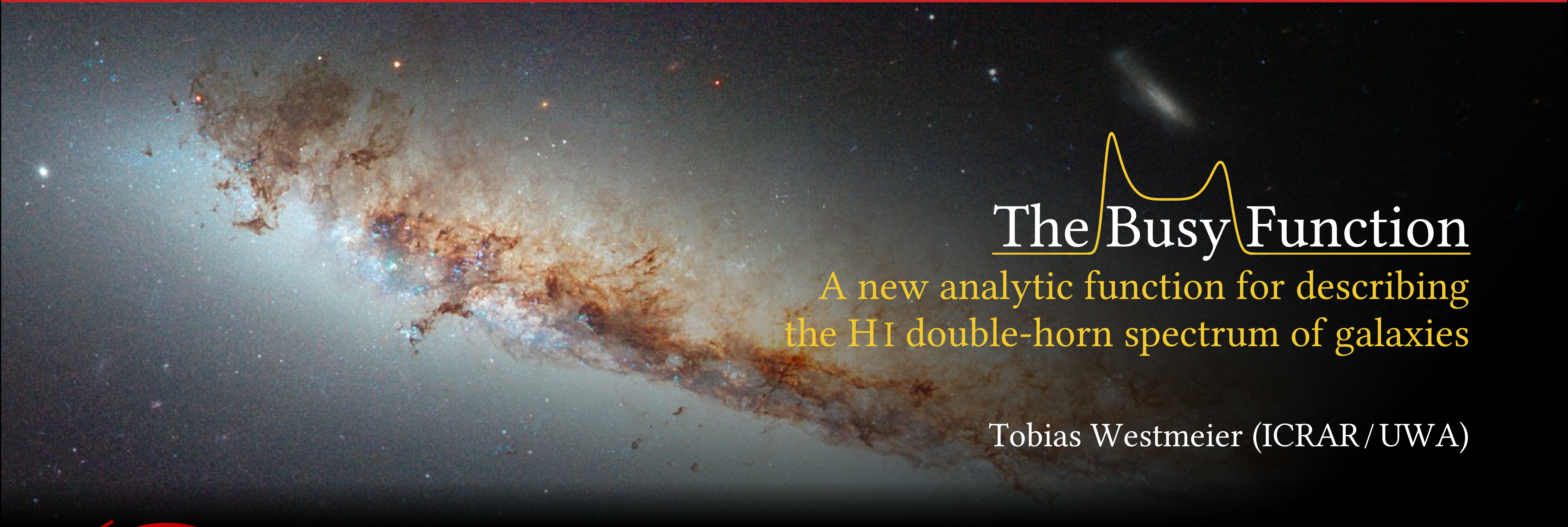


The Busy Function

A new analytic function for describing the H I double-horn spectrum of galaxies

Tobias Westmeier (ICRAR / UWA)

ICRAR
International Centre for Radio Astronomy Research

Curtin University

The University of Western Australia
Achieving International Excellence

The Government of Western Australia

Motivation

Motivation

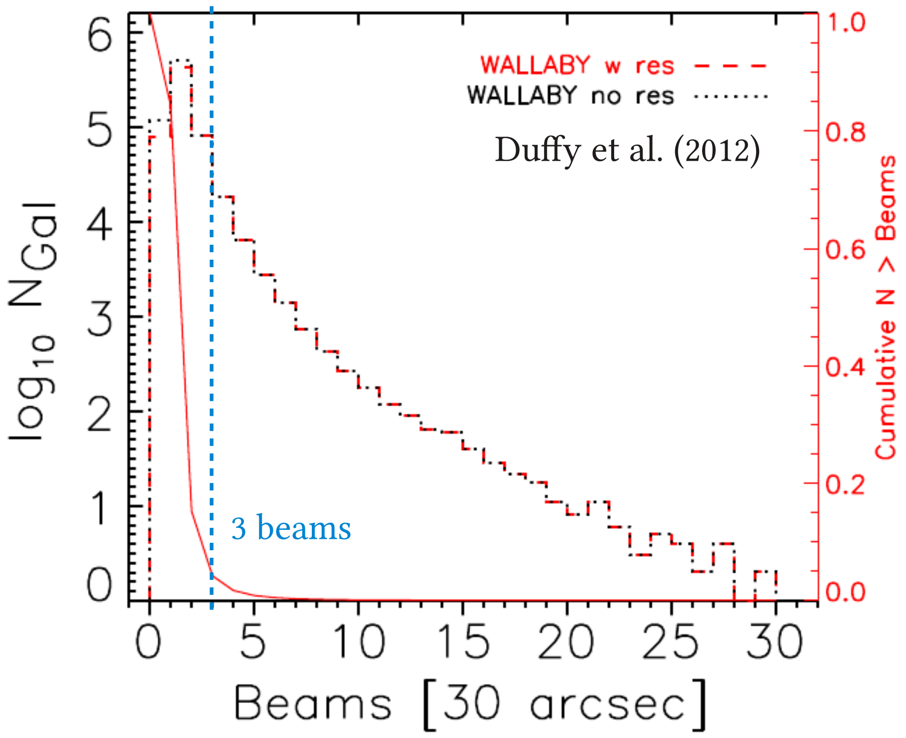
- Future H I surveys:
  - Wallaby, Dingo, Wnshs, etc.
  - Large number of galaxies.
- Parametrisation of galaxies:
  - Accurate.
  - Automated.

Runva® HWE15000 industrial winch

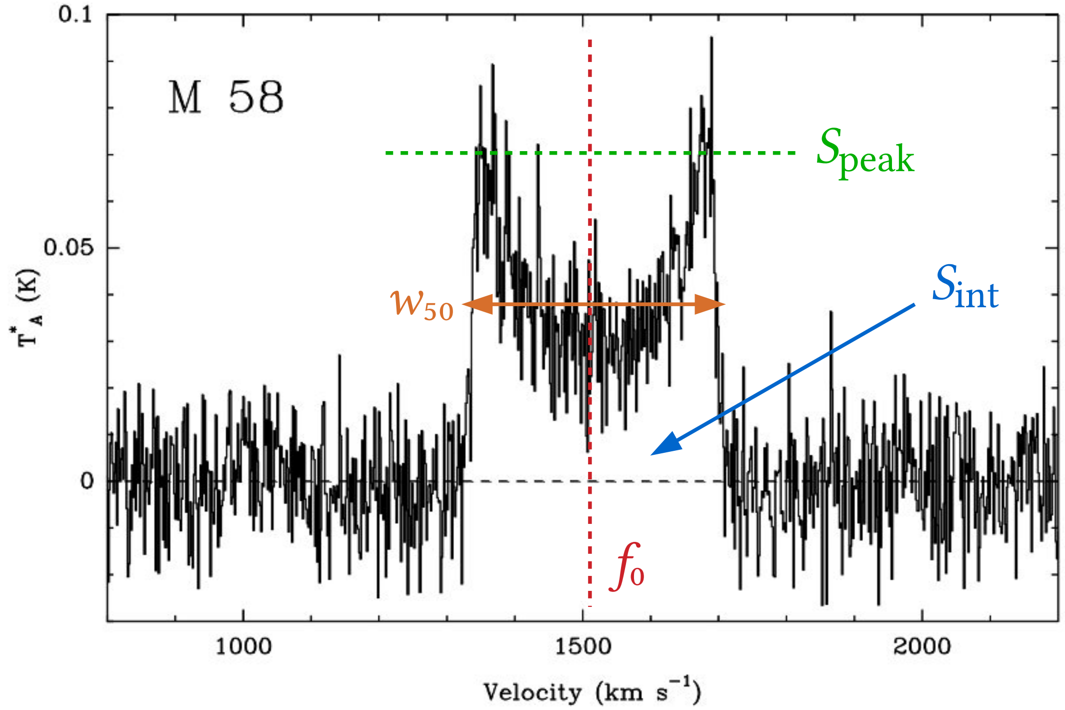
Dingo (*Canis lupus*). (Credit: Sam Fraser-Smith)

Bennett's Wallaby (*Macropus rufogriseus*). (Credit: J. J. Harrison)

- Future H I surveys:
  - WALLABY, DINGO, WNSHS, etc.
  - Large number of galaxies.
- Parametrisation of galaxies:
  - Accurate.
  - Automated.
- Most galaxies barely resolved:
  - 95% less than 3 beams across. (Duffy et al. 2012)
  - Parametrisation of integrated H I spectrum sufficient.

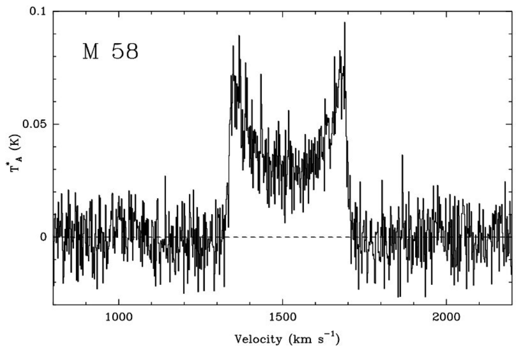


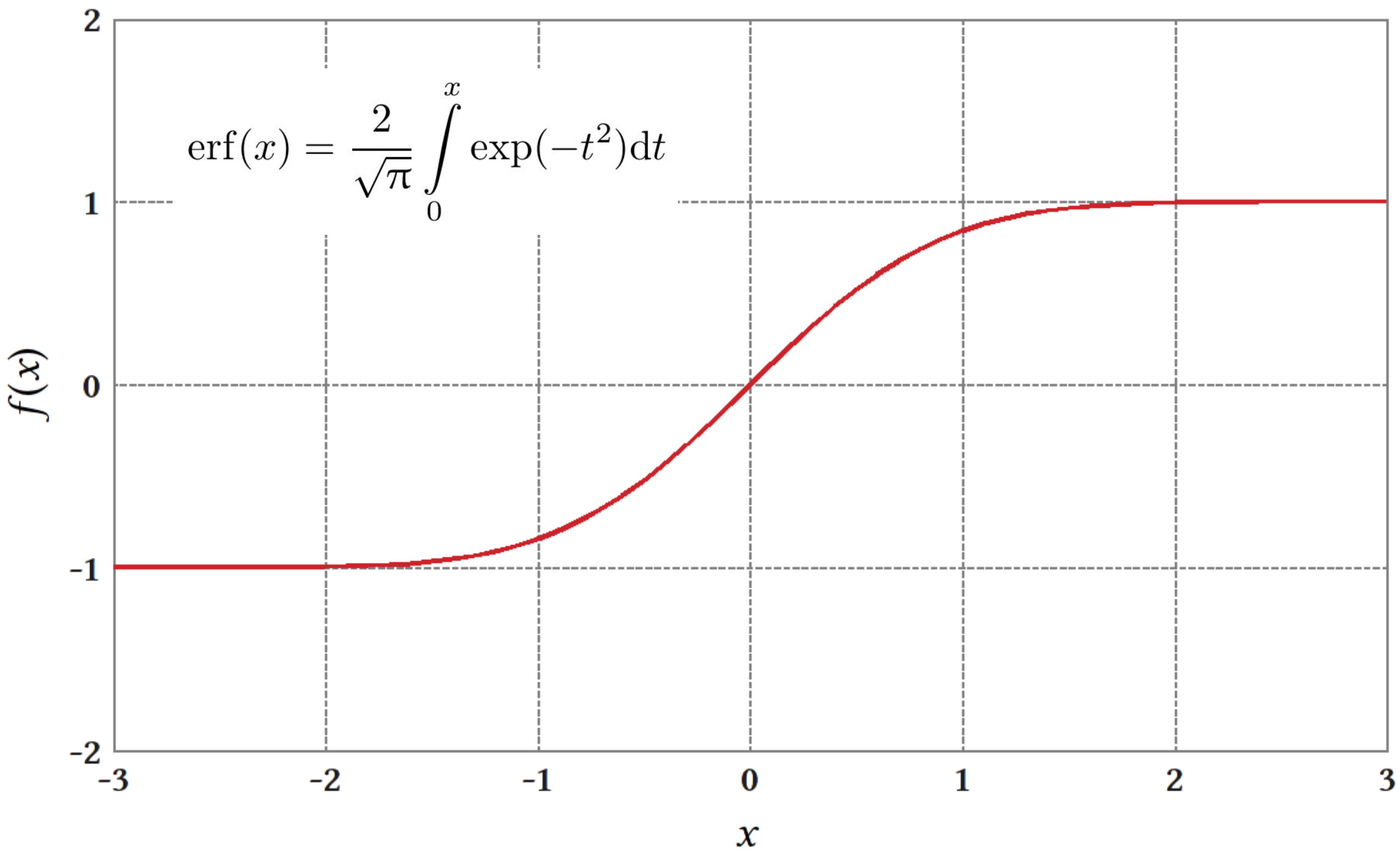
- Integrated H I spectra:
  - Characteristic double-horn profile.
  - Line centroid: f_0
  - Line width: w_{50}, w_{20}, …
  - Peak flux density: $S_{\rm peak}$
  - Integrated flux: $S_{\rm int}$
- Direct parametrisation suffers from large statistical and systematic errors.
- Better: fitting of analytic model to spectrum.
  - Less sensitive to individual noise peaks.
  - Covariance matrix of fit to derive uncertainties of parameters.



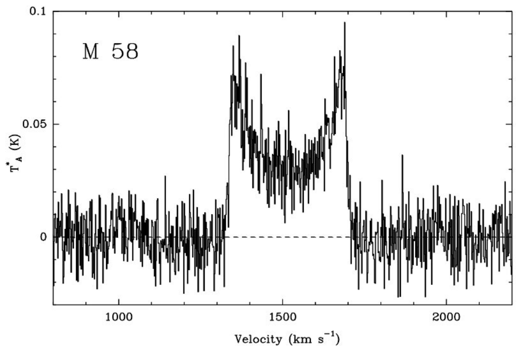
The Busy Function

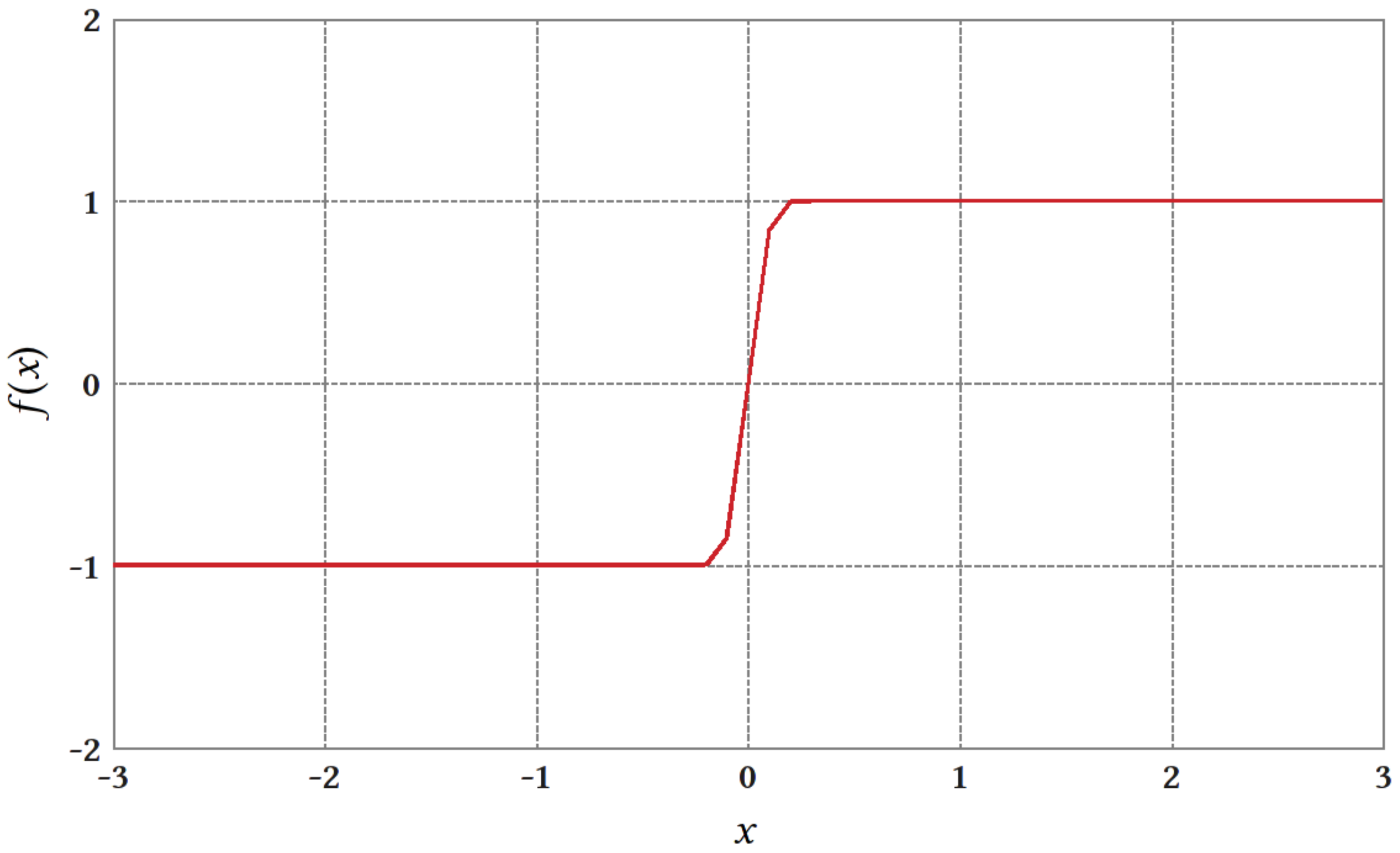
- Error function:
  - $f(x) = \mathrm{erf}(x)$



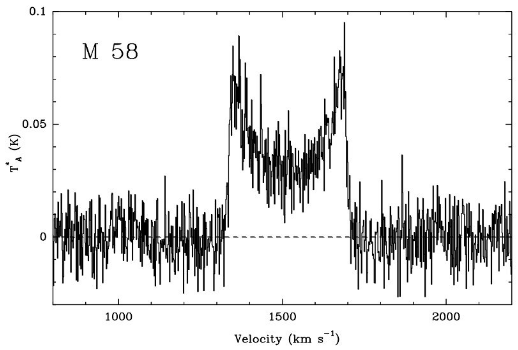


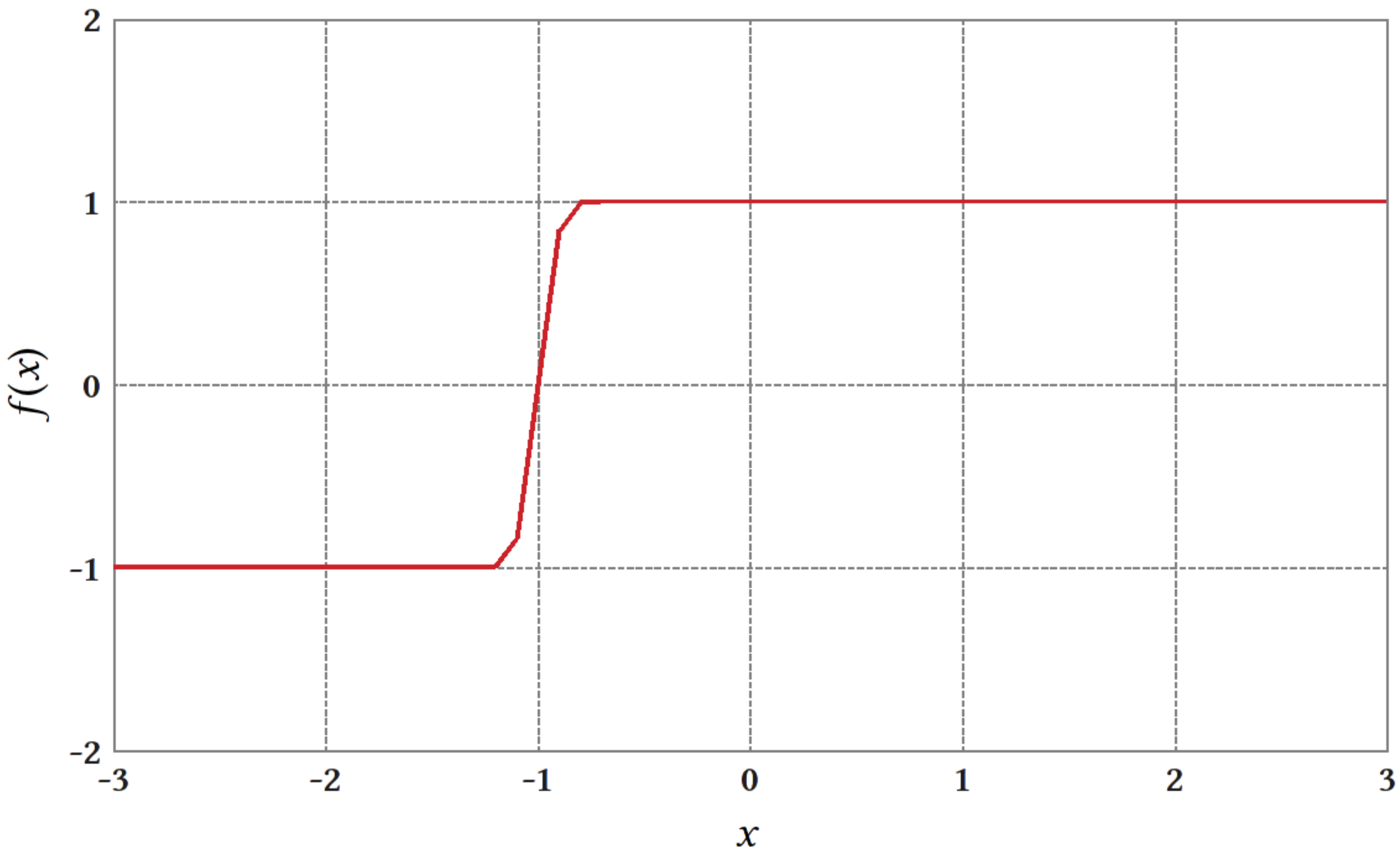
- ★ Error function:
  - ▶ $f(x) = \mathrm{erf}(x)$
- ★ Modifications:
  - ▶ $f(x) = \mathrm{erf}(bx)$



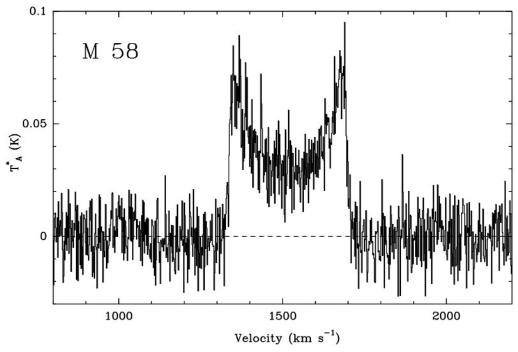


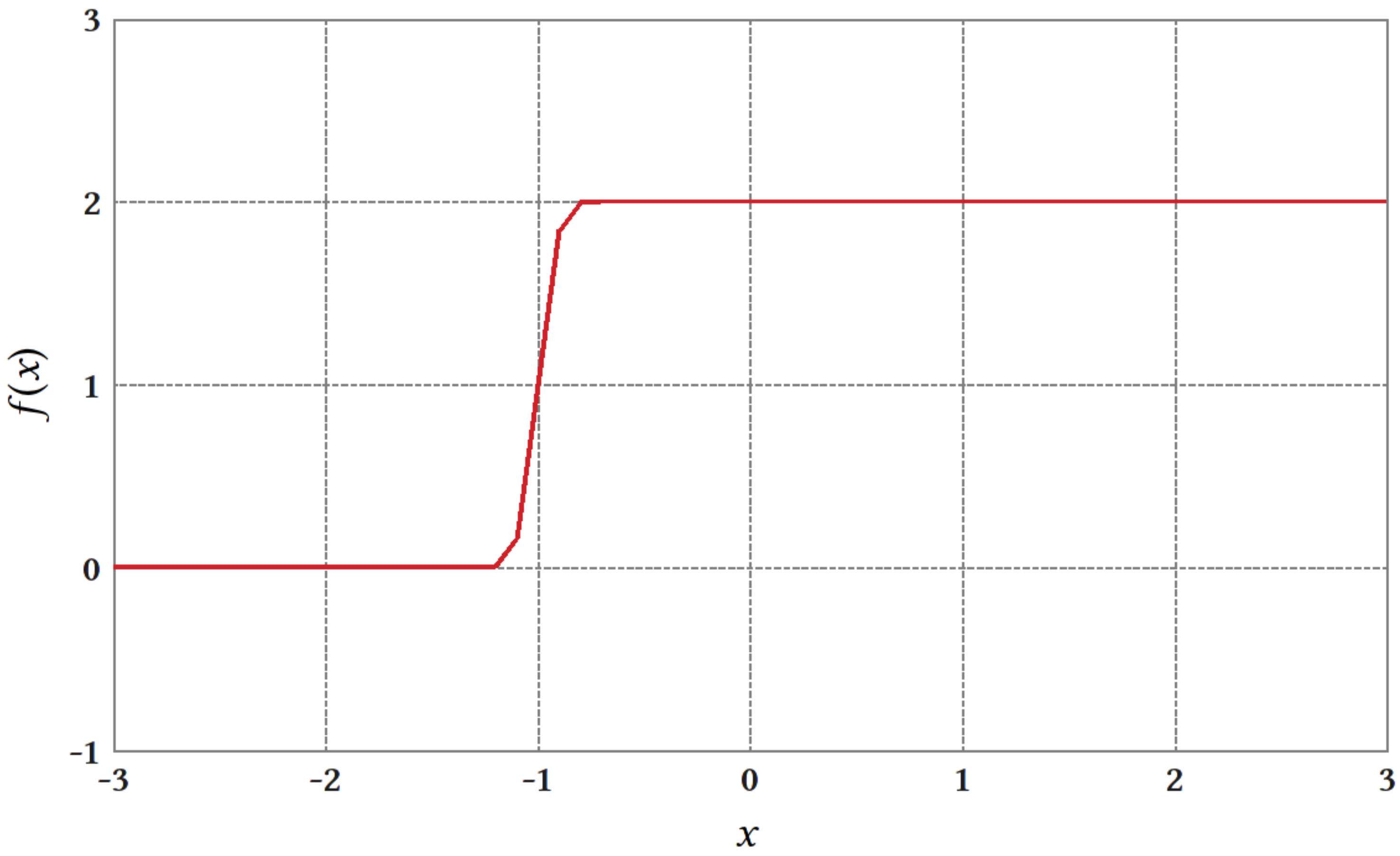
- Error function:
  - $f(x) = \mathrm{erf}(x)$
- Modifications:
  - $f(x) = \mathrm{erf}(bx)$
  - $f(x) = \mathrm{erf}(b[w + x])$



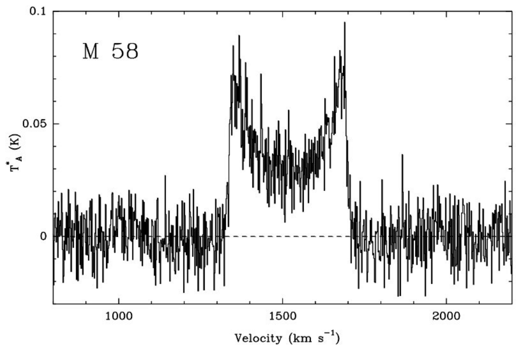


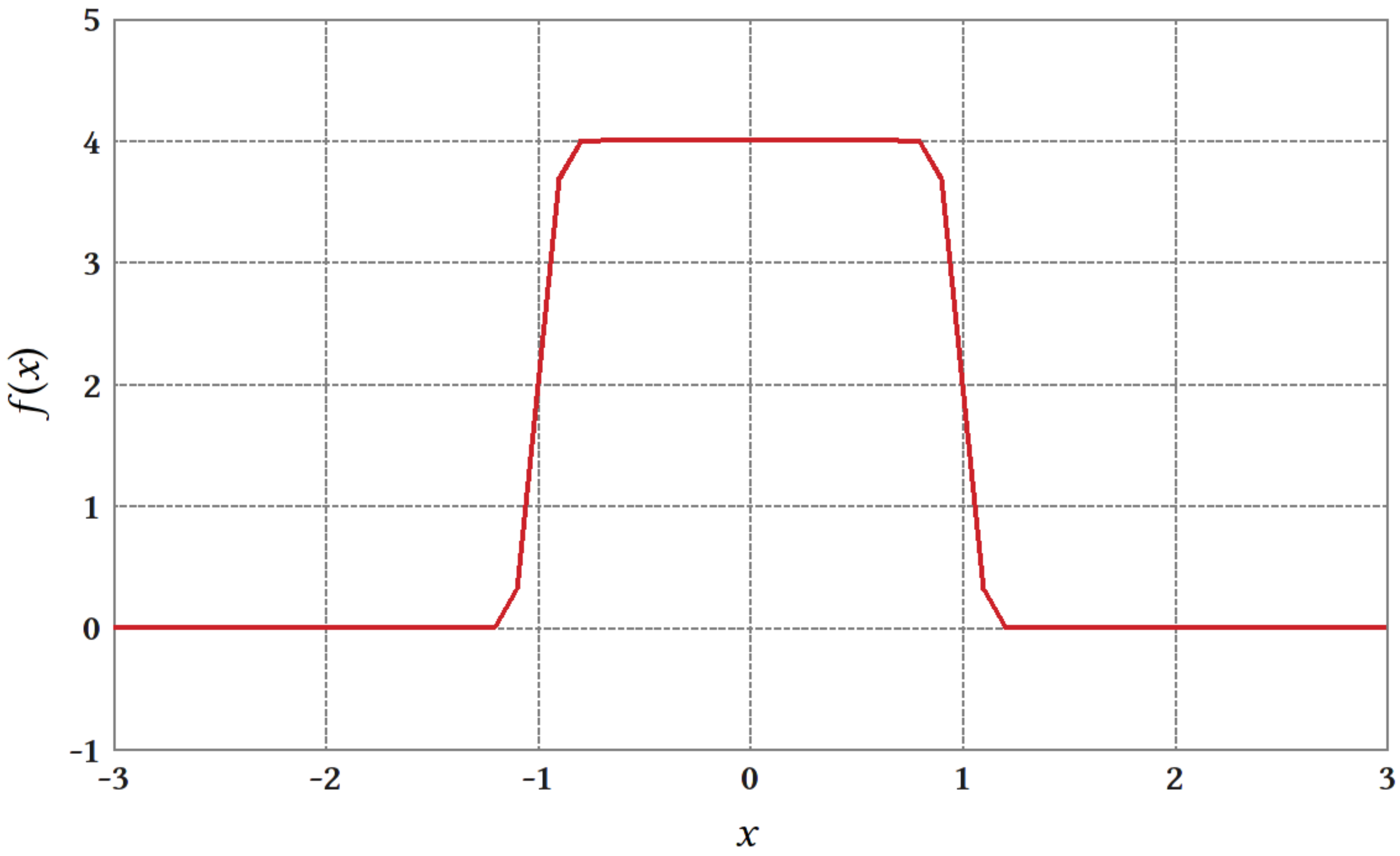
- Error function:
  - $f(x) = \mathrm{erf}(x)$
- Modifications:
  - $f(x) = \mathrm{erf}(bx)$
  - $f(x) = \mathrm{erf}(b[w + x])$
  - $f(x) = \mathrm{erf}(b[w + x]) + 1$



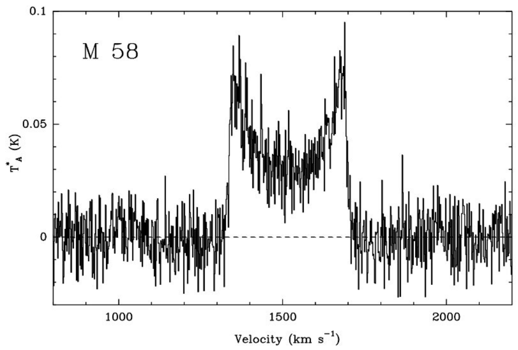


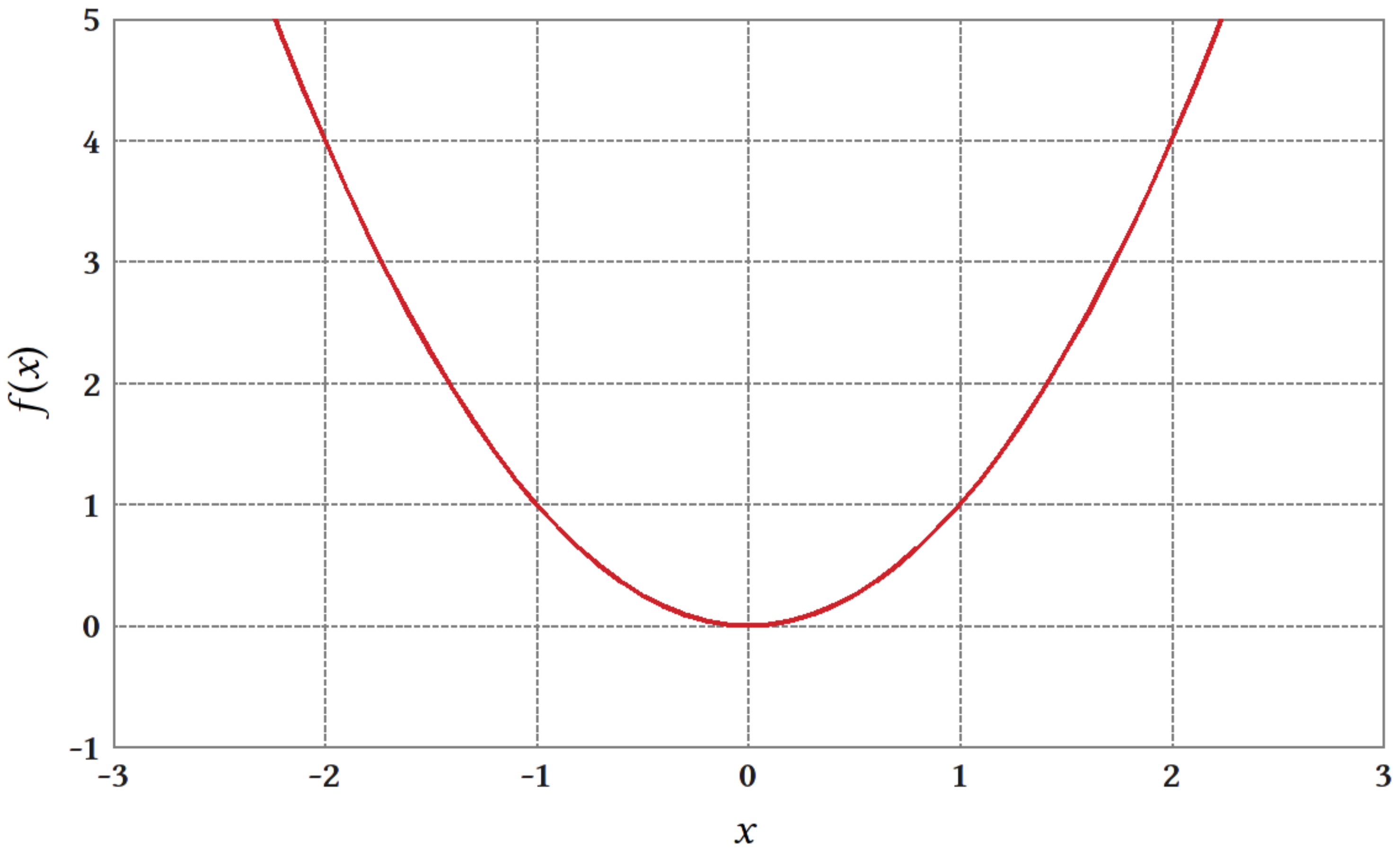
- Error function:
  - $f(x) = \mathrm{erf}(x)$
- Modifications:
  - $f(x) = \mathrm{erf}(bx)$
  - $f(x) = \mathrm{erf}(b[w + x])$
  - $f(x) = \{\mathrm{erf}(b[w + x]) + 1\} \times \{\mathrm{erf}(b[w - x]) + 1\}$



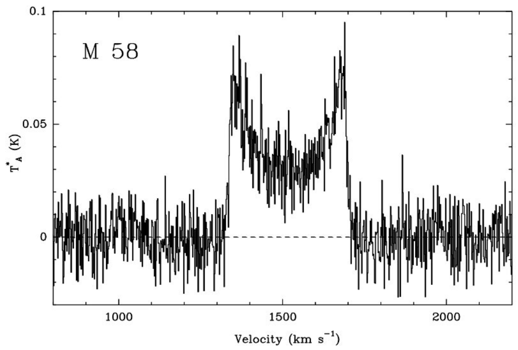


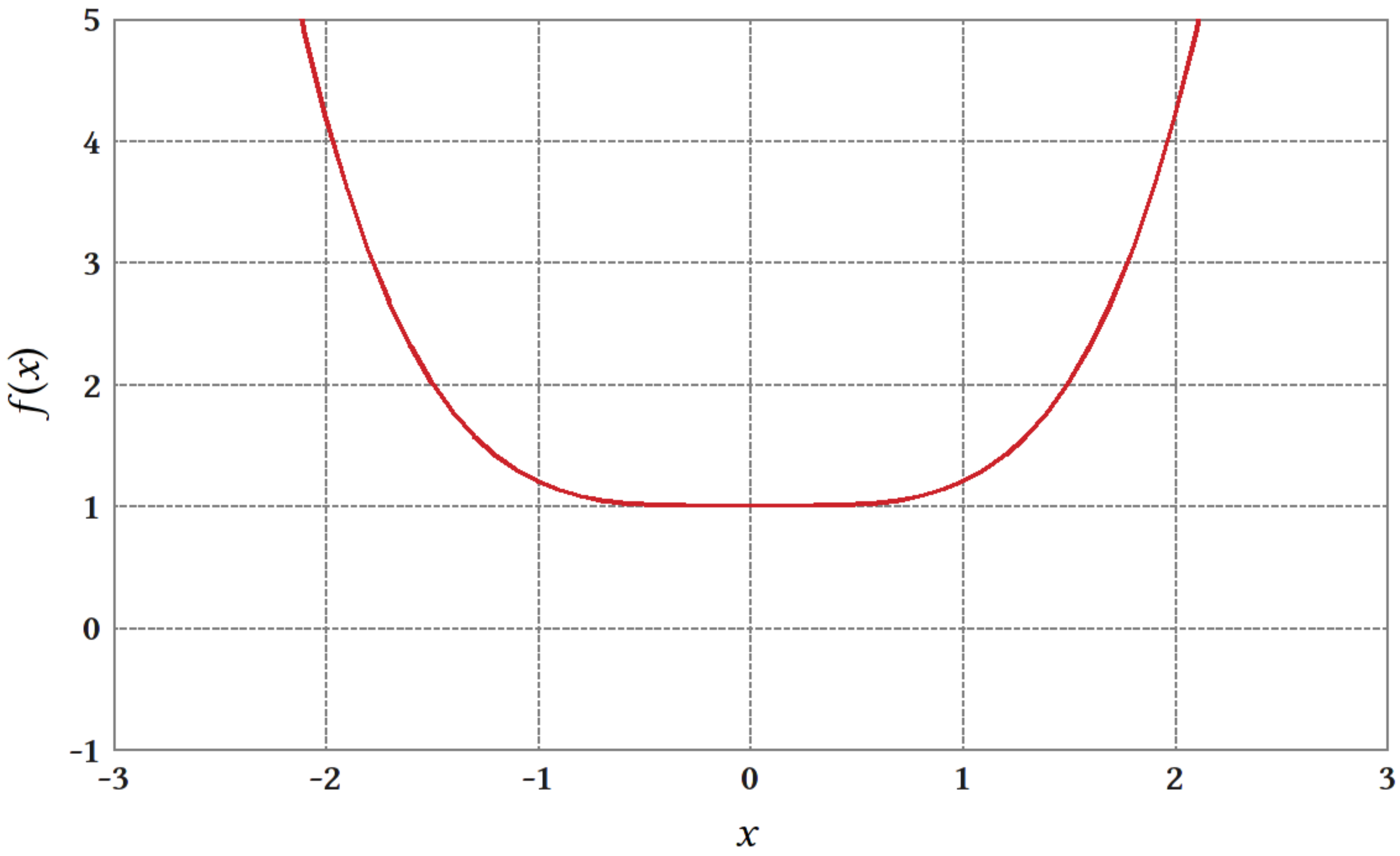
- Polynomial:
  - $f(x) = x^2$



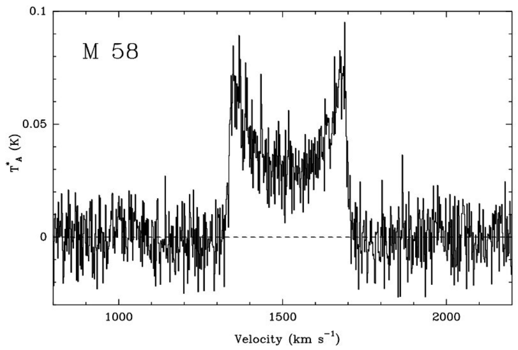


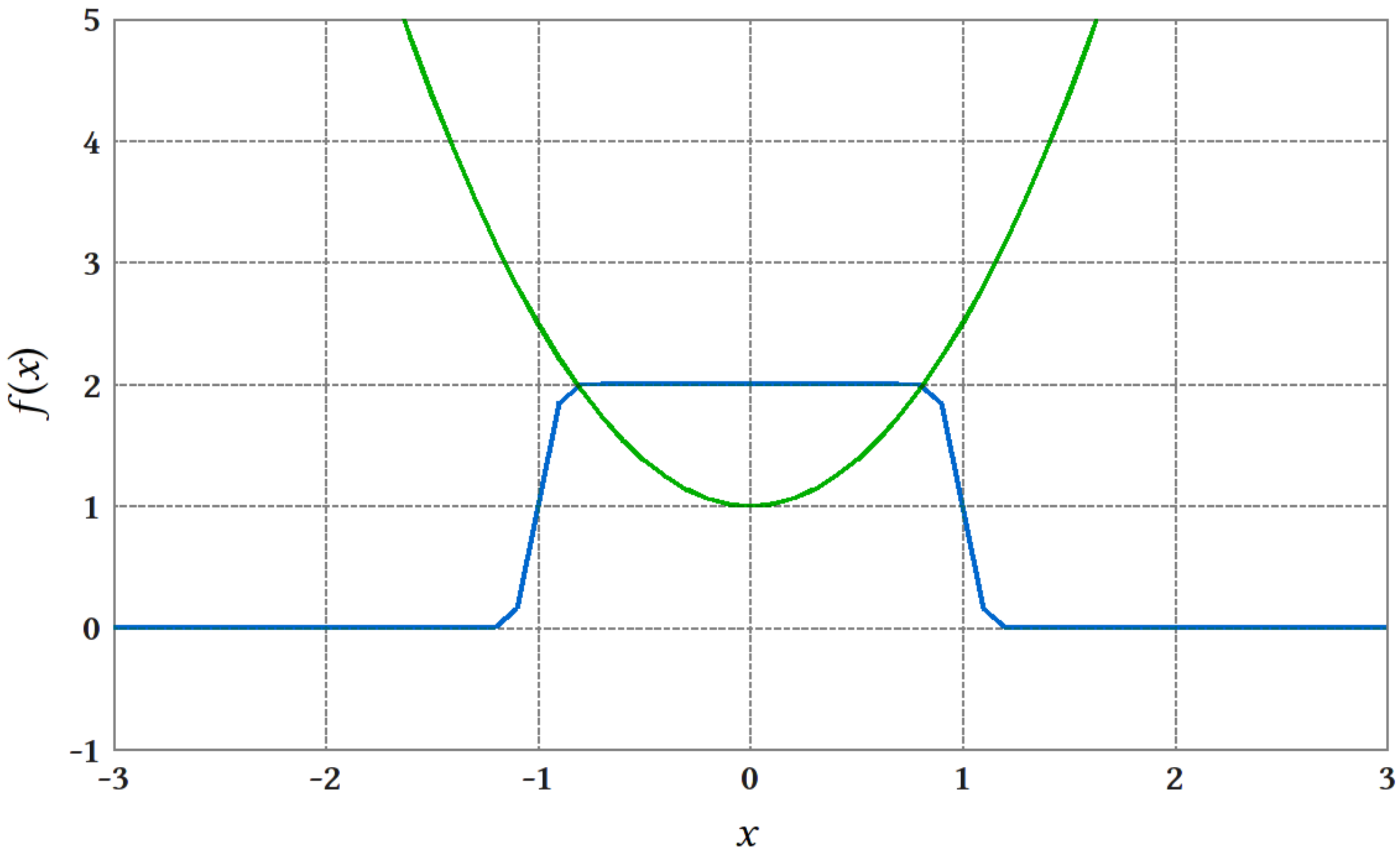
- Polynomial:
  - $f(x) = x^2$
- Modifications:
  - $f(x) = cx^n + d$



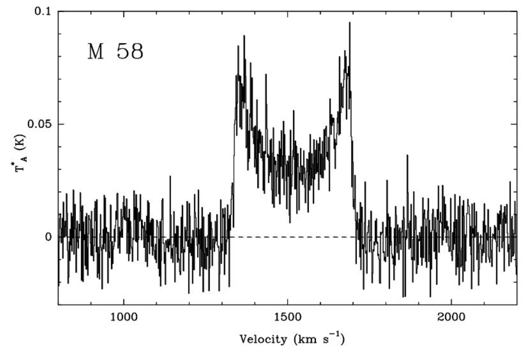


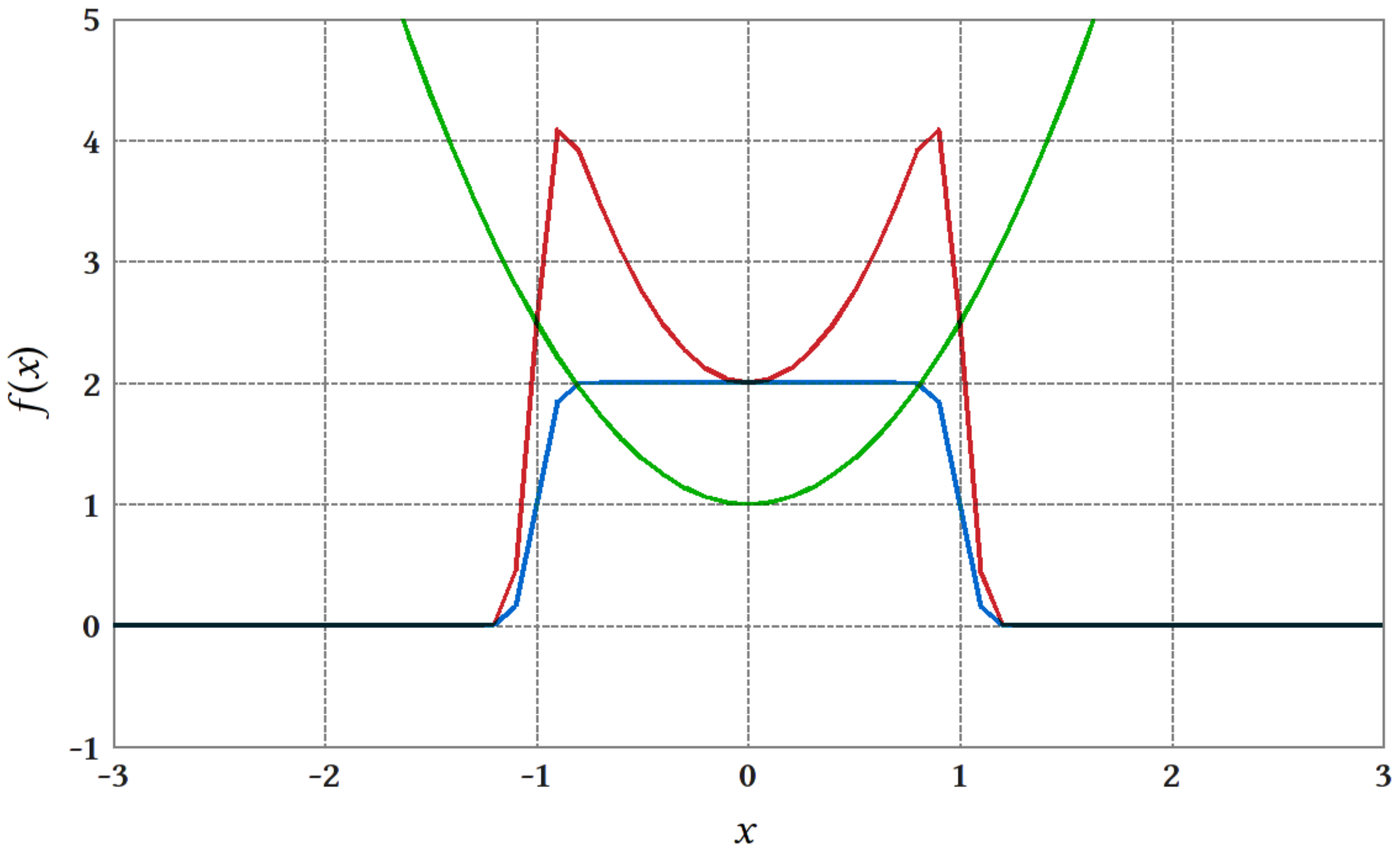
- Error functions:
  - $f(x) = \{\mathrm{erf}(b[w+x]) + 1\} \times \{\mathrm{erf}(b[w-x]) + 1\} \times a$
- Polynomial:
  - $g(x) = cx^n + d$



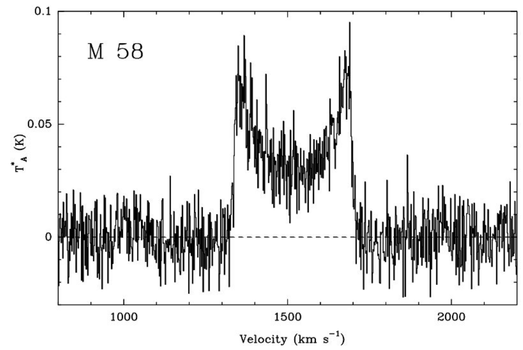


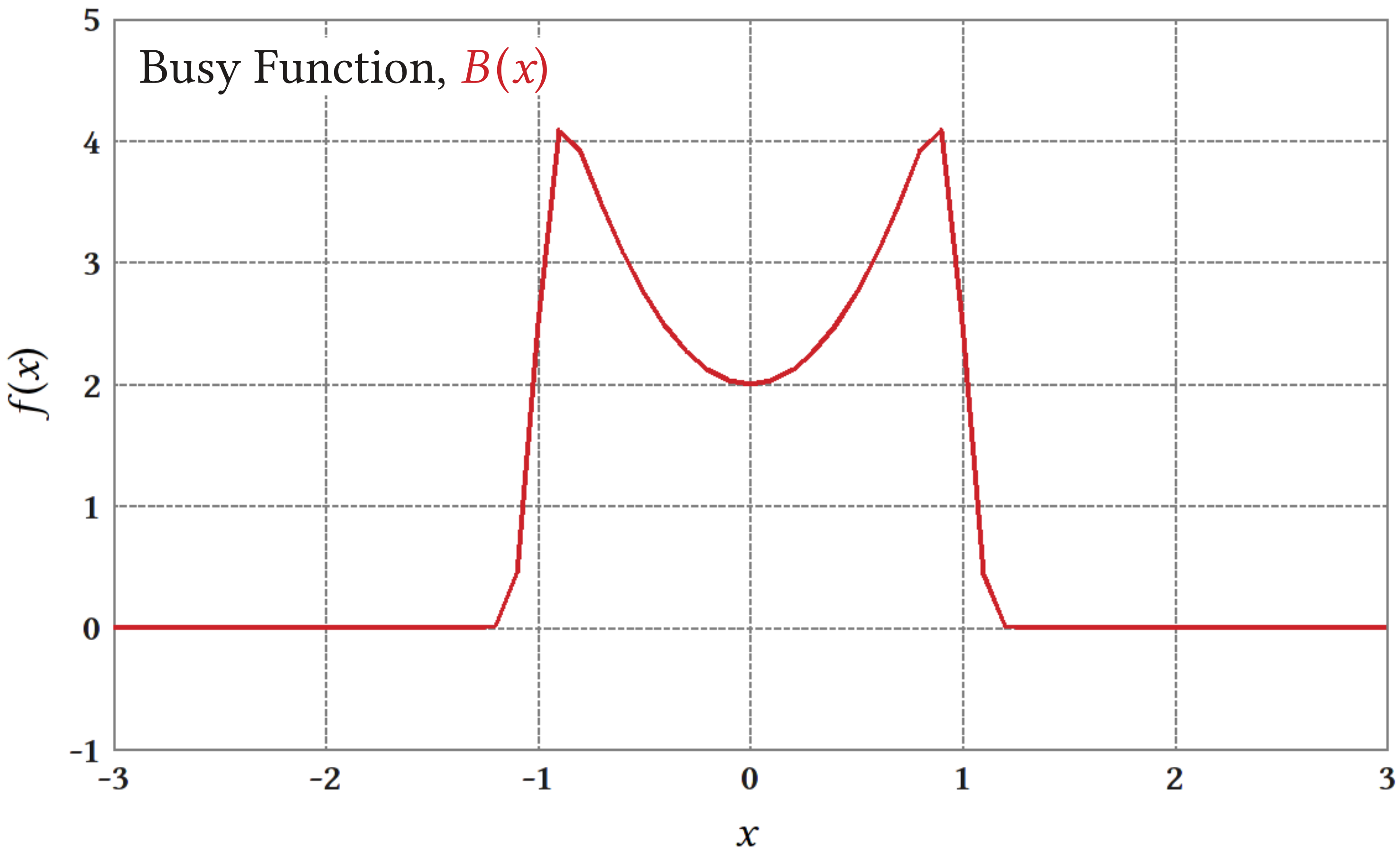
- ★ Error functions:
  - ▶ $f(x) = \{\mathrm{erf}(b[w+x]) + 1\} \times \{\mathrm{erf}(b[w-x]) + 1\} \times a$
- ★ Polynomial:
  - ▶ $g(x) = cx^{n} + d$
- ★ Combination:
  - ▶ $B(x) = f(x) \times g(x)$



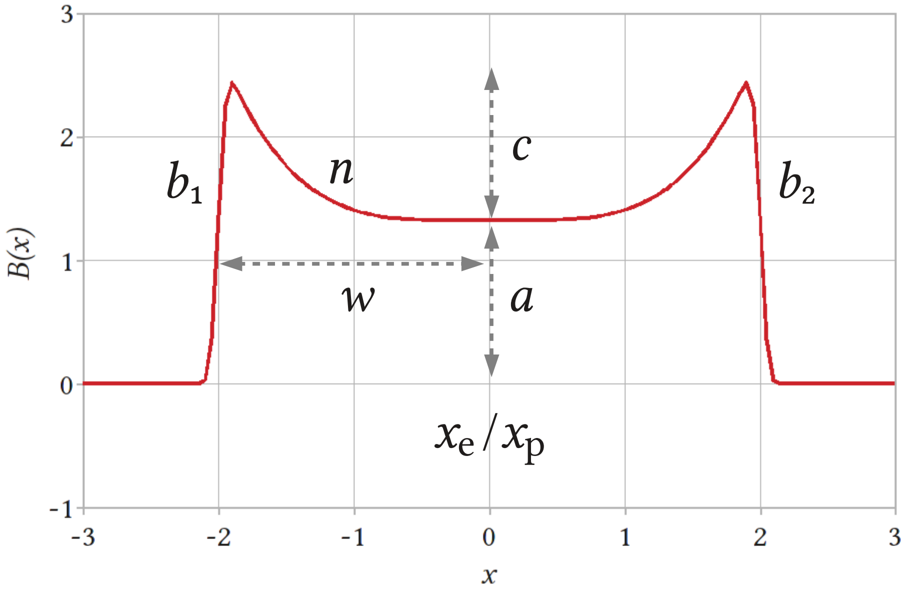


- Error functions:
  - $f(x) = \{\mathrm{erf}(b[w+x]) + 1\} \times \{\mathrm{erf}(b[w-x]) + 1\} \times a$
- Polynomial:
  - $g(x) = cx^n + d$
- Combination:
  - $B(x) = f(x) \times g(x)$

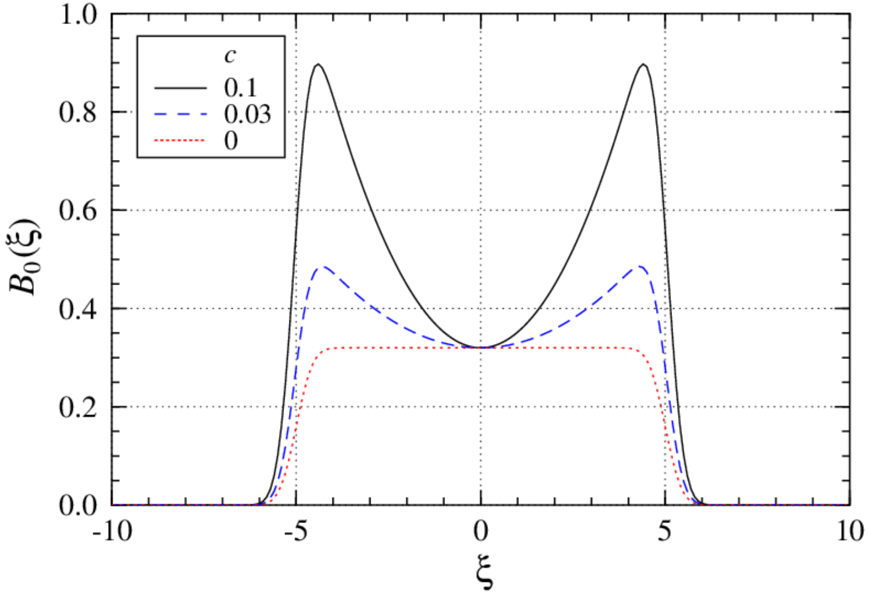


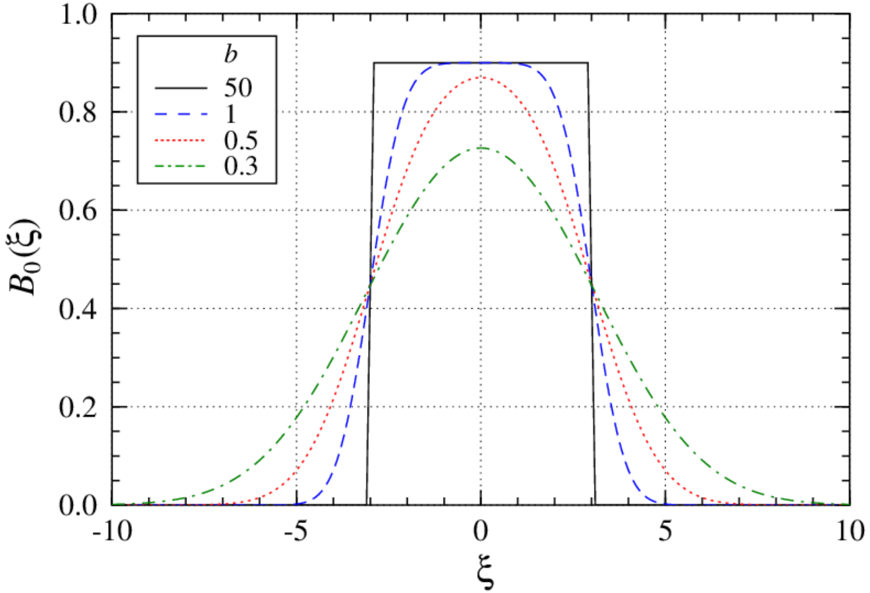


- General form of the Busy Function:
  - $B(x) = [a/4] \times [\mathrm{erf}(b_1\{w + x - x_e\}) + 1] \times [\mathrm{erf}(b_2\{w - x + x_e\}) + 1] \times [c|x - x_p|^n + 1]$
- Maximum of 8 free parameters:
  - a — *scaling factor*
  - b_1, b_2 — *steepness of line flanks*
  - w — *half-width of profile*
  - x_e, x_p — *centroid of error functions / polynomial*
  - c — *scaling factor of polynomial trough*
  - n — *order of polynomial*

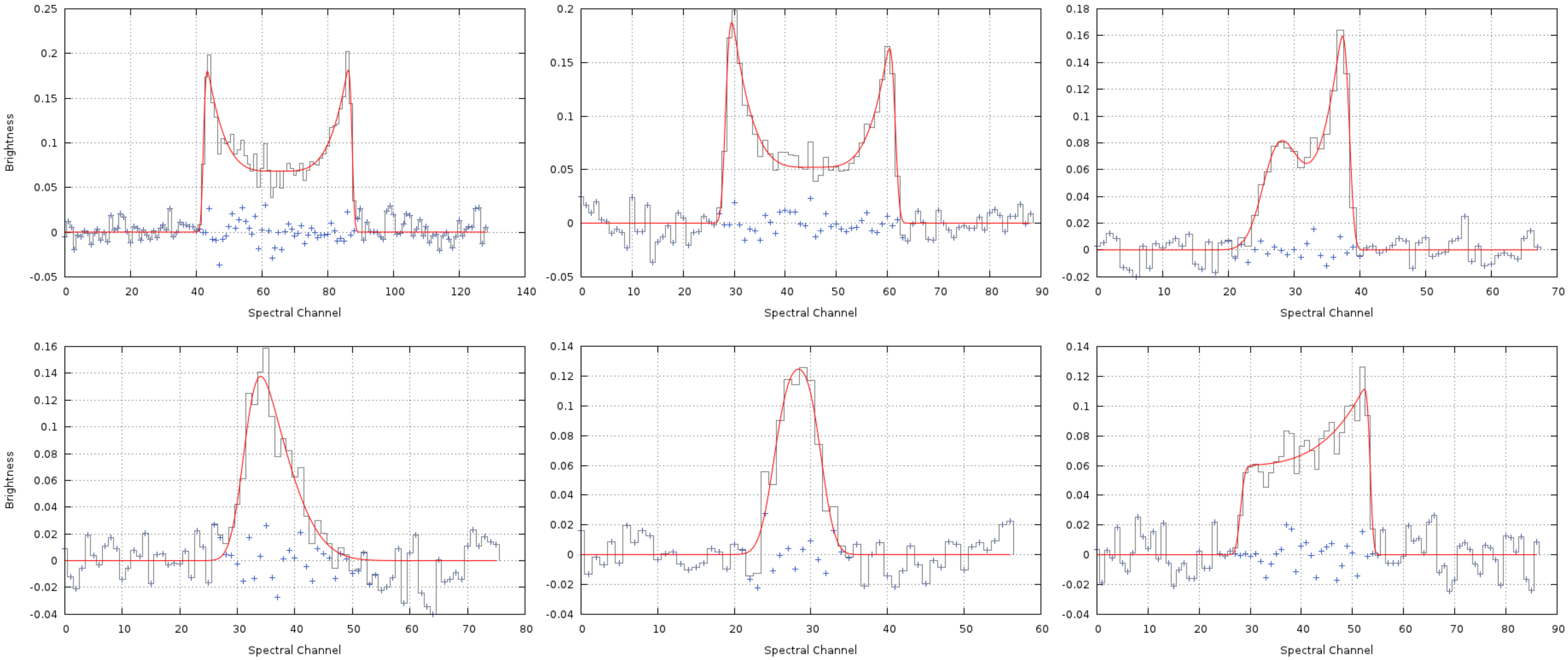


- Properties of the Busy Function:
  - Very versatile; can mimic various shapes:
    - Double-peaked
    - Flat-topped
    - Gaussian





The Busy Function

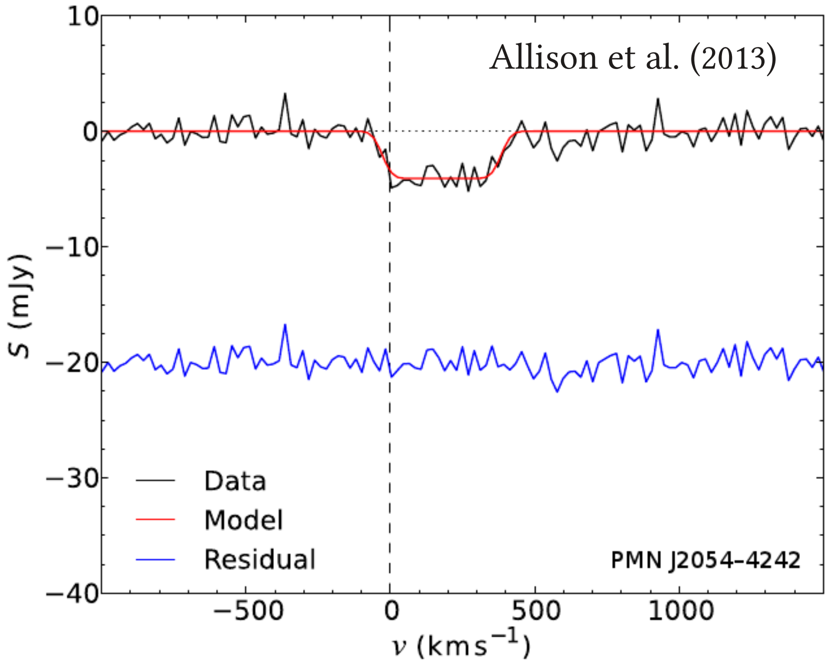


Applications of the Busy Function

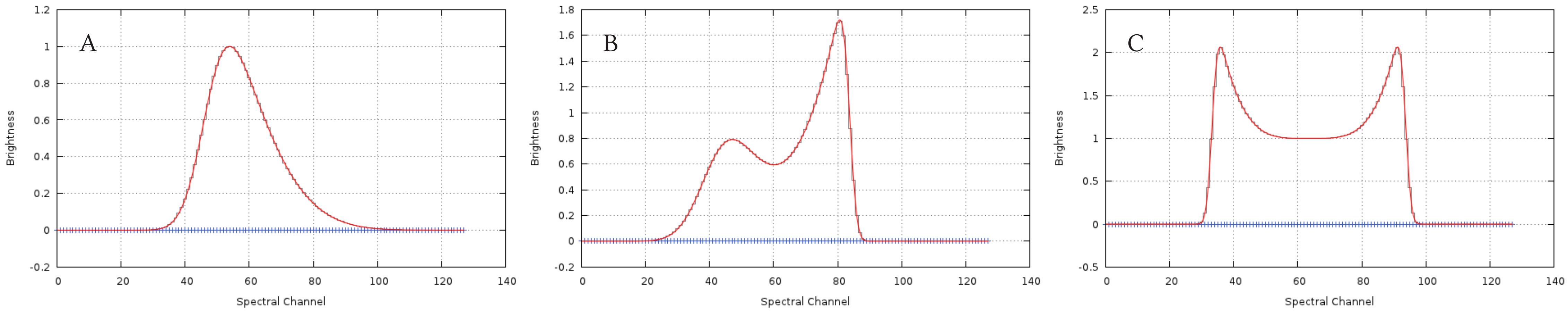
Applications of the Busy Function

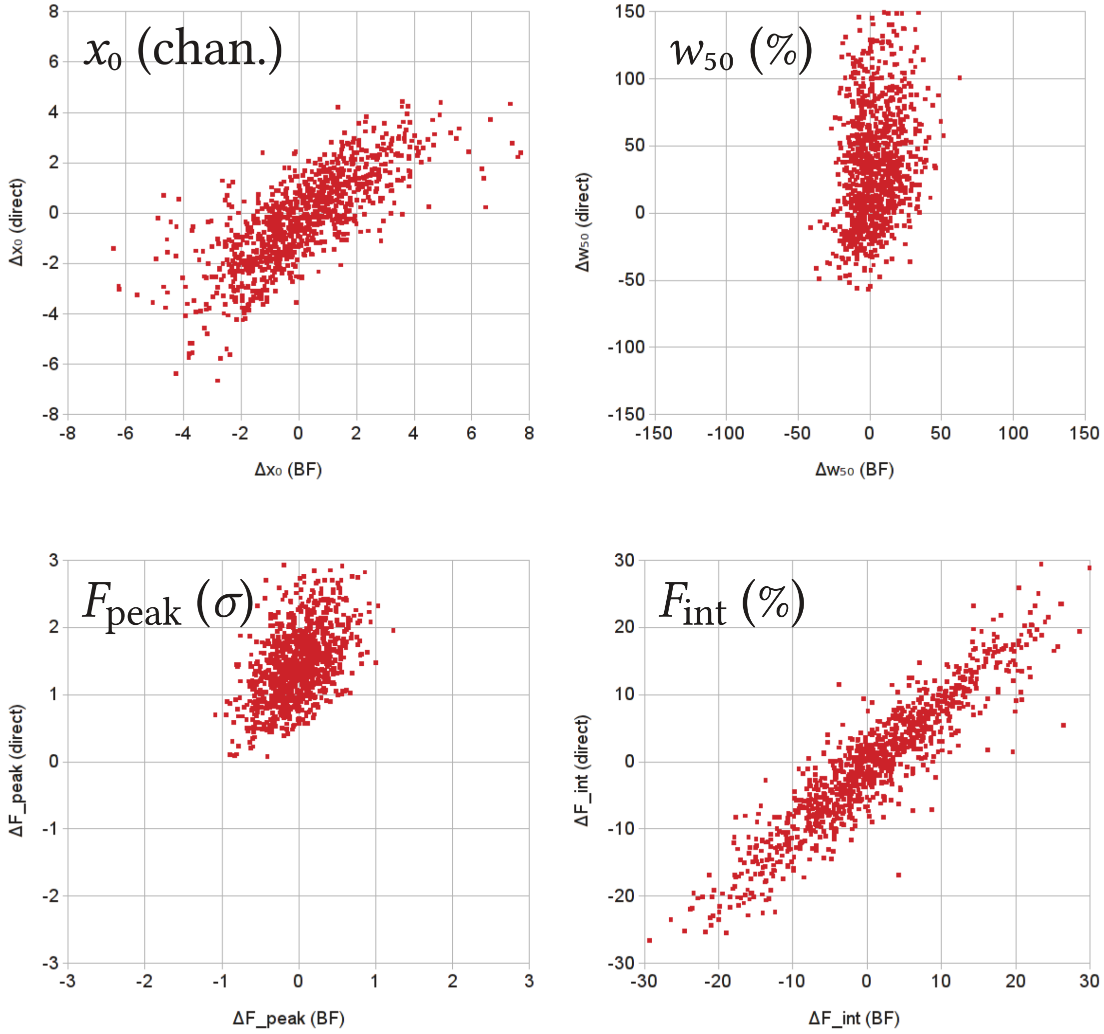
- Parametrisation of galaxies in large surveys
  - Advantage of fitting over ‘direct’ parametrisation:
    - Less sensitive to noise peaks $\longrightarrow$ more accurate for spectra of low S/N.
    - Covariance matrix $\longrightarrow$ calculation of statistical uncertainties.
    - Decomposition of confused sources possible.

- Parametrisation of galaxies in large surveys
  - Advantage of fitting over 'direct' parametrisation:
    - Less sensitive to noise peaks ⟶ more accurate for spectra of low S/N.
    - Covariance matrix ⟶ calculation of statistical uncertainties.
    - Decomposition of confused sources possible.
- Generation of mock profiles
  - Potential use:
    - Simulations / mock data cubes.
    - Filters for matched filtering techniques in source finding.
- Applications beyond H I emission
  - H I stacking and absorption
  - CO spectra

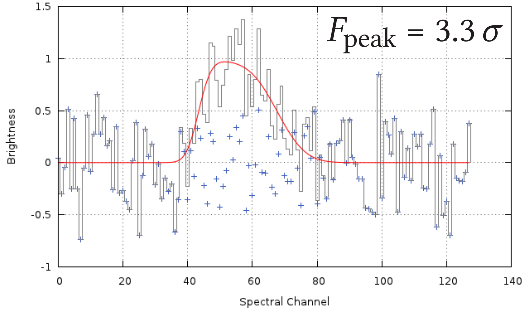


- Systematic testing on model spectra
  - Generation via Busy Function models.
  - Three scenarios:
    - A: single, asymmetric profile
    - B: asymmetric double-horn profile
    - C: broad, symmetric double-horn profile
  - 1 000 iterations with Gaussian noise added.
    - Galaxy parameters (w_{50}, w_{20}, F_{int}, etc.) derived numerically from BF parameters.



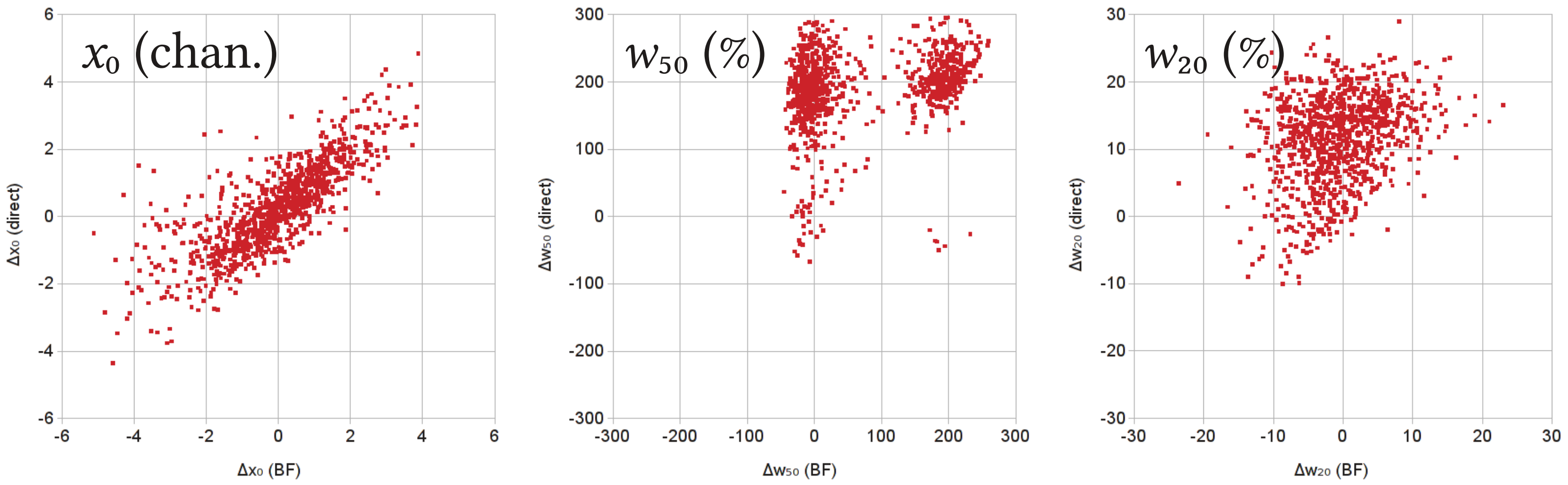


★ A: single, asymmetric profile

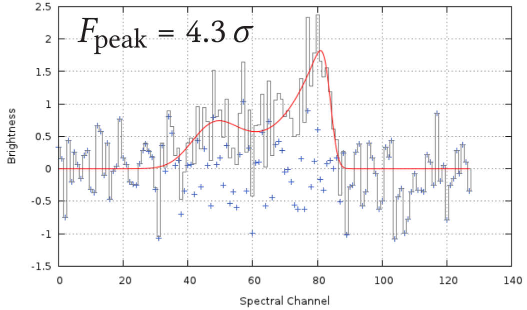


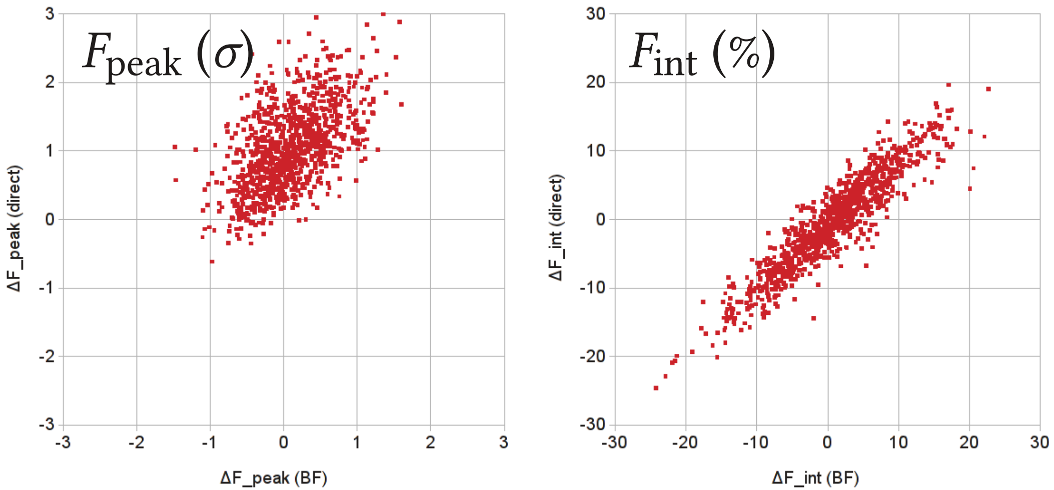
		Busy Function fitting		Direct parametrisation	
		P_0	σ	P_0	σ
x_0	(chan.)	-0.060	1.817	-0.193	1.798
w_{50}	(%)	1.508	14.453	26.331	43.365
w_{20}	(%)	-2.240	11.650	50.209	11.607
F_peak	(σ)	-0.014	0.337	1.393	0.506
F_int	(%)	0.319	9.279	-0.480	9.265

Applications of the Busy Function



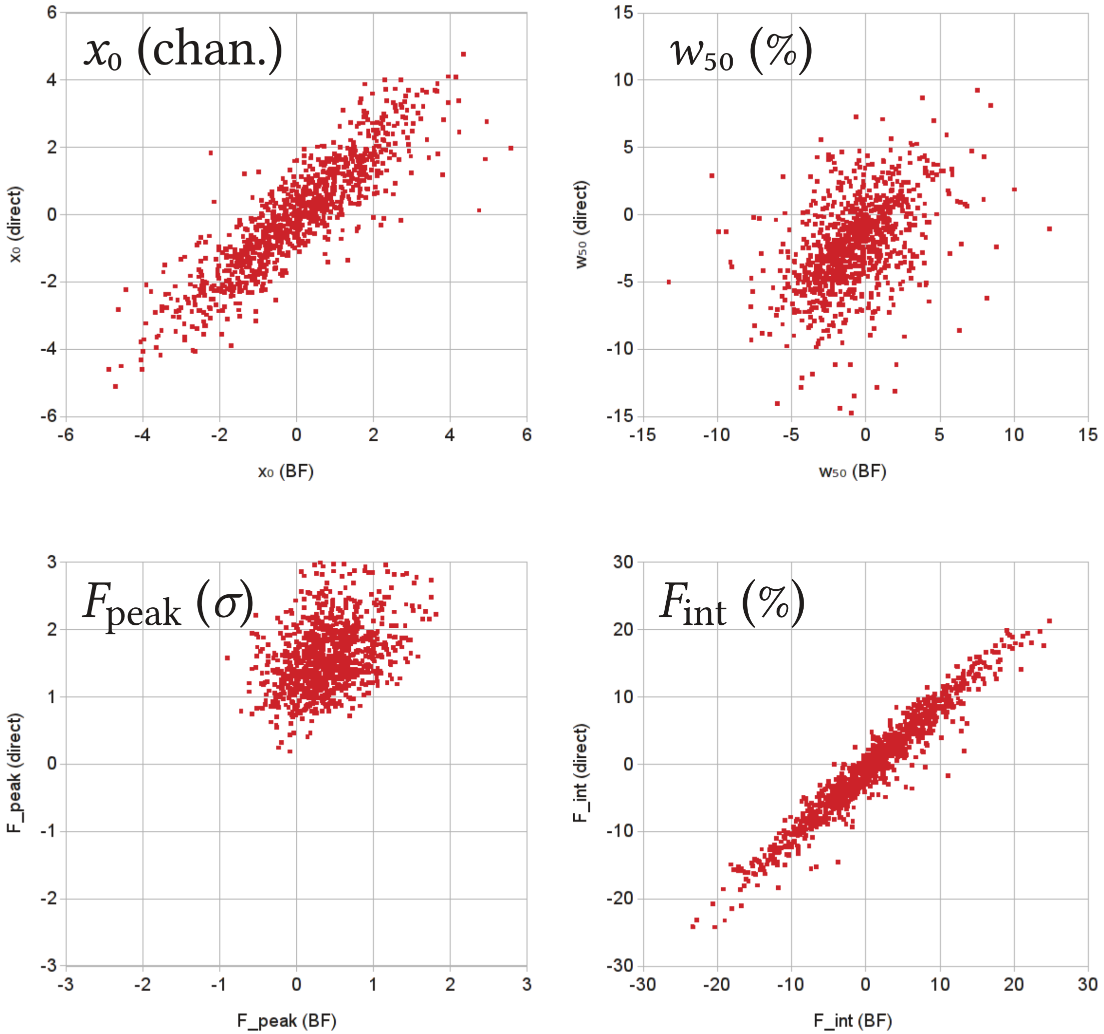
★ B: asymmetric double-horn profile



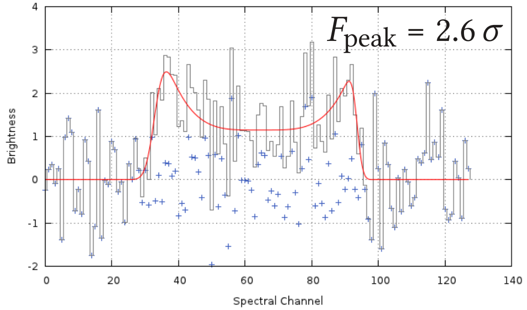


		Busy Function fitting P_0	Busy Function fitting σ	Direct parametrisation P_0	Direct parametrisation σ
x_0	(chan.)	0.029	1.382	0.198	1.286
w_{50}	(%)	-7.662	17.323	199.198	35.196
w_{20}	(%)	-1.000	5.789	12.668	6.025
F_peak	(σ)	0.061	0.487	0.988	0.598
F_int	(%)	0.939	6.946	-0.132	6.950

Applications of the Busy Function

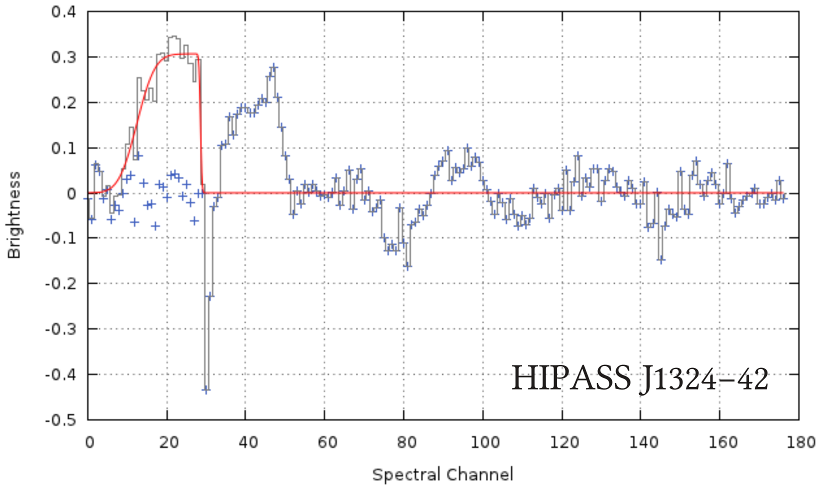


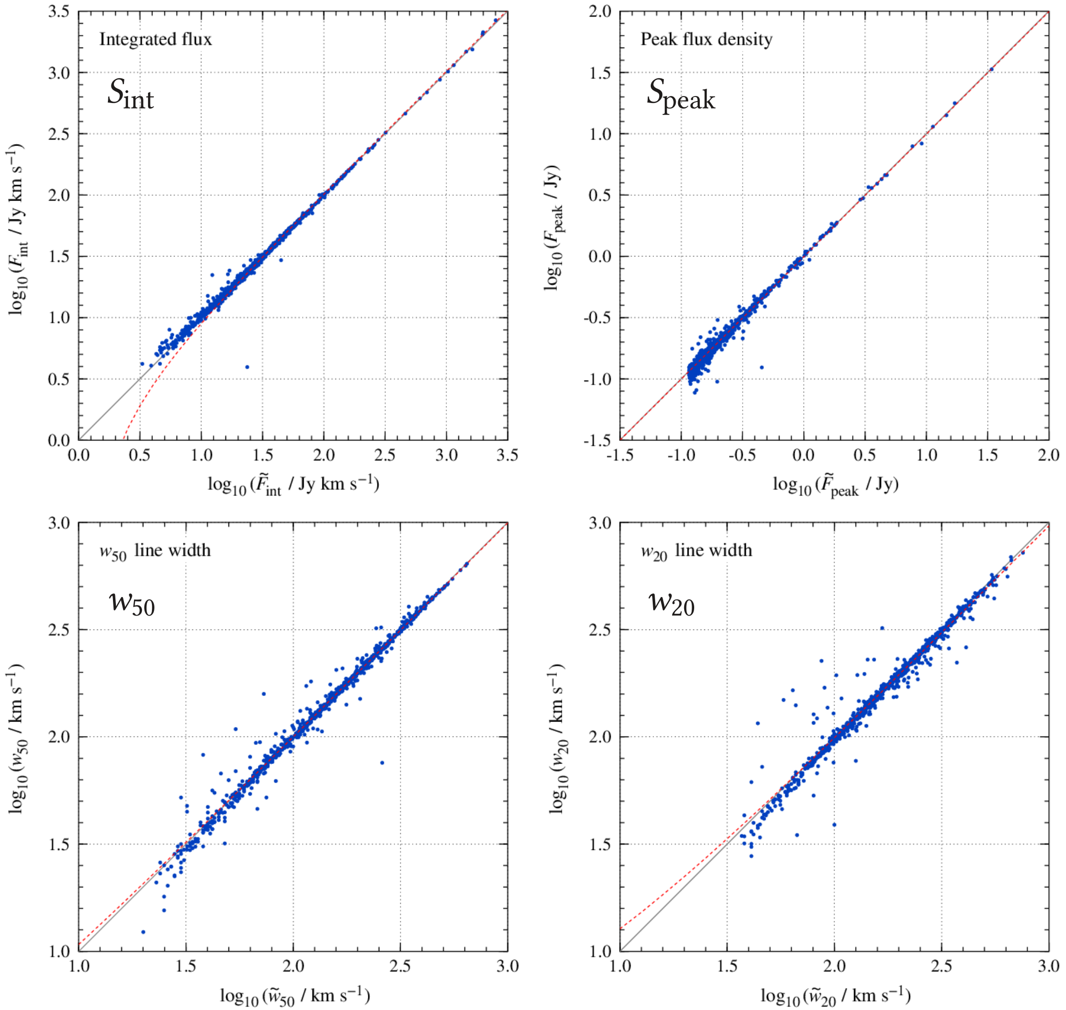
- ★ C: broad, symmetric double-horn profile



		Busy Function fitting		Direct parametrisation	
		P_0	σ	P_0	σ
x_0	(chan.)	0.027	1.607	0.013	1.643
w_{50}	(%)	-0.950	2.312	-2.294	3.046
w_{20}	(%)	-1.115	2.989	0.633	2.841
F_peak	(σ)	0.365	0.391	1.531	0.468
F_int	(%)	1.152	7.866	0.194	7.957

- Example: HIPASS Bright Galaxy Catalogue (Koribalski et al. 2004)
  - Automated fitting of all 1000 BGC sources.
  - Basic parametrisation based on fit.
    - Parameters: v_0, w_{50}, w_{20}, S_{peak}, S_{int}.





Applications of the Busy Function

★ Example: HIPASS Bright Galaxy Catalogue (Koribalski et al. 2004)

- Strategies:
  - "Direct" vs. Busy Function parametrisation.
  - Different S/N: original, 5, and 3.

★ Conclusions:

- Generally accurate recovery of galaxy parameters from Busy Function fit.
- Busy Function much more accurate than 'direct' method for low S/N.
- No difference for bright sources (⟶ special case of peak flux).
- No difference for integrated flux.

Line width (w_{20})

	original		$S/N = 5$		$S/N = 3$	
	direct	BF fit	direct	BF fit	direct	BF fit
best-fit slope	0.96	0.96	0.90	0.87	0.89	0.80
best-fit intercept ($\mathrm{km\,s^{-1}}$)	8.7	5.6	62.2	39.4	94.9	83.7
within 5% of cat. (%)	94.0	86.0	34.6	47.7	20.7	26.5
within 10% of cat. (%)	94.6	93.2	53.0	68.1	36.8	45.4
within 25% of cat. (%)	96.7	97.9	72.2	82.0	56.7	65.0

Integrated flux (F_{int})

	original		$S/N = 5$		$S/N = 3$	
	direct	BF fit	direct	BF fit	direct	BF fit
best-fit slope	1.02	1.03	0.96	1.05	0.92	1.07
best-fit intercept ($\mathrm{Jy\,km\,s^{-1}}$)	−1.08	−1.35	2.01	−1.02	4.49	0.14
within 5% of cat. (%)	89.8	83.3	35.8	37.1	22.3	23.2
within 10% of cat. (%)	96.9	95.5	65.6	66.9	42.5	43.2
within 25% of cat. (%)	98.9	99.1	93.2	91.4	77.4	76.4

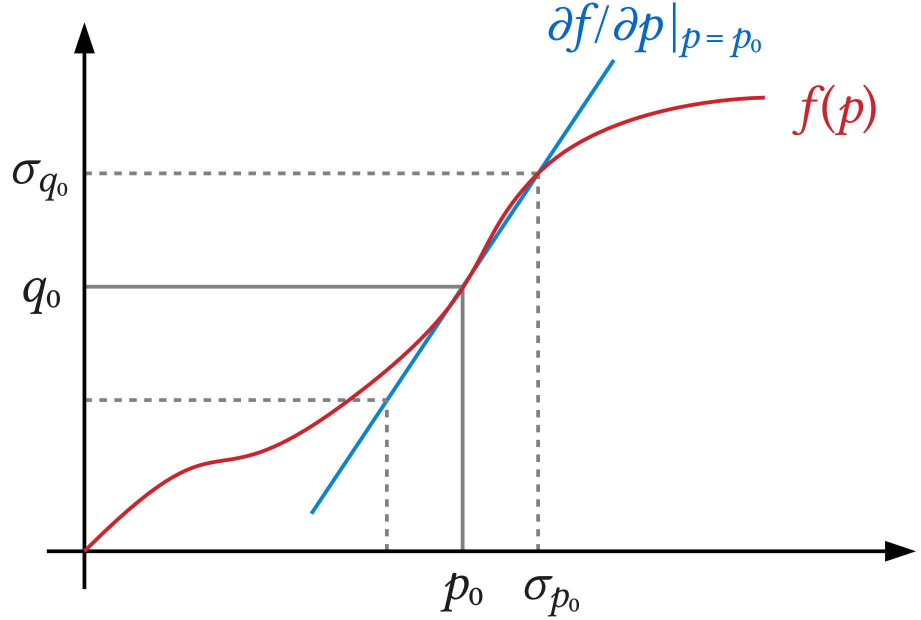
Peak flux density (F_{peak})

	original		$S/N = 5$		$S/N = 3$	
	direct	BF fit	direct	BF fit	direct	BF fit
best-fit slope	1.00	0.99	1.14	1.04	1.27	1.09
best-fit intercept (Jy)	0.00	0.00	−0.01	−0.01	0.01	0.00
within 5% of cat. (%)	99.0	68.6	21.9	28.3	5.4	18.1
within 10% of cat. (%)	99.2	88.0	43.3	54.5	14.7	32.8
within 25% of cat. (%)	99.7	99.1	86.3	91.9	47.1	72.2

Calculation of Uncertainties

★ Linear propagation of covariance matrix

- Assumptions:
  - Parameter vector, $\vec{p}$, and covariance matrix, $\mathrm{C}_{\vec{p}}$, from Busy Function fit.
  - Derived parameters given by differentiable function, $\vec{q} = \vec{f}(\vec{p})$.
- Jacobian matrix of $\vec{f}(\vec{p})$:
  - $\mathcal{J}_{ij} \equiv \partial f_j / \partial p_i \approx [f_j(\vec{p} + \varepsilon_i \hat{e}_i) - f_j(\vec{p})] / \varepsilon_i$
- Error propagation law:
  - $\mathrm{C}_{\vec{q}} = \mathbf{J} \cdot \mathrm{C}_{\vec{p}} \cdot \mathbf{J}^{\mathrm{T}}$
- Advantages:
  - Full covariance matrix "for free".
  - No analytic form of $\vec{f}(\vec{p})$ required.



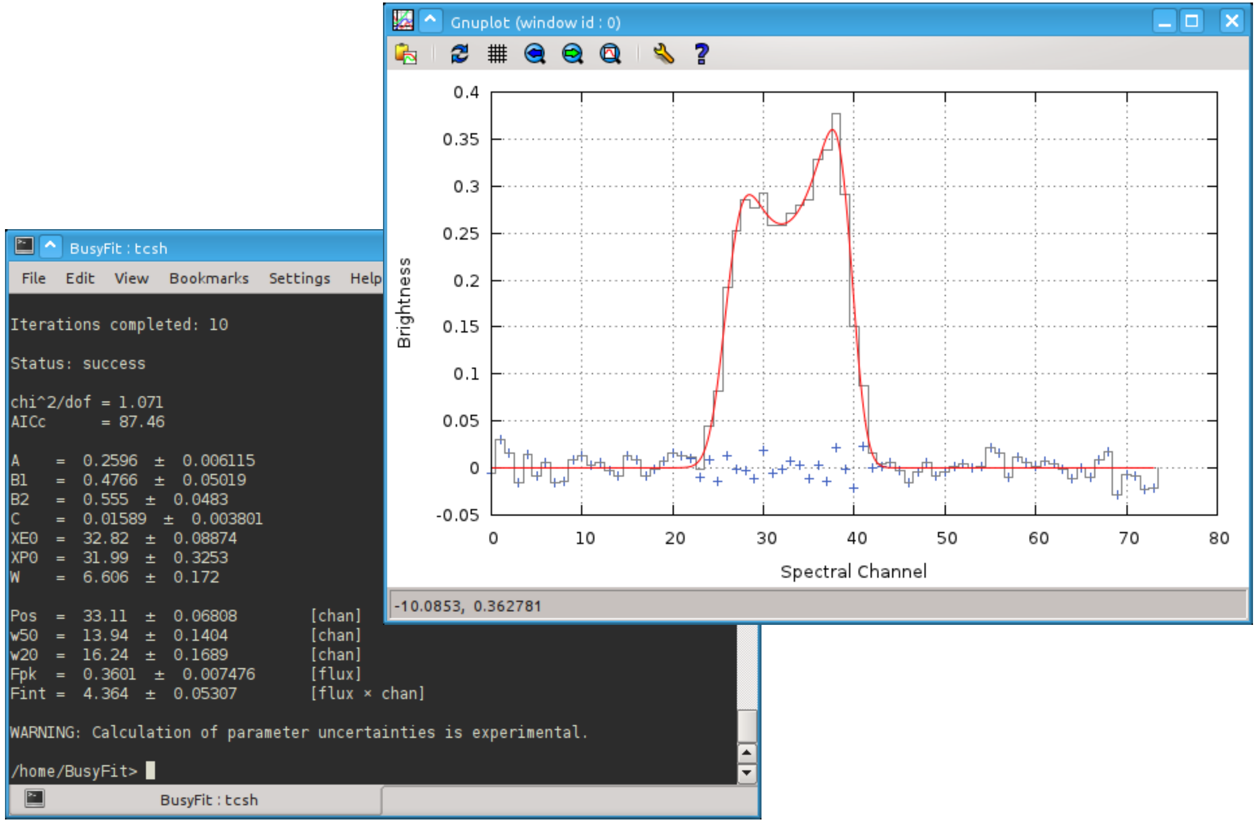
Parameter	Value	Uncert. par. var.	Uncert. err. prop.	Difference (per cent)
Centroid	30.164	0.1216	0.1213	−0.25
w_{50}	51.612	0.08452	0.08648	2.32
w_{20}	53.485	0.1303	0.1245	−4.45
F_{peak}	0.348	0.004646	0.005157	11.00
F_{int}	9.618	0.07491	0.07485	−0.08

Results of test on NGC 3351.

Busy Function Fitting Software

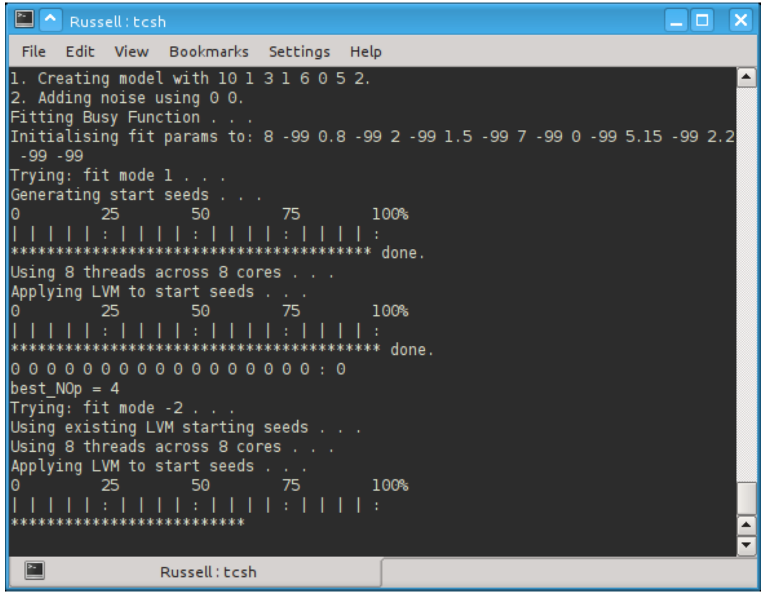
★ BusyFit

- Stand-alone software.
- Requirements:
  - Written in C++ using GSL.
  - Should run on all Linux / Unix systems (e.g. Ubuntu, Mac OS).
- Features:
  - Reads spectrum from text file.
  - Automatic or manual determination of initial parameters.
  - Calculation of uncertainties.
  - Plotting of results using gnuplot.
- Download & instructions:
  - http://www.atnf.csiro.au/people/Tobias.Westmeier/tools_software_busyfit.php

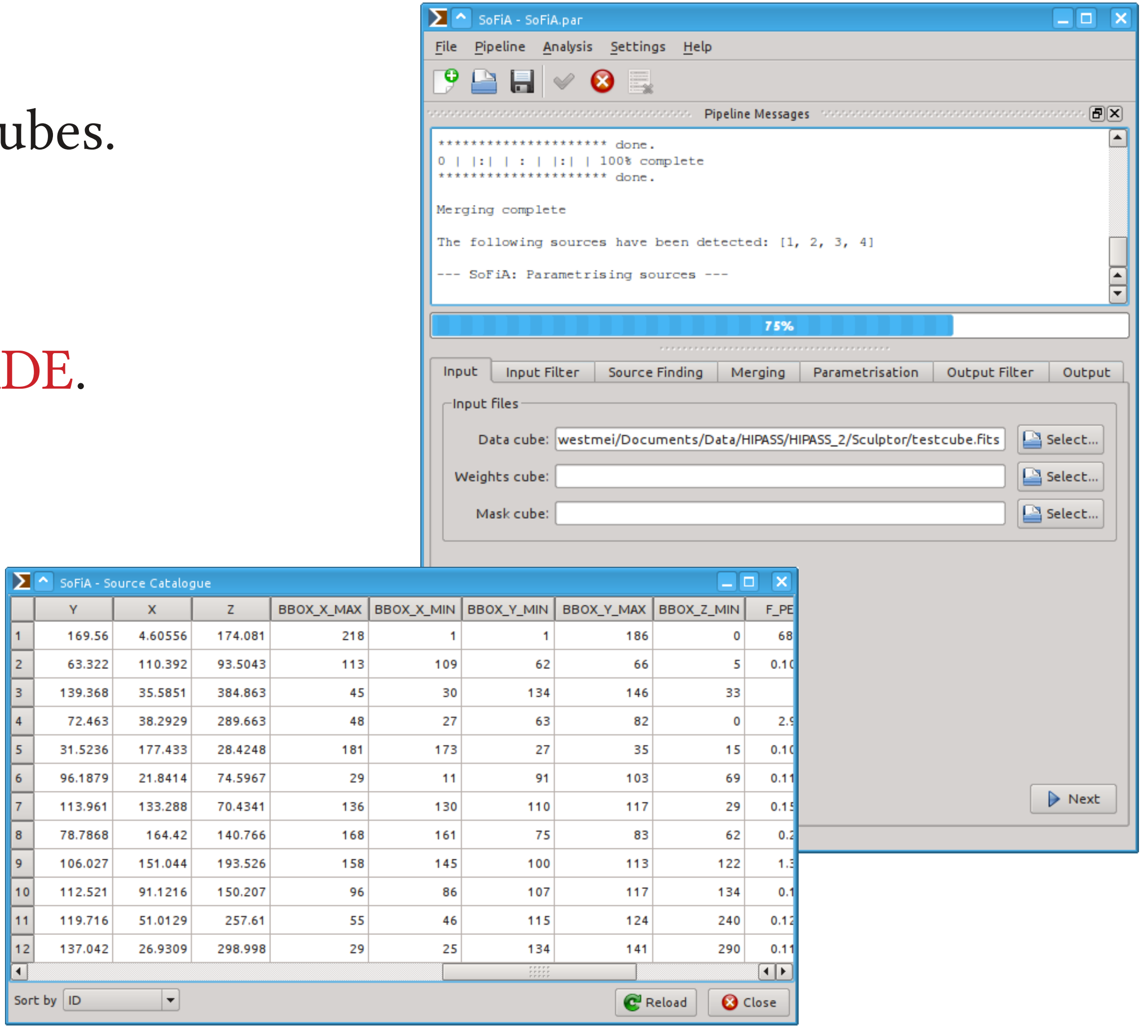


★ BFfit

- Written by Russell Jurek.
- C/C++ library with Python wrapper.
- Requirements:
  - Written in C/C++.
  - Needs cfitsio, cpgplot, and OpenMP.
- Features:
  - Variable number of free parameters based on Akaike Information Criterion.
  - Uses random initial parameters; no initial estimates needed.
  - Makes use of multiple CPU cores.
  - Comes with Python wrapper.
- Download & instructions:
  - http://code.google.com/p/busy-function-fitting/

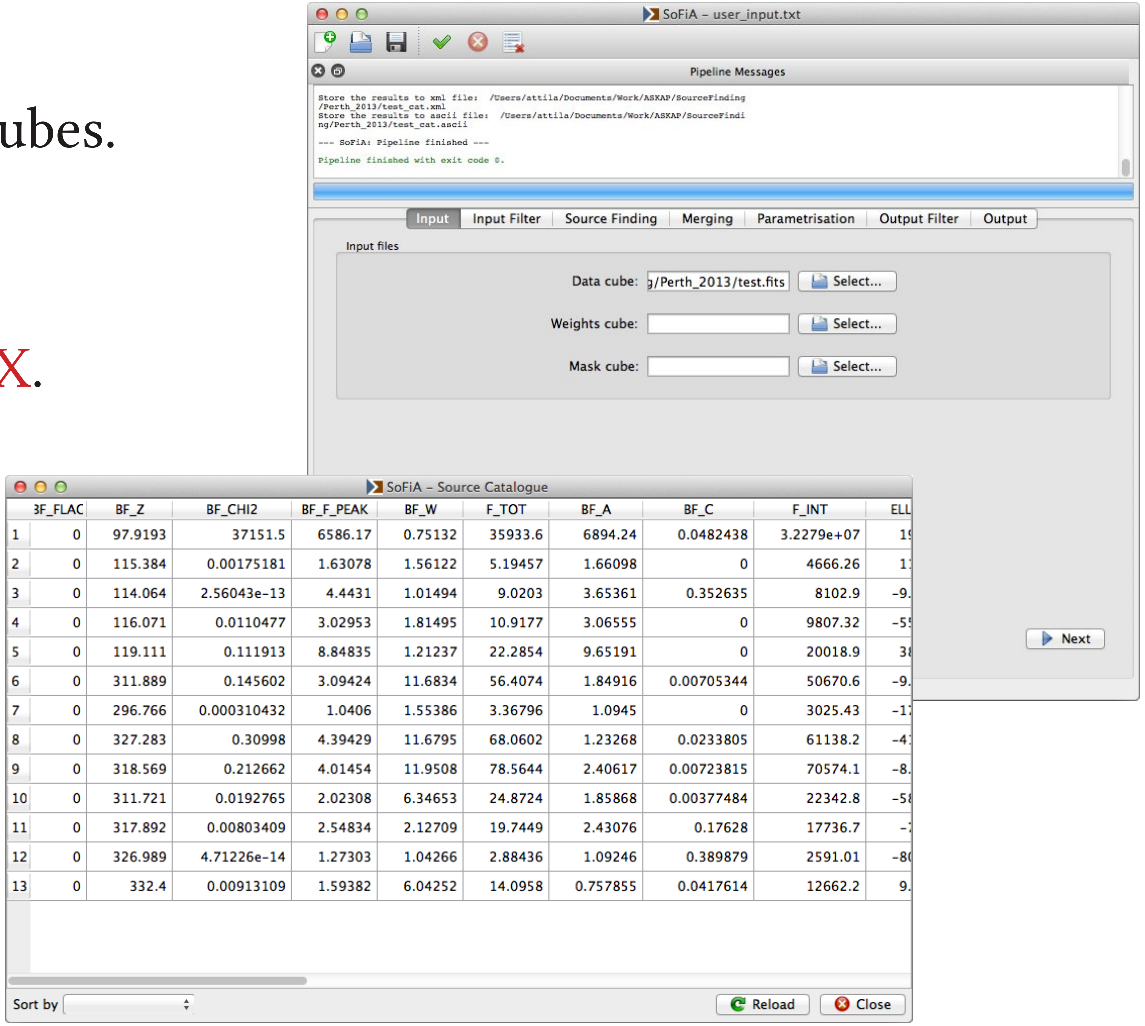


- SoFiA
  - Source Finding Application for H I data cubes.
  - Shiny graphical user interface (GUI).
  - Easy creation of parameter files.
  - Perfect desktop integration, e.g. Linux / KDE.



★ SoFiA

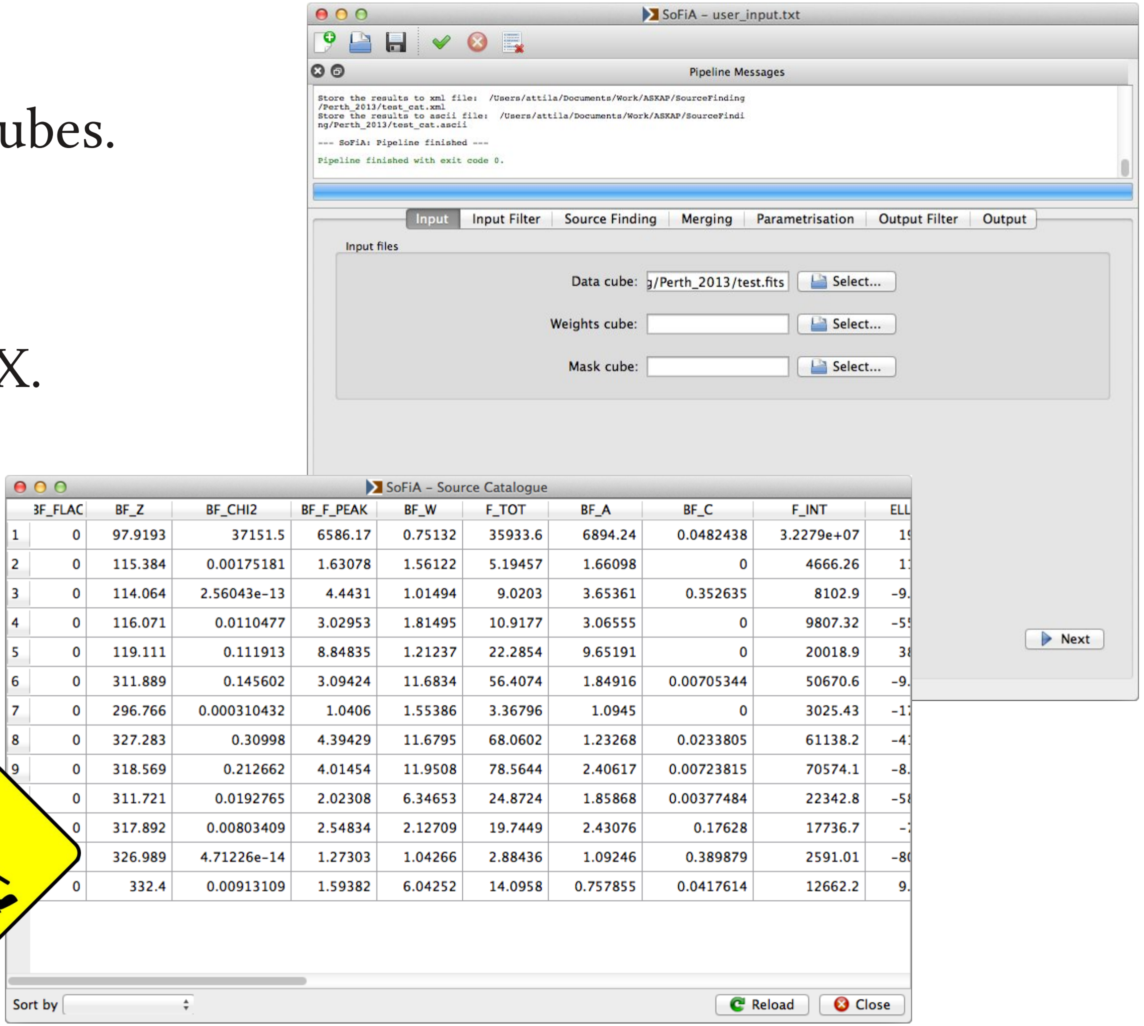
- Source Finding Application for H I data cubes.
- Shiny graphical user interface (GUI).
- Easy creation of parameter files.
- Perfect desktop integration, e.g. Mac OS X.

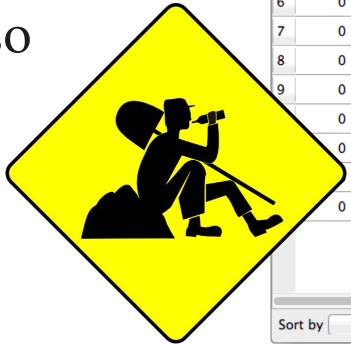


	3F_FLAC	BF_Z	BF_CHI2	BF_F_PEAK	BF_W	F_TOT	BF_A	BF_C	F_INT	ELL
1	0	97.9193	37151.5	6586.17	0.75132	35933.6	6894.24	0.0482438	3.2279e+07	1
2	0	115.384	0.00175181	1.63078	1.56122	5.19457	1.66098	0	4666.26	1
3	0	114.064	2.56043e-13	4.4431	1.01494	9.0203	3.65361	0.352635	8102.9	-9.
4	0	116.071	0.0110477	3.02953	1.81495	10.9177	3.06555	0	9807.32	-5
5	0	119.111	0.111913	8.84835	1.21237	22.2854	9.65191	0	20018.9	3
6	0	311.889	0.145602	3.09424	11.6834	56.4074	1.84916	0.00705344	50670.6	-9.
7	0	296.766	0.000310432	1.0406	1.55386	3.36796	1.0945	0	3025.43	-1
8	0	327.283	0.30998	4.39429	11.6795	68.0602	1.23268	0.0233805	61138.2	-4
9	0	318.569	0.212662	4.01454	11.9508	78.5644	2.40617	0.00723815	70574.1	-8.
10	0	311.721	0.0192765	2.02308	6.34653	24.8724	1.85868	0.00377484	22342.8	-5
11	0	317.892	0.00803409	2.54834	2.12709	19.7449	2.43076	0.17628	17736.7	-
12	0	326.989	4.71226e-14	1.27303	1.04266	2.88436	1.09246	0.389879	2591.01	-8
13	0	332.4	0.00913109	1.59382	6.04252	14.0958	0.757855	0.0417614	12662.2	9.

Busy Function Fitting Software

- ★ SoFiA
  - ▸ Source Finding Application for H I data cubes.
  - ▸ Shiny graphical user interface (GUI).
  - ▸ Easy creation of parameter files.
  - ▸ Perfect desktop integration, e.g. Mac OS X.
- ★ Download SoFiA
  - ▸ Current stable version: 0.2.1
    - https://github.com/SoFiA-Admin/SoFiA
  - ▸ SoFiA *wiki* and *bug reports* also available on GitHub.
  - ▸ Work in progress!





Summary & Conclusions

Summary & Conclusions

- Busy Function:
  - Simple analytic function for describing different galaxy H I profiles.
  - Applications:
    - Automated fitting and parametrisation of galaxies.
    - Profiles for simulations / matched filtering.
  - Testing on models and HIPASS data:
    - Much more accurate than “direct” parametrisation method, in particular for w_{50}, w_{20}, and S_{peak} of faint sources.
  - Uncertainties from covariance matrix.
  - Software:
    - BusyFit (stand-alone software by *T. Westmeier*)
    - BFfit (C / C++ / Python library by *R. Jurek*)
    - SoFiA (source finder, includes Busy Function fitting)

The busy function: a new analytic function for describing the integrated 21-cm spectral profile of galaxies

T. Westmeier,[1]★ R. Jurek,[2]★ D. Obreschkow,[1] B. S. Koribalski[2] and L. Staveley-Smith[1]

[1]*ICRAR, M468, The University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia*
[2]*Australia Telescope National Facility, CSIRO Astronomy and Space Science, PO Box 76, Epping, NSW 1710, Australia*

Accepted 2013 November 20. Received 2013 November 17; in original form 2013 September 27

ABSTRACT

Accurate parametrization of galaxies detected in the 21-cm H I emission is of fundamental importance to the measurement of commonly used indicators of galaxy evolution, including the Tully–Fisher relation and the H I mass function. Here, we propose a new analytic function, named the ‘busy function’, that can be used to accurately describe the characteristic double-horn H I profile of many galaxies. The busy function is a continuous, differentiable function that consists of only two basic functions, the error function, $\mathrm{erf}(x)$, and a polynomial, $|x|^n$, of degree $n \geq 2$. We present the basic properties of the busy function and illustrate its great flexibility in fitting a wide range of H I profiles from the Gaussian profiles of dwarf galaxies to the broad, asymmetric double-horn profiles of spiral galaxies. Applications of the busy function include the accurate and efficient parametrization of observed H I spectra of galaxies and the construction of spectral templates for simulations and matched-filtering algorithms. We demonstrate the busy function’s power by automatically fitting it to the H I spectra of 1000 galaxies from the HI Parkes All-Sky Survey (HIPASS) Bright Galaxy Catalog, using our own c/c++ implementation, and comparing the resulting parameters with the catalogued ones. We also present two methods for determining the uncertainties of observational parameters derived from the fit.

Key words: line: profiles – methods: data analysis – radio lines: galaxies.

1 INTRODUCTION

Observations of the 21-cm emission line of neutral hydrogen provide measurements of several important parameters of galaxies, including their redshift, mass and rotational velocity, as well as evolutionary indicators such as the Tully–Fisher relation (Tully & Fisher 1977) and the H I mass function (Zwaan et al. 1997). While high-resolution H I maps have been obtained for several hundred nearby galaxies using radio interferometers, the vast majority of catalogued H I properties of galaxies has been extracted from integrated spectra obtained with single-dish telescopes. Over the next decade, H I surveys with the Square Kilometre Array (SKA; Dewdney et al. 2009) and some of its precursor and pathfinder instruments, such as Australian Square Kilometre Array Pathfinder (ASKAP; DeBoer et al. 2009) and Apertif (Oosterloo et al. 2009), will probe larger volumes of space to much greater depth than ever before. As a result, the number of H I-detected galaxies is expected to rise from currently $\gtrsim$30 000 (HYPERLEDA; Paturel et al. 2003) to more than half a million galaxies predicted for the planned all-sky surveys Widefield ASKAP *L*-band Legacy All-sky Blind surveY (WALLABY) and Westerbork Northern Sky HI Survey (WNSHS; Duffy et al. 2012). Yet, even in interferometric surveys like WALLABY, 95 per cent of all expected detections will be less than three beams across (assuming a beam size of 30 arcsec) and hence only marginally resolved, highlighting the importance of accurate parametrization methods based on the integrated H I spectrum.

Integrated H I line profiles encode much physical information. For example, (i) the frequency centroid of the H I line measures the cosmological redshift plus peculiar motion of the galaxy, (ii) the integral of the H I line provides a direct measure of the total H I mass (Roberts 1962), (iii) the linewidth traces the projected circular velocity of the galaxy and hence its dynamical mass (Casertano & Shostak 1980) and (iv) the small-scale structure of the line profile encodes information on turbulent motion and warps (Sancisi 1976). Furthermore, the shape of the H I line depends on disc asymmetries (Baldwin, Lynden-Bell & Sancisi 1980), extra-planar gas (Swaters, Sancisi & van der Hulst 1997; Heald et al. 2011), tidal tails and companions. Finally, observational settings, such as spectral resolution and noise level, also affect observed H I lines. The efficient extraction of all this information from thousands of noisy H I spectra requires a quick and accurate parametrization method.

2014, MNRAS, 438, 1176

Thank You!

Busy Function

Russell Jurek
Danail Obreschkow
Bärbel Koribalski
Lister Staveley-Smith

SoFiA

Lars Flöer
Nadine Giese
Russell Jurek
Martin Meyer
Attila Popping
Paolo Serra
Benjamin Winkel