allowed minimum spatial separation (in pixels) between two detections for them to be merged into one. Only used if `flagAdjacent = false`.

`threshVelocity` [7. | `float` | ≥ 0]:
The maximum allowed minimum channel separation between two detections for them to be merged into one.

WCS parameters

`spectralType ["" | string | A valid WCS type]:`
The user can specify an alternative WCS spectral type that the spectral axis can be expressed in. This specification should conform to the standards described in Greisen et al. (2006), although it is possible to provide just the first four letters (the "S-type", e.g. 'VELO').

`restFrequency [-1 | float | any]:`
If provided, this will be used in preference to the rest frequency given in the FITS header to calculate velocities and related quantities. A negative value (such as the default) will mean this is not used and the FITS header value, if present, is used instead.

`spectralUnits ["" | string | A valid units string]:`
The user can specify the units of the spectral axis, overriding those given in the FITS header. If the spectral type is being changed, these units should be appropriate for that quantity. If not provided, the FITS header information is used.

Other parameters

`spectralMethod [peak | string | peak/sum]:`
This indicates which method is used to plot the output spectra: `peak` means plot the spectrum containing the detection's peak pixel; `sum` means sum the spectra of each detected spatial pixel, and correct for the beam size. Any other choice defaults to `peak`.

`pixelCentre [centroid | string | centroid/peak/average]:`
Which of the three ways of expressing the "centre" of a detection (see §5.2.1 for a description of the options) to use for the X, Y, & Z columns in the output list. Alternatives are: `centroid, peak, average`.

`sortingParam [vel | string | see text for options]:`
The parameter on which to sort the output list of detected objects. Options are: xvalue, yvalue, zvalue, ra, dec, vel, w50, iflux, pflux (integrated and peak flux respectively), or snr. If the parameter begins with a '-' (e.g. '-vel'), the order of the sort is reversed.

`drawBorders [true | bool | true/false/1/0]:`
A flag indicating whether to draw borders around the detected objects in the moment maps included in the output (see for example Fig. 2).

`drawBlankEdges [true | bool | true/false/1/0]:`
A flag indicating whether to draw the dividing line between BLANK and non-BLANK pixels on the 2D images (see for example Fig. 3).

`verbose [true | bool | true/false/1/0]:`

A flag indicating whether to print the progress of any computationally intensive algorithms (e.g. reconstruction, searching or merging algorithms) to the screen.

C Example parameter files

This is what a typical parameter file would look like.

```
imageFile       /home/mduchamp/fountain.fits
logFile         logfile.txt
outFile         results.txt
spectraFile     spectra.ps
flagSubsection  false
flagOutputRecon false
flagOutputResid 0
flagTrim        1
flaggedChannels 75-112
flagGrowth      1
growthCut       1.5
flagATrous      1
reconDim        1
scaleMin        1
snrRecon        4
flagFDR         1
alphaFDR        0.1
snrCut          3
threshSpatial   3
threshVelocity  7
```

Note that, as in this example, the flag parameters can be entered as strings (`true`/`false`) or integers (`1`/`0`). Also, note that it is not necessary to include all these parameters in the file, only those that need to be changed from the defaults (as listed in Appendix B), which in this case would be very few. A minimal parameter file might look like:

```
imageFile       /home/mduchamp/fountain.fits
flagLog         false
flagATrous      1
snrRecon        3
snrCut          2.5
minChannels     4
```

This will reconstruct the cube with a lower SNR value than the default, select objects at a lower threshold, with a looser minimum channel requirement, and not keep a log of the intermediate detections.

The following page demonstrates how the parameters are presented to the user, both on the screen at execution time, and in the output and log files. On each line, there is a description on the parameter, the relevant parameter name that is used in the input file (if there is one that the user can enter), and the value of the parameter being used.

```
# ---- Parameters ----
# Image to be analysed.............................[imageFile]  =  fountain.fits
# Intermediate Logfile...............................[logFile]  =  duchamp-Logfile.txt
# Final Results file.................................[outFile]  =  duchamp-Results.txt
# Header for results file.........................[headerFile]  =  duchamp-Results.hdr
# Spectrum file..................................[spectraFile]  =  duchamp-Spectra.ps
# Text file with ascii spectral data.........[spectraTextFile]  =  duchamp-Spectra.txt
# VOTable file.......................................[votFile]  =  duchamp-Results.xml
# Karma annotation file............................[karmaFile]  =  duchamp-Results.ann
# DS9 annotation file................................[ds9File]  =  duchamp-Results.reg
# CASA annotation file..............................[casaFile]  =  duchamp-Results.crf
# 0th Moment Map...................................[momentMap]  =  duchamp-MomentMap.ps
# Detection Map.................................[detectionMap]  =  duchamp-DetectionMap.ps
# Display a map in a pgplot xwindow?.............[flagXOutput]  =  true
# Saving reconstructed cube?.................[flagOutputRecon]  =  true --> fountain.RECON-1-1-4-1-8-0.005.fits
# Saving residuals from reconstruction?......[flagOutputResid]  =  true --> latestResid.fits
# Saving mask cube?...........................[flagOutputMask]  =  true --> latestmask2.fits
# Saving 0th moment to FITS file?........[flagOutputMomentMap]  =  true --> latestmom0.fits
# Saving 0th moment mask to FITS file?..[flagOutputMomentMask]  =  true --> latestmom0mask.fits
# Saving baseline values to FITS file?....[flagOutputBaseline]  =  false
# ------
# Type of searching performed.....................[searchType]  =  spectral
# Blank Pixel Value...........................................  =  -8.00061
# Trimming Blank Pixels?............................[flagTrim]  =  false
# Searching for Negative features?..............[flagNegative]  =  false
# Channels flagged by user...................[flaggedChannels]  =  75-112
# Area of Beam (pixels).......................................  =  14.6848   (beam: 3.6 x 3.6 pixels)
# Removing baselines before search?.............[flagBaseline]  =  false
# Smoothing data prior to searching?..............[flagSmooth]  =  false
# Using A Trous reconstruction?...................[flagATrous]  =  true
# Number of dimensions in reconstruction............[reconDim]  =  1
# Scales used in reconstruction............[scaleMin-scaleMax]  =  1-8
# SNR Threshold within reconstruction...............[snrRecon]  =  4
# Residual convergence criterion............[reconConvergence]  =  0.005
# Filter being used for reconstruction............[filterCode]  =  1 (B3 spline function)
# Using Robust statistics?...................[flagRobustStats]  =  true
# Using FDR analysis?................................[flagFDR]  =  false
# SNR Threshold (in sigma).............................[snrCut]  =  3.5
# Minimum # Pixels in a detection.....................[minPix]  =  5
# Minimum # Channels in a detection..............[minChannels]  =  3
# Minimum # Voxels in a detection..................[minVoxels]  =  7
# Growing objects after detection?................[flagGrowth]  =  false
# Using Adjacent-pixel criterion?...............[flagAdjacent]  =  true
# Max. velocity separation for merging........[threshVelocity]  =  7
# Reject objects before merging?.......[flagRejectBeforeMerge]  =  false
# Merge objects in two stages?...........[flagTwoStageMerging]  =  false
# Method of spectral plotting.................[spectralMethod]  =  peak
# Type of object centre used in results..........[pixelCentre]  =  centroid
# --------------------
```

D Example results file

This the typical content of an output file, after running *Duchamp* with the parameters illustrated on the previous page. The table is wide, and has been split into four sections to make it easier to read.

```
#-------------------------------------------------------------------------------------------
# ObjID           Name      X      Y      Z         RA       DEC         RA        DEC      VRAD
#                                                                   [deg]      [deg]   [km/s]
#-------------------------------------------------------------------------------------------
      1 J060930-215738  58.9 140.4 114.5 06:09:30.9 -21:57:38  92.378723 -21.960721  223.964
      2 J060143-250033  86.0  94.9 117.9 06:01:43.0 -25:00:33  90.429334 -25.009433  269.253
      3 J060217-254714  84.0  83.2 118.0 06:02:17.6 -25:47:14  90.573276 -25.787385  270.507
      4 J060605-271848  71.4  60.2 121.3 06:06:05.8 -27:18:48  91.524232 -27.313549  313.411
      5 J061118-213635  52.5 145.5 162.6 06:11:18.6 -21:36:35  92.827691 -21.609736  858.050
      6 J060034-285857  89.7  35.3 202.5 06:00:34.0 -28:58:57  90.141557 -28.982718 1384.415
      7 J061700-272250  35.0  58.6 216.5 06:17:00.9 -27:22:50  94.253896 -27.380808 1570.316
      8 J055847-252517  95.9  88.6 232.9 05:58:47.5 -25:25:17  89.697895 -25.421585 1786.082
      9 J060053-214223  88.9 144.3 233.2 06:00:53.3 -21:42:23  90.222109 -21.706623 1789.639
     10 J060444-260644  75.8  78.3 233.2 06:04:44.7 -26:06:44  91.186272 -26.112238 1790.501
     11 J061706-272510  34.7  58.0 235.2 06:17:06.8 -27:25:10  94.278350 -27.419558 1816.205
     12 J060106-233956  88.0 115.0 235.5 06:01:06.8 -23:39:56  90.278502 -23.665813 1820.251
     13 J061209-214921  49.6 142.3 269.4 06:12:09.6 -21:49:21  93.039947 -21.822543 2267.524
     14 J060924-223303  59.4 131.5 295.3 06:09:24.2 -22:33:03  92.350788 -22.550861 2609.616
     15 J055505-295651 107.4  20.7 367.5 05:55:05.4 -29:56:51  88.772554 -29.947667 3561.487
```

```
----------------------------------------------------------------------------------
   MAJ    MIN     PA     w_RA    w_DEC    w_50    w_20  w_VRAD     F_int    eF_int
[deg]   [deg]  [deg] [arcmin] [arcmin]  [km/s]  [km/s]  [km/s] [Jy km/s] [Jy km/s]
----------------------------------------------------------------------------------
  0.43  0.280 113.19    52.43    31.33  25.752  40.905  65.957     9.767     0.127
  0.32  0.166  91.63    28.00    16.02  23.753  42.244  26.383     1.687     0.055
  0.22  0.013 179.43    20.01    15.99  28.934  40.975  26.383     1.232     0.049
  0.52  0.380  87.97    48.26    27.61  28.263  42.272  26.383     5.916     0.104
  0.27  0.235 140.56    24.32    19.63  87.851 111.107 118.722    27.796     0.169
  0.22  0.186 133.22    15.93    24.07 175.885 194.031 197.870    12.535     0.155
  0.15  0.096 138.15    12.26     7.65  65.960 329.015  52.765     0.647     0.039
  0.20  0.181  90.68    19.91    16.11 232.786 279.536 211.061     4.453     0.102
  0.32  0.203 117.37    27.95    24.13  93.004 226.209 197.870    16.373     0.175
  0.25  0.199 134.05    16.11    19.91 211.882 228.161 211.061    12.293     0.161
  0.18  0.108 176.19    16.39    11.53  60.490 327.217  52.765     1.243     0.054
  0.27  0.230  90.44    27.96    28.07 199.616 248.932 263.826    50.402     0.256
  0.14  0.094 128.55    16.21    11.72 403.215 436.588 382.548     5.813     0.122
  0.46  0.145   9.67    20.56    43.79  13.748  25.718  26.383     1.688     0.053
  0.24  0.188 155.90    15.76    16.25  30.543  45.907  39.574     3.131     0.071
```

```
-----------------------------------------------------------------------------------------
     F_tot    eF_tot    F_peak S/Nmax  X1  X2  Y1  Y2  Z1  Z2 Nvoxel Nchan Nspatpix Flag
[Jy/beam] [Jy/beam] [Jy/beam]
-----------------------------------------------------------------------------------------
    10.873     0.142     0.213  16.17  53  65 136 143 113 118    116     4       57    F
     1.878     0.062     0.124   8.99  83  89  93  96 117 119     22     3       17    -
     1.372     0.054     0.118   8.60  82  86  82  85 117 119     17     3       13    -
     6.586     0.115     0.150  11.29  65  76  57  63 120 122     77     3       46    -
    30.943     0.188     0.410  31.17  50  55 143 147 158 167    205    10       27    E
    13.954     0.173     0.173  13.03  88  91  33  38 195 210    172    16       20    -
     0.720     0.044     0.078   4.92  34  36  58  59 215 219     11     5        5    -
     4.957     0.113     0.115   7.88  93  97  87  90 222 238     74    12       10    -
    18.227     0.195     0.166  11.99  86  92 142 147 224 239    219    16       29    E
    13.685     0.179     0.155   9.18  74  77  76  80 225 241    185    17       16    -
     1.384     0.060     0.093   5.73  33  36  57  59 233 237     21     5        9    -
    56.108     0.285     0.297  21.00  85  91 112 118 226 246    469    21       34    -
     6.471     0.135     0.101   6.10  48  51 141 143 257 286    106    30        9    -
     1.879     0.059     0.177  12.81  57  61 127 137 295 297     20     3       20    -
     3.485     0.079     0.169  12.66 106 109  19  22 366 369     36     4       14    -
```

```
------------------------------------------------------------
  X_av  Y_av  Z_av X_cent Y_cent Z_cent X_peak Y_peak Z_peak

------------------------------------------------------------
  59.1 140.4 114.6   58.9  140.4  114.5     59    140    114
  86.0  94.9 117.9   86.0   94.9  117.9     85     95    118
  84.1  83.2 118.0   84.0   83.2  118.0     84     83    118
  71.2  60.2 121.2   71.4   60.2  121.3     72     61    121
  52.5 145.4 162.5   52.5  145.5  162.6     53    146    164
  89.7  35.3 202.7   89.7   35.3  202.5     90     36    197
  35.0  58.5 216.6   35.0   58.6  216.5     35     59    215
  95.9  88.6 232.6   95.9   88.6  232.9     96     89    237
  88.8 144.3 232.7   88.9  144.3  233.2     89    144    233
  75.8  78.3 233.2   75.8   78.3  233.2     76     78    240
  34.7  58.0 235.0   34.7   58.0  235.2     35     58    236
  88.1 115.0 235.8   88.0  115.0  235.5     88    115    231
  49.6 142.3 270.0   49.6  142.3  269.4     50    142    259
  59.2 131.6 295.5   59.4  131.5  295.3     60    132    295
 107.5  20.7 367.5  107.4   20.7  367.5    108     21    367
```

A good trick for those using UNIX/Linux is to make use of the `a2ps` command. The following works well:

```
> a2ps -1 -r -f5 -o duchamp-Results.ps duchamp-Results.txt
```

and produces a postscript file `duchamp-Results.ps`.

The table is preceded by information on the parameters used (see the previous section), the statistics and threshold values, and the number of detections. If `flagSeparateHeader` is set to `true`, this information is put in a separate file. This information looks like the following:

```
# Results of the Duchamp source finder v.1.5: Wed Aug 28 11:07:28 2013
#
# ---- Parameters ----
      (... omitted for clarity -- see previous page for examples...)
# --------------------
#
# --------------------
# Summary of statistics:
# Detection threshold = 0.0461842 Jy/beam
# Noise level = 0.000122074, Noise spread = 0.0131606
# Full stats:
#  Mean   = 0.000421091 Std.Dev. = 0.013392
#  Median = 0.000122074 MADFM    = 0.00887669 (= 0.0131606 as std.dev.)
# --------------------
# Total number of detections = 15
# --------------------
```

E Example VOTable output

This is part of the VOTable, in XML format, corresponding to a search with wavelet reconstruction.

```
<?xml version="1.0"?>
<VOTABLE version="1.1" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
 xsi:noNamespaceSchemaLocation="http://www.ivoa.net/xml/VOTable/VOTable/v1.1">
  <COOSYS ID="J2000" equinox="J2000" system="eq_FK5"/>
  <RESOURCE name="Duchamp Output">
    <TABLE name="Detections">
      <DESCRIPTION>Detected sources and parameters from running the Duchamp source finder.</DESCRIPTION>
      <PARAM name="imageFile" ucd="meta.file;meta.fits" datatype="char" arraysize="55" value="/Users/whi550/ObsData/cubes/H201_abcde_luther_chop.fits"/>
      <PARAM name="flagSubsection" ucd="meta.code" datatype="boolean" value="0"/>
      <PARAM name="flagStatSec" ucd="meta.code" datatype="boolean" value="0"/>
      <PARAM name="searchType" ucd="meta.note" datatype="char" arraysize="8" value="spectral"/>
      <PARAM name="flagNegative" ucd="meta.code" datatype="boolean" value="0"/>
      <PARAM name="flagBaseline" ucd="meta.code" datatype="boolean" value="0"/>
      <PARAM name="flagRobustStats" ucd="meta.code" datatype="boolean" value="1"/>
      <PARAM name="flagFDR" ucd="meta.code" datatype="boolean" value="0"/>
      <PARAM name="snrCut" ucd="stat.snr;phot;stat.min" datatype="float" value="3.5"/>
      <PARAM name="flagGrowth" ucd="meta.code" datatype="boolean" value="0"/>
      <PARAM name="minPix" ucd="" datatype="int" value="5"/>
      <PARAM name="minChannels" ucd="" datatype="int" value="3"/>
      <PARAM name="minVoxels" ucd="" datatype="int" value="7"/>
      <PARAM name="flagAdjacent" ucd="meta.code" datatype="boolean" value="1"/>
      <PARAM name="threshVelocity" ucd="" datatype="float" value="7"/>
      <PARAM name="flagRejectBeforeMerge" ucd="" datatype="boolean" value="0"/>
      <PARAM name="flagTwoStageMerging" ucd="" datatype="boolean" value="0"/>
      <PARAM name="pixelCentre" ucd="" datatype="char" arraysize="8" value="centroid"/>
      <PARAM name="flagSmooth" ucd="meta.code" datatype="boolean" value="0"/>
      <PARAM name="flagATrous" ucd="meta.code" datatype="boolean" value="1"/>
      <PARAM name="reconDim" ucd="" datatype="int" value="1"/>
      <PARAM name="scaleMin" ucd="" datatype="int" value="1"/>
      <PARAM name="scaleMax" ucd="" datatype="int" value="8"/>
      <PARAM name="snrRecon" ucd="" datatype="float" value="4"/>
      <PARAM name="reconConvergence" ucd="" datatype="float" value="0.005"/>
      <PARAM name="filterCode" ucd="" datatype="int" value="1"/>
      <PARAM name="thresholdActual" ucd="" datatype="float" units="Jy/beam" value="0.0461842"/>
      <PARAM name="noiseMeanActual" ucd="" datatype="float" units="Jy/beam" value="0.000122074"/>
      <PARAM name="noiseSpreadActual" ucd="" datatype="float" units="Jy/beam" value="0.0131606"/>
      <FIELD name="ObjID" ID="col_objnum" ucd="meta.id" datatype="int" unit="" width="8"/>
      <FIELD name="Name" ID="col_name" ucd="meta.id;meta.main" datatype="char" unit="" arraysize="15"/>
      <FIELD name="X" ID="col_x" ucd="pos.cartesian.x" datatype="float" unit="" width="6" precision="1"/>
      <FIELD name="Y" ID="col_y" ucd="pos.cartesian.y" datatype="float" unit="" width="6" precision="1"/>
      <FIELD name="Z" ID="col_z" ucd="pos.cartesian.z" datatype="float" unit="" width="6" precision="1"/>
      <FIELD name="RA" ID="col_rajd" ucd="pos.eq.ra;meta.main" ref="J2000" datatype="float" unit="deg" width="11" precision="6"/>
      <FIELD name="DEC" ID="col_decjd" ucd="pos.eq.dec;meta.main" ref="J2000" datatype="float" unit="deg" width="11" precision="6"/>
      <FIELD name="VRAD" ID="col_vel" ucd="phys.veloc;spect.dopplerVeloc.rad;meta.main" datatype="float" unit="km/s" width="9" precision="3"/>
      <FIELD name="MAJ" ID="col_maj" ucd="phys.angSize.smajAxis" ref="J2000" datatype="float" unit="deg" width="6" precision="2"/>
      <FIELD name="MIN" ID="col_min" ucd="phys.angSize.sminAxis" ref="J2000" datatype="float" unit="deg" width="7" precision="3"/>
      <FIELD name="PA" ID="col_pa" ucd="pos.posAng;phys.angSize" ref="J2000" datatype="float" unit="deg" width="7" precision="2"/>
      <FIELD name="w_RA" ID="col_wra" ucd="phys.angSize;pos.eq.ra" ref="J2000" datatype="float" unit="arcmin" width="9" precision="2"/>
      <FIELD name="w_DEC" ID="col_wdec" ucd="phys.angSize;pos.eq.dec" ref="J2000" datatype="float" unit="arcmin" width="9" precision="2"/>
      <FIELD name="w_50" ID="col_w50" ucd="spect.line.width.50;phys.veloc;spect.dopplerVeloc.rad" datatype="float" unit="km/s" width="8" precision="3"/>
      <FIELD name="w_20" ID="col_w20" ucd="spect.line.width.20;phys.veloc;spect.dopplerVeloc.rad" datatype="float" unit="km/s" width="8" precision="3"/>
      <FIELD name="w_VRAD" ID="col_wvel" ucd="spect.line.width.full;phys.veloc;spect.dopplerVeloc.rad" datatype="float" unit="km/s" width="8" precision="3"/>
      <FIELD name="Integrated_Flux" ID="col_fint" ucd="phot.flux.density.integratedspect.line.intensity" datatype="float" unit="Jy.km/s" width="10" precision="3"/>
      <FIELD name="Integrated_Flux_Error" ID="col_efint" ucd="phot.flux.density.integrated;stat.err" datatype="float" unit="Jy.km/s" width="10" precision="3"/>
      <FIELD name="Peak_Flux" ID="col_fpeak" ucd="phot.flux;stat.maxspect.line.intensity" datatype="float" unit="Jy/beam" width="10" precision="3"/>
      <FIELD name="S/Nmax" ID="col_snrmax" ucd="stat.snr;phot.flux" datatype="float" unit="" width="7" precision="2"/>
      <FIELD name="X1" ID="col_x1" ucd="pos.cartesian.x;stat.min" datatype="int" unit="" width="4"/>
      <FIELD name="X2" ID="col_x2" ucd="pos.cartesian.x;stat.max" datatype="int" unit="" width="4"/>
      <FIELD name="Y1" ID="col_y1" ucd="pos.cartesian.y;stat.min" datatype="int" unit="" width="4"/>
      <FIELD name="Y2" ID="col_y2" ucd="pos.cartesian.y;stat.max" datatype="int" unit="" width="4"/>
      <FIELD name="Z1" ID="col_z1" ucd="pos.cartesian.z;stat.min" datatype="int" unit="" width="4"/>
      <FIELD name="Z2" ID="col_z2" ucd="pos.cartesian.z;stat.max" datatype="int" unit="" width="4"/>
      <FIELD name="Nvoxel" ID="col_nvox" ucd="phys.size;instr.pixel;em.bin" datatype="int" unit="" width="7"/>
      <FIELD name="Nchan" ID="col_nch" ucd="spect.line.width.full;em.bin" datatype="int" unit="" width="6"/>
      <FIELD name="Nspatpix" ID="col_spsize" ucd="phys.angArea;instr.pixel" datatype="int" unit="" width="9"/>
      <FIELD name="Flag" ID="col_flag" ucd="meta.code.qual" datatype="char" unit="" arraysize="5"/>
      <FIELD name="X_av" ID="col_xav" ucd="pos.cartesian.x;stat.mean" datatype="float" unit="" width="6" precision="1"/>
      <FIELD name="Y_av" ID="col_yav" ucd="pos.cartesian.y;stat.mean" datatype="float" unit="" width="6" precision="1"/>
      <FIELD name="Z_av" ID="col_zav" ucd="pos.cartesian.z;stat.mean" datatype="float" unit="" width="6" precision="1"/>
      <FIELD name="X_Centroid" ID="col_xcent" ucd="pos.cartesian.x;stat.centroid" datatype="float" unit="" width="7" precision="1"/>
      <FIELD name="Y_Centroid" ID="col_ycent" ucd="pos.cartesian.y;stat.centroid" datatype="float" unit="" width="7" precision="1"/>
      <FIELD name="Z_Centroid" ID="col_zcent" ucd="pos.cartesian.z;stat.centroid" datatype="float" unit="" width="7" precision="1"/>
      <FIELD name="X_peak" ID="col_xpeak" ucd="pos.cartesian.x;phot.flux;stat.max" datatype="int" unit="" width="7"/>
      <FIELD name="Y_peak" ID="col_ypeak" ucd="pos.cartesian.y;phot.flux;stat.max" datatype="int" unit="" width="7"/>
      <FIELD name="Z_peak" ID="col_zpeak" ucd="pos.cartesian.z;phot.flux;stat.max" datatype="int" unit="" width="7"/>
      <DATA>
        <TABLEDATA>
        <TR>
          <TD>       1</TD><TD> J060930-215738</TD><TD>  58.9</TD><TD> 140.4</TD><TD> 114.5</TD><TD>  92.378723</TD>
<TD> -21.960721</TD><TD>  223.964</TD><TD>  0.43</TD><TD>  0.280</TD><TD> 113.19</TD><TD>    52.43</TD><TD>    31.33</TD>
<TD>  25.752</TD><TD>  40.905</TD><TD>  65.957</TD><TD>     9.767</TD><TD>     0.127</TD><TD>     0.213</TD><TD>  16.17</TD>
<TD>  53</TD><TD>  65</TD><TD> 136</TD><TD> 143</TD><TD> 113</TD><TD> 118</TD><TD>    116</TD><TD>     4</TD><TD>       57</TD>
<TD>    F</TD><TD>  59.1</TD><TD> 140.4</TD><TD> 114.6</TD><TD>   58.9</TD><TD>  140.4</TD><TD>  114.5</TD><TD>     59</TD>
<TD>    140</TD><TD>    114</TD>
        </TR>
... truncated ...
        </TABLEDATA>
      </DATA>
    </TABLE>
  </RESOURCE>
</VOTABLE>
```

F Example Karma Annotation file output

This is the format of the Karma Annotation file, showing the locations of the detected objects. This can be loaded by the plotting tools of the Karma package (for instance, `kvis`) as an overlay on the FITS file.

```
# Duchamp Source Finder v.1.5
# Results for FITS file: /home/mduchamp/fountain.fits
# imageFile             /home/mduchamp/fountain.fits
# flagSubsection        0
# flagStatSec           0
# searchType            spectral
# flagNegative          0
# flagBaseline          0
# flagRobustStats       1
# flagFDR               0
# snrCut                3.5
# flagGrowth            0
# minPix                5
# minChannels           3
# minVoxels             7
# flagAdjacent          1
# threshVelocity        7
# flagRejectBeforeMerge 0
# flagTwoStageMerging   0
# pixelCentre           centroid
# flagSmooth            0
# flagATrous            1
# reconDim              1
# scaleMin              1
# scaleMax              8
# snrRecon              4
# reconConvergence      0.005
# filterCode            1
# Detection threshold used = 0.0461842
# Mean of noise background = 0.000122074
# Std. Deviation of noise background = 0.0131606
#  [Using robust methods]
COLOR RED
PA STANDARD
COORD W
LINE 92.839864 -22.144360 92.767849 -22.145538
LINE 92.838756 -22.077549 92.766775 -22.078727
LINE 92.767849 -22.145538 92.695836 -22.146680
...
TEXT 92.378723 -21.960721 1
```

G Example DS9 Region file output

This is the format of the DS9 region file, showing the locations of the detected objects. This can be loaded by the visualisation tool SAOImage DS9 as an overlay on the FITS file, and should be compatible with CASA's casaviewer.

```
# Duchamp Source Finder v.1.5
# Results for FITS file: /home/mduchamp/fountain.fits
# imageFile             /home/mduchamp/fountain.fits
# flagSubsection        0
# flagStatSec           0
# searchType            spectral
# flagNegative          0
# flagBaseline          0
# flagRobustStats       1
# flagFDR               0
# snrCut                3.5
# flagGrowth            0
# minPix                5
# minChannels           3
# minVoxels             7
# flagAdjacent          1
# threshVelocity        7
# flagRejectBeforeMerge 0
# flagTwoStageMerging   0
# pixelCentre           centroid
# flagSmooth            0
# flagATrous            1
# reconDim              1
# scaleMin              1
# scaleMax              8
# snrRecon              4
# reconConvergence      0.005
# filterCode            1
# Detection threshold used = 0.0461842
# Mean of noise background = 0.000122074
# Std. Deviation of noise background = 0.0131606
#  [Using robust methods]
global color=red
fk5
line 92.839864 -22.144360 92.767849 -22.145538
line 92.838756 -22.077549 92.766775 -22.078727
line 92.767849 -22.145538 92.695836 -22.146680
...
text 92.378723 -21.960721 {1}
```

H Example CASA Region file output

This is the format of the CASA region file, showing the locations of the detected objects. This can be loaded in casapy, and should be able to be used in the casaviewer image viewer (once that functionality is made available)

```
#CRTFv0
# Duchamp Source Finder v.1.5
# Results for FITS file: /home/mduchamp/fountain.fits
# imageFile             /home/mduchamp/fountain.fits
# flagSubsection        0
# flagStatSec           0
# searchType            spectral
# flagNegative          0
# flagBaseline          0
# flagRobustStats       1
# flagFDR               0
# snrCut                3.5
# flagGrowth            0
# minPix                5
# minChannels           3
# minVoxels             7
# flagAdjacent          1
# threshVelocity        7
# flagRejectBeforeMerge 0
# flagTwoStageMerging   0
# pixelCentre           centroid
# flagSmooth            0
# flagATrous            1
# reconDim              1
# scaleMin              1
# scaleMax              8
# snrRecon              4
# reconConvergence      0.005
# filterCode            1
# Detection threshold used = 0.0461842
# Mean of noise background = 0.000122074
# Std. Deviation of noise background = 0.0131606
#  [Using robust methods]
global color=red, coord=J2000
#
box[[92.842095deg,-22.277967deg], [91.899837deg,-21.755824deg]], label=’1’
line[[92.839864deg,-22.144360deg], [92.767849deg,-22.145538deg]]
line[[92.838756deg,-22.077549deg], [92.766775deg,-22.078727deg]]
line[[92.767849deg,-22.145538deg], [92.695836deg,-22.146680deg]]
...
text[[92.378723deg,-21.960721deg], ’1’]
```

I Robust statistics for a Normal distribution

The Normal, or Gaussian, distribution for mean μ and standard deviation σ can be written as

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\; e^{-(x-\mu)^2/2\sigma^2}.$$

When one has a purely Gaussian signal, it is straightforward to estimate σ by calculating the standard deviation (or rms) of the data. However, if there is a small amount of signal present on top of Gaussian noise, and one wants to estimate the σ for the noise, the presence of the large values from the signal can bias the estimator to higher values.

An alternative way is to use the median (m) and median absolute deviation from the median (s) to estimate μ and σ. The median is the middle of the distribution, defined for a continuous distribution by

$$\int_{-\infty}^{m} f(x)dx = \int_{m}^{\infty} f(x)dx.$$

From symmetry, we quickly see that for the continuous Normal distribution, $m = \mu$. We consider the case henceforth of $\mu = 0$, without loss of generality.

To find s, we find the distribution of the absolute deviation from the median, and then find the median of that distribution. This distribution is given by

$$\begin{aligned} g(x) &= \text{distribution of } |x| \\ &= f(x) + f(-x),\ x \geq 0 \\ &= \sqrt{\frac{2}{\pi\sigma^2}}\, e^{-x^2/2\sigma^2},\ x \geq 0. \end{aligned}$$

So, the median absolute deviation from the median, s, is given by

$$\int_0^s g(x)dx = \int_s^\infty g(x)dx.$$

If we use the identity

$$\int_0^\infty e^{-x^2/2\sigma^2} dx = \sqrt{\pi\sigma^2/2}$$

we find that

$$\int_s^\infty e^{-x^2/2\sigma^2} dx = \sqrt{\pi\sigma^2/2} - \int_0^s e^{-x^2/2\sigma^2} dx.$$

Hence, to find s we simply solve the following equation (setting $\sigma = 1$ for simplicity – equivalent to stating x and s in units of σ):

$$\int_0^s e^{-x^2/2} dx - \sqrt{\pi/8} = 0.$$

This is hard to solve analytically (no nice analytic solution exists for the finite integral that I'm aware of), but straightforward to solve numerically, yielding the value of $s = 0.6744888$. Thus, to estimate σ for a Normally distributed data set, one can calculate s, then divide by 0.6744888 (or multiply by 1.4826042) to obtain the correct estimator.

Note that this is different to solutions quoted elsewhere, specifically in Meyer et al. (2004), where the same robust estimator is used but with an incorrect conversion to standard deviation – they assume $\sigma = s\sqrt{\pi/2}$. This, in fact, is the conversion used to convert the *mean* absolute deviation from the mean to the standard deviation. This means that the cube noise in the HIPASS catalogue (their parameter $\mathrm{Rms}_{\mathrm{cube}}$) should be 18% larger than quoted.

J Gaussian noise and the wavelet scale

The key element in the wavelet reconstruction of an array is the thresholding of the individual wavelet coefficient arrays. This is usually done by choosing a level to be some number of standard deviations above the mean value.

However, since the wavelet arrays are produced by convolving the input array by an increasingly large filter, the pixels in the coefficient arrays become increasingly correlated as the scale of the filter increases. This results in the measured standard deviation from a given coefficient array decreasing with increasing scale. To calculate this, we need to take into account how many other pixels each pixel in the convolved array depends on.

To demonstrate, suppose we have a 1-D array with N pixel values given by $F_i,\ i = 1, ..., N$, and we convolve it with the B$_3$-spline filter, defined by the set of coefficients $\{1/16, 1/4, 3/8, 1/4, 1/16\}$. The flux of the ith pixel in the convolved array will be

$$F'_i = \frac{1}{16}F_{i-2} + \frac{1}{4}F_{i-1} + \frac{3}{8}F_i + \frac{1}{4}F_{i+1} + \frac{1}{16}F_{i+2}$$

and the flux of the corresponding pixel in the wavelet array will be

$$W'_i = F_i - F'_i = \frac{-1}{16}F_{i-2} - \frac{1}{4}F_{i-1} + \frac{5}{8}F_i - \frac{1}{4}F_{i+1} - \frac{1}{16}F_{i+2}$$

Now, assuming each pixel has the same standard deviation $\sigma_i = \sigma$, we can work out the standard deviation for the wavelet array:

$$\sigma'_i = \sigma\sqrt{\left(\frac{1}{16}\right)^2 + \left(\frac{1}{4}\right)^2 + \left(\frac{5}{8}\right)^2 + \left(\frac{1}{4}\right)^2 + \left(\frac{1}{16}\right)^2} = 0.72349\ \sigma$$

Thus, the first scale wavelet coefficient array will have a standard deviation of 72.3% of the input array. This procedure can be followed to calculate the necessary values for all scales, dimensions and filters used by *Duchamp*.

Calculating these values is clearly a critical step in performing the reconstruction. The method used by Starck and Murtagh (2002) was to simulate data sets with Gaussian noise, take the wavelet transform, and measure the value of σ for each scale. We take a different approach, by calculating the scaling factors directly from the filter coefficients by taking the wavelet transform of an array made up of a 1 in the central pixel and 0s everywhere else. The scaling value is then derived by taking the square root of the sum (in quadrature) of all the wavelet coefficient values at each scale. We give the scaling factors for the three filters available to *Duchamp* below. These values are hard-coded into *Duchamp*, so no on-the-fly calculation of them is necessary.

Memory limitations prevent us from calculating factors for large scales, particularly for the three-dimensional case (hence the smaller table). To calculate factors for higher scales than those available, we divide the previous scale's factor by either $\sqrt{2}$, 2, or $\sqrt{8}$ for 1D, 2D and 3D respectively.

	B_3 Spline	Triangle	Haar
	$\{\frac{1}{16},\frac{1}{4},\frac{3}{8},\frac{1}{4},\frac{1}{16}\}$	$\{\frac{1}{4},\frac{1}{2},\frac{1}{4}\}$	$\{0,\frac{1}{2},\frac{1}{2}\}$
1 dimension			
1	0.723489806	0.612372436	0.707106781
2	0.285450405	0.330718914	0.5
3	0.177947535	0.211947812	0.353553391
4	0.122223156	0.145740298	0.25
5	0.0858113122	0.102310944	0.176776695
6	0.0605703043	0.0722128185	0.125
7	0.0428107206	0.0510388224	0.0883883476
8	0.0302684024	0.0360857673	0.0625
9	0.0214024008	0.0255157615	0.0441941738
10	0.0151336781	0.0180422389	0.03125
11	0.0107011079	0.0127577667	0.0220970869
12	0.00756682272	0.00902109930	0.015625
13	0.00535055108	0.00637887978	0.0110485435
2 dimension			
1	0.890796310	0.800390530	0.866025404
2	0.200663851	0.272878894	0.433012702
3	0.0855075048	0.119779282	0.216506351
4	0.0412474444	0.0577664785	0.108253175
5	0.0204249666	0.0286163283	0.0541265877
6	0.0101897592	0.0142747506	0.0270632939
7	0.00509204670	0.00713319703	0.0135316469
8	0.00254566946	0.00356607618	0.00676582347
9	0.00127279050	0.00178297280	0.00338291173
10	0.000636389722	0.000891478237	0.00169145587
11	0.000318194170	0.000445738098	0.000845727933
3 dimension			
1	0.956543592	0.895954449	0.935414347
2	0.120336499	0.192033014	0.330718914
3	0.0349500154	0.0576484078	0.116926793
4	0.0118164242	0.0194912393	0.0413398642
5	0.00413233507	0.00681278387	0.0146158492
6	0.00145703714	0.00240175885	0.00516748303
7	0.000514791120	0.000848538128	0.00182698115