

# Report of TWG 4 – source finding and cataloguing

(23/12/2013 – by Tobias Westmeier on behalf of TWG 4)

Recent activity in TWG 4 has been focusing on two main projects. Firstly, we decided to merge the different source finding and parametrisation algorithms developed over the past few years into a single, stand-alone source finding pipeline called SoFiA. Secondly, we developed an analytic function, the so-called ‘Busy Function’, that can be used to fit the integrated HI spectra of galaxies for the purpose of accurate parametrisation. Both projects have been a great success and are described in detail in this report.

## SoFiA: the new HI source finding pipeline

Following the recent busy week of our international HI source finding collaboration in Perth, we are proud to present the first public release of SoFiA, the new HI source finding and parametrisation pipeline. The acronym ‘SoFiA’ stands for ‘Source Finding Application’ and is derived from the Greek word for wisdom (σοφία).



The purpose of SoFiA is to implement all of the source finding and parametrisation algorithms that our group has developed over the past few years in a single, user-friendly pipeline. Not only does this pipeline allow us to apply these algorithms to existing data sets in a simple and coherent way, but we will also be able to extensively test the developed algorithms and compare their performance to that of existing software.

The pipeline is written entirely in a combination of Python and C++. While the pipeline itself and most of the input, output and filtering modules are implemented in Python, time-critical components such as automatic merging of detections and source parametrisation have been written in C++ to minimise processing time. In addition, SoFiA has a strictly modular structure, allowing the users to turn individual components of the pipeline on and off and even integrate selected components into their own software. Another innovation of SoFiA is its easy-to-use graphical user interface (GUI) that allows the user to interactively change parameter settings and invoke the pipeline with a single mouse click (Fig. 1). While the pipeline can still be entirely controlled and run from the command line, the GUI is designed to make that process easier and much more convenient.

The currently released version of SoFiA is a beta version that does not come with the full set of features yet, but should be fully functional and free of major bugs and errors. The following features and algorithms have already been implemented:

- Input
  - Support for FITS data cubes
  - Application of weights cube prior to source finding
  - Spatial and spectral smoothing with different kernels
  - Automatic noise weighting
- Source finding
  - Threshold-based finder
  - Smooth + Clip finder
  - Automatic merging of detections into sources
- Source parametrisation
  - Mask optimisation for accurate flux measurement

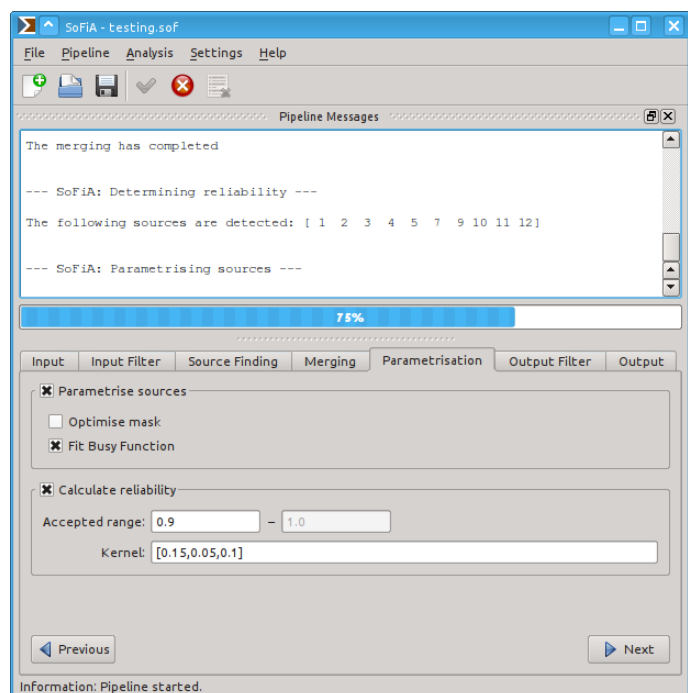


Fig. 1: Screenshot of the SoFiA user interface on a Linux system with the pipeline process running.

- Parametrisation based on Busy Function fitting (Westmeier et al. 2013)
- Calculation of reliability based on negative detections (Serra et al. 2012)
- Output
  - Different catalogue formats (ASCII, VO table)
  - Additional data products such as mask, moment maps and sub-cubes

Several algorithms and methods are yet to be implemented, including the 2D–1D wavelet filtering (Flöer & Winkel 2012), the CNHI finder (Jurek 2012), calculation of uncertainties, support for additional data formats (HDF5, CASA, etc.), additional catalogue formats (SQL), etc. All of these will be reserved for future releases of SoFiA.

The current beta version of SoFiA can be obtained from

[http://www.atnf.csiro.au/people/Tobias.Westmeier/tools\\_software\\_sofia.php](http://www.atnf.csiro.au/people/Tobias.Westmeier/tools_software_sofia.php).

The release comes with a basic installation guide and should compile and run on all common Linux and Unix systems, such as Ubuntu or Mac OS, provided that all required libraries and packages are installed. SoFiA also comes with its own integrated user manual. We hope that SoFiA will be useful to the wider HI community. If you plan to test SoFiA on your data, we would appreciate any reports of errors found in the software as well as suggestions for improvement of both the functionality and user experience.

Lastly, I would like to thank the entire team who worked tirelessly throughout two enjoyable busy weeks in Bonn and Perth this year to complete the SoFiA pipeline, in particular *Lars Flöer, Nadine Giese, Russell Jurek, Martin Meyer, Attila Popping, Paolo Serra and Benjamin Winkel*. Their great expertise and hard work have been essential in the successful completion and timely release of SoFiA.

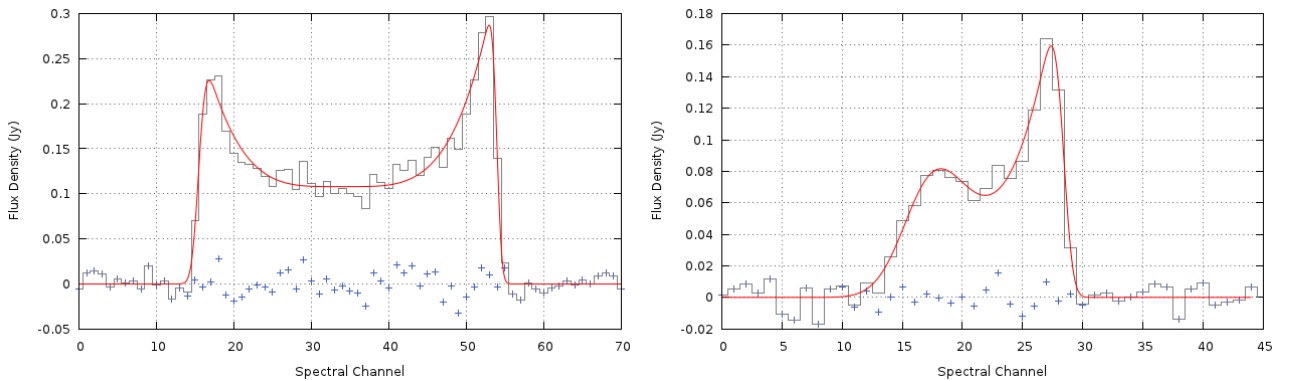


Fig. 2: Integrated HI spectra of HIPASS J1329–17 (left) and J1324–19 (right) with the fitted Busy Function shown as the red curve. The blue crosses show the residuals between the data and the fit.

## HI spectral-line fitting with the Busy Function

One of the most fundamental problems faced by HI surveys like WALLABY is the accurate parametrisation of galaxies, e.g. for studies of the HI mass function or the Tully–Fisher relation. As most galaxies detected by WALLABY are expected to be only marginally resolved, most physical information can in principle be extracted from the integrated spectrum alone. A low signal-to-noise ratio, however, will often result in large uncertainties in the derived parameters. A principal solution to this problem could be the fitting of a model to the integrated spectrum, because the fit would be much less affected by single noise peaks.

As part of our international HI source finding collaboration we have developed a new function, called the ‘Busy Function’, that can be used to fit the typical double-horn profile of galaxies (Westmeier et al. 2013). The function consists of only two types of basic functions: (1) two error functions describing the flanks of the line and (2) a simple polynomial for the central trough. In its simplest form the Busy Function reads

$$B(\xi) = [a/4] \times [\text{erf}(b(w + \xi)) + 1] \times [\text{erf}(b(w - \xi)) + 1] \times [c\xi^2 + 1]$$

with  $\xi \equiv x - x_0$ , where  $x$  is the spectral parameter, e.g. frequency or velocity. The function’s free parameters control the overall amplitude ( $a$ ), the steepness of the line flanks ( $b$ ), the depth of the central trough ( $c$ ), the half-width of the profile ( $w$ ), and the line centroid ( $x_0$ ). While this simple form of the Busy Function is symmetric, it can easily be expanded to describe asymmetric profiles, e.g. by introducing separate steepness

parameters for the two flanks or separate centroids for the error functions and the polynomial. In addition, the shape of the trough can be varied by changing the order of the polynomial, e.g. a fourth-order polynomial will produce a broader trough with sharper peaks.

Two examples of an asymmetric form of the Busy Function fitted to the integrated HI spectra of HIPASS J1329–17 and J1324–19 are shown in Fig. 2. HIPASS J1329–17 is an example of a fairly symmetric spectrum with a deep, broad trough and sharp peaks, whereas the spectrum of HIPASS J1324–19 is asymmetric with a relatively shallow trough and smoother peaks. Both types of spectra are well described by the Busy Function, but its flexibility goes well beyond the fitting of double-horn profiles: it can be demonstrated that with the right choice of parameters the Busy Function can be made to closely resemble a Gaussian function as well. Hence, both Gaussian and double-horn line profiles can be fitted with just a single function.

Several potential applications of the Busy Function come to mind. Its intended purpose is the fitting of integrated HI spectra of galaxies for the purpose of accurate parametrisation. There are several advantages to this approach over direct parametrisation methods. Firstly, the Busy Function is fitted to the spectrum as a whole and hence is less susceptible to individual noise peaks. This will lead to more accurate parameter measurements, particularly in the case of spectra at low signal-to-noise ratio. Secondly, least-squares fitting algorithms like the Levenberg–Marquardt algorithm will produce a full parameter covariance matrix from which statistical uncertainties of the function’s free parameters and any derived physical parameters can be determined.

Other potential applications of the Busy Function include the creation of templates for simulations or matched filtering algorithms. In addition, the function can potentially be used beyond HI emission lines, e.g. in HI absorption, HI stacking, or CO emission of galaxies. To support the wider use of the Busy Function, we have released two software packages for spectral line fitting. A combined C, C++, and Python library that can be integrated into existing software packages was developed by *Russell Jurek* and can be obtained from the project’s website at

<http://code.google.com/p/busy-function-fitting/>.

An easy-to-use, stand-alone fitting tool was created by *Tobias Westmeier* and can be obtained from

[http://www.atnf.csiro.au/people/Tobias.Westmeier/tools\\_software\\_busyfit.php](http://www.atnf.csiro.au/people/Tobias.Westmeier/tools_software_busyfit.php).

A detailed description of the Busy Function and its properties as well as the result of basic tests of its usefulness as a galaxy parametrisation tool can be found in Westmeier et al. (2013).

## References

- **2D–1D wavelet reconstruction as a tool for source finding in spectroscopic imaging surveys**  
Flöer, L., Winkel, B., 2012, *PASA*, 29, 244
- **The Characterised Noise HI Source Finder: detecting HI galaxies using a novel implementation of matched filtering**  
Jurek, R., 2012, *PASA*, 29, 251
- **Using negative detections to estimate source finder reliability**  
Serra, P., Jurek, R., Flöer, L., 2012, *PASA*, 29, 296
- **The busy function: a new analytic function for describing the integrated 21-cm spectral profile of galaxies**  
Westmeier, T., Jurek, R., Obreschkow, D., Koribalski, B.S., Staveley-Smith, L., 2013, *MNRAS*, in press (<http://arxiv.org/abs/1311.5308>)