Unsupervised Source Extraction for the Effelsberg-Bonn HI Survey

Lars Flöer AlfA, University of Bonn, Germany 7th PHISCC Workshop ASTRON / Dwingeloo, 2014



Argelander-Institut für Astronomie





EBHIS

- Complete $\delta > -5^{\circ}$ HI survey with Effelsberg 100-m out to $z \approx 0.07$
 - Galactic and extragalactic survey in one pass
 - Northern complement to GASS and HIPASS
 - EBHIS and GASS will supersede LAB

- Survey Team
 - PD Dr. Jürgen Kerp (PI, AlfA)
 - Dr. Benjamin Winkel (MPIfR)
 - Dr. Nadya Ben Bekhti (AlfA)
 - Dr. Peter Kalberla (a.D.)
 - Lars Flöer (AlfA)
 - Daniel Lenz (AlfA)

EBHIS

- 10.8 arcmin
- -600 km/s < v_{LSR} < 18,000 km/s
- Galactic Survey
 - 1.45 km/s
 - 90 mK
- Extragalactic Survey
 - 10.24 km/s
 - · 23 mJy/beam



State of the Art

- Human labor
 - HIPASS & ALFALFA
 - Inspect $\geq 10^5$ candidates
 - Parametrize $\sim 10^{3-4}$ sources

- WALLABY / WNSHS will detect ~800,000 galaxies
- Need for unsupervised source extraction
 - Source finding, parametrization and classification

- Extract significant structure at all scales
- 2D-1D transform adapted to spectroscopic data Starck et al. 2009 Flöer & Winkel 2012
- Robust against common defects
- Will be part of SoFiA

- Extract significant structure at all scales
- 2D-1D transform adapted to spectroscopic data Starck et al. 2009 Flöer & Winkel 2012
- Robust against common defects
- Will be part of SoFiA



NGC 3628

- Extract significant structure at all scales
- 2D-1D transform adapted to spectroscopic data Starck et al. 2009 Flöer & Winkel 2012
- Robust against common defects
- Will be part of SoFiA







Parameterization

- Requirements for EBHIS pipeline
 - Automated
 - Flexible
 - Robust

	Lexikon
	Q parameterization S
Alle Deutsch Apple Wikipedia Englisch »	
parameterization	<pre>parameterize pə'ramıt(ə)rʌız (also parametrize pə'ramıtrʌız) verb [with obj.] technical describe or represent in terms of a parameter or parameters. the nuclear charge distribution can be parameterized directly using a suitable mathematical form. DERIVATIVES parameterization [-,ramıtərʌı'zɛıʃ(ə)n noun</pre>

Baseline Fitting

- · Polynomials
 - Require selection of degree
 - Sensitive to outliers
- Robust spline smoothing Garcia, 2010
 - Parameter-free
 - Insensitive to outliers
 - Handles missing data



Line Width

- Commonly measured by w_{50}
 - Accuracy depends on peak SNR and mask
 - Smoothing increases peak SNR but lowers resolution
- Bilateral filtering
 Tomasi & Manduchi 1998
 - Smoothing based on proximity and similarity
 - Preserves edges



Parameter Accuracy compared to HIPASS



- Easy for humans, hard for machines
- Artificial Neural Network
 - Learn complex decision functions from training data
 - Revitalized field of research

- Training data set
 - 7,500 simulated sources
 - 7,500 EBHIS artifacts
 - 38 Parameters
- Output interpreted as probabilities
- 0.6% Loss
- 99.7% Reliability



- Training data set
 - 7,500 simulated sources
 - 7,500 EBHIS artifacts
 - 38 Parameters
- Output interpreted as probabilities
- 0.6% Loss
- 99.7% Reliability



- Training data set
 - 7,500 simulated sources
 - 7,500 EBHIS artifacts
 - 38 Parameters
- Output interpreted as probabilities
- 0.6% Loss
- 99.7% Reliability



Summary

- EBHIS as a testbed for automated source extraction
- Source finding with wavelet denoising
 - Adapted to 2D-1D data
 - Robust against common defects
- Automated, time-tested parametrization algorithms
 - Baseline estimation with robust splines
 - Adaptive smoothing for line width measurement
- Classification with Artificial Neural Networks
 - Training with a hybrid data set
 - High reliability and completeness