All-sky selection of WISE YSO candidates

G. Marton[1,*], S. Zahorecz[2,3], L. V. Tóth[3], Cs. Kiss[1]

Abstract

The Wide-field Infrared Survey Explorer (WISE) observed 100% of the sky at 4 wavelengths (3.6, 4.6, 12 & 22 μm). Sources listed in the ALLWISE Catalog with S/N>5 in all bands were investigated. An 8 dimensional parameter space (including associated 2MASS data) was built to identify potential Young Stellar Object (YSO) candidates. Using Support Vector Machines (SVM), a machine learning algorithm we were able to efficiently remove the sources identified as contaminants, and we compiled a catalog of 1,141,437 candidates. Comparison with SIMBAD database showed that the predicted contamination of this selection is well below 1%. We conclude that SVM outperforms other simple colour-colour and colour-magnitude methods. The spatial distribution of the candidates has been investigated in correlation with interstellar bubbles identified by the Milky Way project. An overdensity of young stars was observed in the direction of the bubbles.

I - Introduction

WISE (Wright et al., 2010) observed 100% of the sky at 4 wavelengths (W1=3.6, W2=4.6, W3=12 & W4=22 μm). 5σ point source sensitivities of 0.08, 0.11, 1 & 6 mJy were achieved in unconfused regions, with angular resolutions of 6.1″, 6.4″, 6.5″ & 12.0″, respectively. The ALLWISE catalogue is an improved version of the WISE Point Source Catalog (Cutri et al., 2012) and it contains information on ~746 million sources. We compiled a subcatalog of ALLWISE detections with S/N>5 in all WISE bands and with existing associated 2MASS J, H & K_s magnitudes. It resulted in a list of 5.5 million sources.

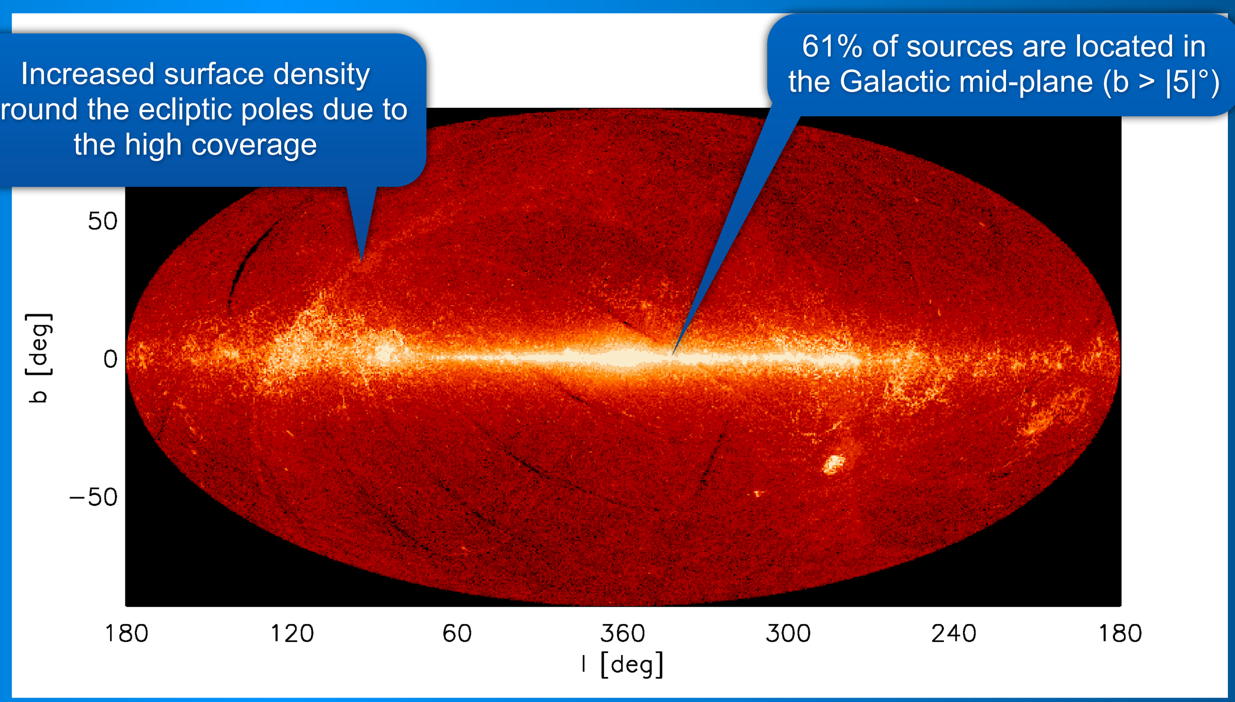

Figure: Surface density distribution of the used ALLWISE sources presented on logarithmic scale. Values are ranging from 10^{-1} to 10^{3}, averaged in 0.5x0.5 degree bins.

II - YSO selection method

For classification and pattern recognition in multidimensional data, one can use several statistical methods. We used the Support Vector Machines, a class of supervised learning algorithm, created as an extension to nonlinear models of the generalized portrait algorithm developed by Vladimir Vapnik (Vapnik 1995), for classification in a multidimensional parameter space. The parameters we used were J-H, H-K_s, K_s-W1, W1-W2, W2-W3, W3-W4 colours, W1 brightness and the extended source flag.

Methods like SVM need a training set to be prepared, which is then used to determine the boundaries in the parameter space between the different object types. Our training set was prepared by using the SIMBAD and VizieR databases. 5″ radius was used to match our sources.

In our classification process we first identified the extragalactic contamination. The remaining sources were classified as ISM related objects, evolved stars, field (or main sequence) stars and YSO candidates.

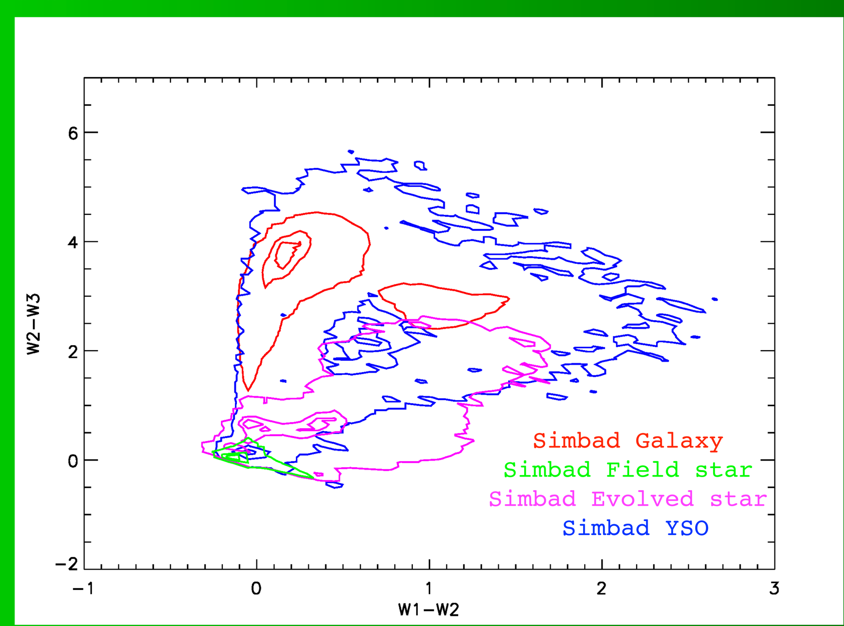

Figure: Example of colour-colour diagram demonstrating that boundaries between different object types are not linear. Surface density contour levels are at 5%, 50% and 75% of maximum. The bin size is 0.1 mag.

III - Results

We were able to compile a catalogue of 1,141,437 sources as potential YSO candidates. 77.3% of the known YSOs were re-classified as YSO candidate. Based on SIMBAD the estimated contamination is under 0.4%. Using the SDSS DR-9 (Adelman-McCarthy et al. 2012) source classification flags we found that the fraction of sources with galaxy flag in our sample is only 11%, while it is 24% among the known SIMBAD YSOs.

We compared our selection to various YSO catalogues from the literature. The goodness of our selection is 85.8% for Class I, 87.5% for Class II and 43.9% for Class III sources. Based on these results we classified our candidates into evolutionary classes.

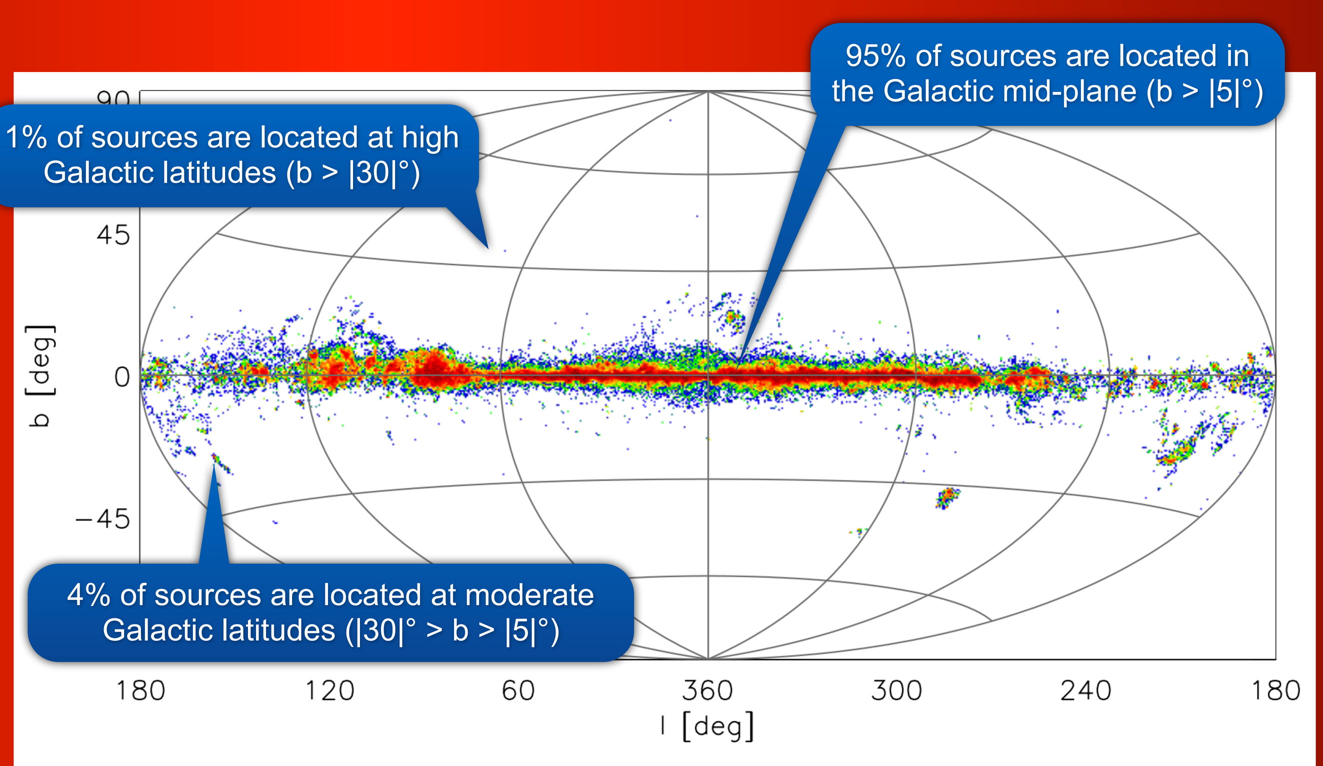

Figure: Surface density distribution (logarithmic scale) of the sources classified as YSO candidates. Surface density values are ranging from 10^{-1} to 10^{3}, averaged in 0.5x0.5 degree bins.

IV - Discussion & Conclusion

The spatial distribution of our Class I and II candidates have been investigated in correlation with the interstellar bubbles identified by the Milky Way Project (MWP, Simpson et al., 2012). We used 3744 of the large bubbles of the MWP catalogue and calculated the YSO surface density inside, on and outside the bubbles, within five times their effective radius. Overlapping bubbles were not included. We found that in case of 1848 bubbles the surface density inside or on the bubble is higher than in their neighborhood. We created Monte-Carlo simulations to test the significancy of the result. Based on 100 simulations the average number of bubbles showing YSO surface density excess was 345±30. In the studied region 36% of the YSO candidates were found to be located in the direction of the bubbles. We conclude that the bubbles and the physical precesses behind bubble formation play an important role in the early stages of star formation.

References:
Adelman-McCarthy, J.K. et al., 2012, ApJS, 203, 21
Cutri, R. M. et al., 2012, WISE All-Sky Data Release, NASA/IPAC IRSA
Vapnik, V. N. 1995, The Nature of Statistical Learning Theory (Springer)
Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K. et al., 2010, AJ, 140, 1868

*email: marton.gabor@csfk.mta.hu
1)Konkoly Observatory, Research Centre for Astronomy and Earth Sciences, Hungarian Academy of Sciences, Konkoly Thege Miklós út 15-17, 1121 Budapest, Hungary
2)European Southern Observatory, Karl Schwarzschild str. 2., D-85748 Garching by München, Germany
3)Eötvös Loránd University, Department of Astronomy, Pázmány Péter sétány 1/A, 1117 Budapest, Hungary

Figure on top: RGB composite image of the MWP bubble located at l=349.8, b=-0.59. The overlayed red dots represent our YSO candidates associated to this bubble. The green circles represent other YSO candidates from our selection. The false-colour image was created by using the WISE 4.6, 12 and 22 μm images.

Acknowledgement:
This work has been supported by the following grants: i) PECS contract no. 98073 of the Hungarian Space Office and the European Space Agency, ii) Hungarian Research Fund (OTKA) grants nr. 101939 and 104607, iii) Bolyai Research Fellowship of the Hungarian Academy of Sciences. This research has made use of the SIMBAD database, operated at CDS, Strasbourg, France. This publication makes use of data products from the Wide-field Infrared Survey Explorer, which is a joint project of the University of California, Los Angeles, and the Jet Propulsion Laboratory/California Institute of Technology, funded by the National Aeronautics and Space Administration.