Wolf-Rayet Stars
W.-R. Hamann, A. Sander, H. Todt, eds.
Potsdam: Univ.-Verlag, 2015
URL: http://nbn-resolving.de/urn:nbn:de:kobv:517-opus4-84268

Finding new Wolf-Rayet stars in the Magellanic Clouds

A. Becker[1], **D. J. Bomans**[1,2], **K. Weis**[1]
[1]*Astronomical Institute of the Ruhr-University Bochum, Germany*
[2]*Research Department Plasmas with Complex Interactions*

Obtaining a complete census of massive, evolved stars in a galaxy would be a key ingredient for testing stellar evolution models. However, as the evolution of stars is also strongly dependent on their metallicity, it is inevitable to have this kind of data for a variety of galaxies with different metallicities. Between 2009 and 2011, we conducted the Magellanic Clouds Massive Stars and Feedback Survey (MSCF); a spatially complete, multi-epoch, broad- and narrow-band optical imaging survey of the Large and Small Magellanic Clouds. With the inclusion of shallow images, we are able to give a complete photometric catalog of stars between B $\approx$ 18 and B $\approx$ 19 mag.

These observations were augmented with additional photometric data of similar spatial resolution from UV to IR (e.g. from GALEX, 2MASS and Spitzer) in order to sample a large portion of the spectral energy distribution of the brightest stars (B $<$ 16 mag) in the Magellanic Clouds. Using these data, were are able to train a machine learning algorithm that gives us a good estimate of the spectral type of tens of thousands of stars.

This method can be applied to the search for Wolf-Rayet-Stars to obtain a sample of candidates for follow-up observations. As this approach can, in principle, be adopted for any resolved galaxy as long as sufficient photometric data is available, it can form an effective alternative method to the classical strategies (e.g. He II filter imaging).

1 Introduction

The Wolf-Rayet star population is of particular interest for stellar evolution modeling. This applies not only to the pure number count and the WC/WN-ratio but also to the specific locus of the Wolf-Rayet stars in the Hertzsprung-Russell-Diagram. Unfortunately the Wolf-Rayet population in most galaxies is barely known. The census of WR stars in the Magellanic Clouds was long considered to be nearly complete, but new Wolf-Rayet surveys of the Magellanic Clouds, e.g. by Massey et al. (2014, 2015), revealed a new subtype of WR stars and increased the number of known WR stars in the LMC by 9.3 % (139 to 152).

The standard technique to find Wolf-Rayet stars in the optical is based on the use of a He II λ4686 filter and a C III λ4650 filter in combination with the matching continuum filter. This image differencing technique is quite expensive as it requires a large amount of observational time; e.g. $\approx$30 hours HST observations to image large parts of M101 in the F469N filter, see Shara et al. (2013). Another promising method is the selection of candidates via their infrared colors as free-free scattering in the dense ionizing wind of WR stars lead to an excess in the infrared (Mauerhan et al. 2011). This however, requires a set of MIR data.

Here we present a new and more general approach to estimate the spectral type of stars that is also capable of finding Wolf-Rayet candidates. It works for all external galaxies where the stellar population can be resolved. Our method relies only on photometric data and has therefore the potential to make use of the already available and upcoming photometric surveys.

2 The method

Between 2009 and 2011 we conducted the Magellanic Clouds Massive Stars and Feedback Survey (Bomans et al. 2015). Using two 6 inch telescopes at the Universitätssternwarte Bochum on Cerro Armazones in Chile we gathered spatial complete high quality images in a large set of photometric filters (u, B, V, R, I, Hα, [OIII] and [SII]). By including shallow exposures (10 to 30 seconds) we were able to measure even the brightest stars in the Magellanic Clouds. To cover a larger wavelength range the dataset was supplemented with photometric data from the UV (GALEX, Swift) to the IR (2MASS, Spitzer, WISE). By doing so we obtained spectral energy distributions (SED) for $\approx$ 80000 stars. Based on these SEDs we estimate the spectral class of each star.

For the classification we used a supervised learning approach. These stars with an already known spectral classification found in the catalog of Skiff (2013) were used as a so-called training sample. Based on this subset an algorithm is trained to recognize the spectral type only based on the multi-wavelength photometry. Once trained, the algorithm can be applied to the rest of the data to estimate the spectral types of all of the other stars. It turned out that a decision forest build up by several decision trees that determine the classification by majority vote led to the best result and could reproduce 98% of the training data. The Large and Small Magellanic Cloud were treated separately to avoid confusion due to different metallicities and differences in the set of filters.

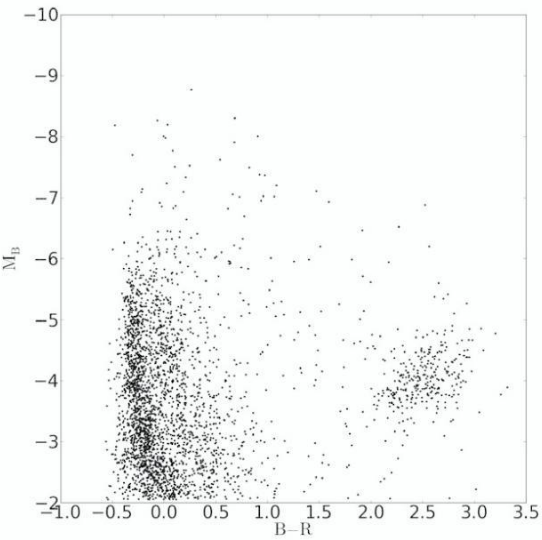

Fig. 1: Distribution of the training sample for the SMC in the color-magnitude-diagram. The region of the foreground stars and the region below the red supergiants are sparsely sampled and will therefore give no reliable classifications.

2.1 Evaluation of the method

To evaluate the results we repeated the training procedure while retaining about 900 objects from the full training sample as a test sample. On this test sample the classification was compared with their true spectral classification as given by Skiff et al. (2013). The result of this test is shown as a histogram in Figure 2. The deviation is measured as the difference between the true spectral class and the estimated spectral class. The value represents the deviation in subtypes (e.g.: An O7 star estimated as O9 deviates by -2; An B2 star estimated as B1 deviates by +1). If we look at all stars we see a deviation of up to 2-3 subtypes with just a few catastrophic outliers and with more than 40% without noticeable deviation. Considering only the O-stars a somewhat larger deviation of about 4 subtypes becomes apparent with a systematic trend to classify them as later types than their actual spectral type.

Systematic errors also occur as a consequence of the training sample. Not all types of stars are equally represented and regions in the color-magnitude-diagram that are sparsely populated with training data will not yield reliable results (see table 1). The smaller sample used in the evaluation might therefore lead to a lower precision compared to the final classification.

2.2 Finding Wolf-Rayet stars

The training sample included 78 of the 139 known Wolf-Rayet stars. Of those the algorithm was able to reproduce all but one. This further proved the reliability of our method and encouraged us to use this method as a new approach to search for Wolf-Rayet candidates. To obtain even better results we fine tuned the method for the identification of Wolf-Rayet stars with a second supervised learning algorithm (multilayer perceptron). This algorithm was just trained on the decision Wolf-Rayet star or no Wolf-Rayet star and applied to the results of the first classification. As an additional cross-check we can compare the results with the observations by Massey et al. (2014, 2015) (see table 1). Unfortunately we do not have an overlap with their newly discovered WN3/O3-type stars as they have a typical brightness that is below the magnitude-cut applied to the algorithm ($B \leq 16$ (17) for the LMC (SMC)). The classification for the OB-stars is generally in good agreement and within the typical errors of 2-3 subtypes, as discussed before. We were also able to predict the new found WR star LMC143-1. The apparent discrepancy for two stars that were predicted as A-supergiants is still under investigation and subject of the current improvement of the method.

Tab. 1: Comparison between the spectral classifications from Massey et al. (2014, 2015) with the results from our method.

ID	Massey	this work
LMC143-1	WN3+O8-9III	WR
LMC173-1	WN3+O7.5	WR
SMC159-2	O9f?p	B1/5III
LMC156-1	O6If	O4V((f))
LMC164-2	O8f?p	WN4.5
LMC173-2	O7.5Iaf	O9.7Ia
LMC174-4	O4Ifc	A1Iab
LMC174-5	B[e]+WN?	Be
LMCe063-1	WN11	A2Ia
LMC143-1	Onfp	O7n(f)

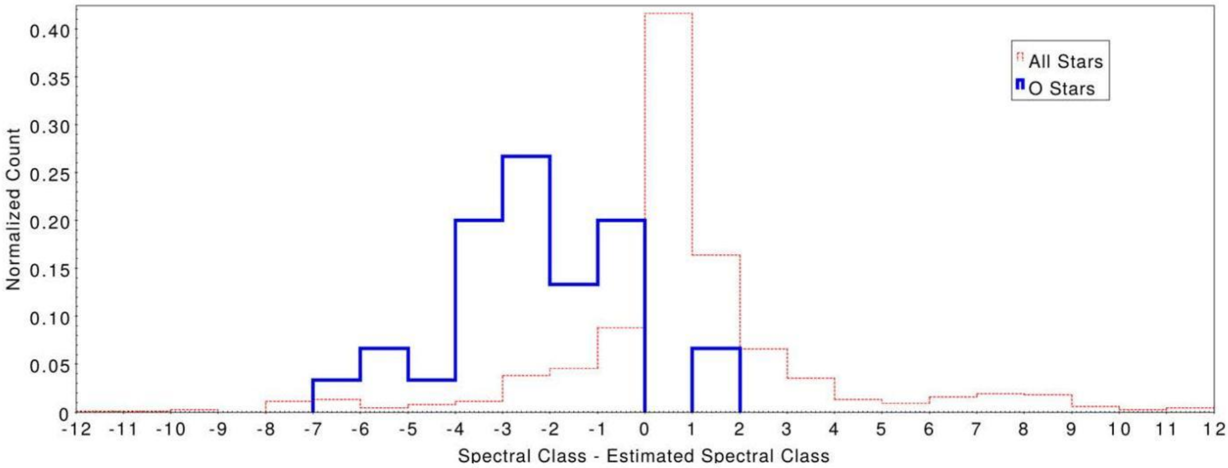

Fig. 2: Deviation of the estimated spectral classification from the true spectral class measured in subtypes. The red dotted line represents all stars and show a deviation of 2-3 subtypes. The blue solid line represents only the O-stars and shows a systematic trend to later spectral types with a deviation of about 4 subtypes.

3 Future prospects

As a next step we will transfer the algorithm to work for other galaxies. M33 would be an obvious target. With its metallicity gradient and a metallicity that is, for some parts of the galaxy, comparable to the LMC it is an ideal object to test the influence of the metallicity on the precision of the estimates. This will help us determine the feasibility of transferring e.g. the SMC training sample to galaxies with lower metallicity where no or few spectral classifications are available (see Bomans et al., this conference). Currently we are still solving some challenges in the preprocessing of the data such as the relative distance between the LMC and M33. This is because we have to match the magnitudes of the LMC photometry to those of M33. And even a small offset, of e.g. 0.2 mag, will result in a complete mismatch in the classification.

4 Summary & Conclusions

With the use of data mining techniques we were able to estimate the spectral type of about 80000 bright stars in the Magellanic Clouds. For that, a training sample, consisting of stars with known spectral classification, was used to train a decision forest that is capable of distinguishing the spectral types by photometric measurements only. The typical precision of our method is currently in the order of 2-3 subtypes. For the O-stars we have a systematic trend to later types with a typical deviation of up to 4 subtypes. This technique was extended to the search for Wolf-Rayet stars in the Magellanic Clouds. The work done so far serves as a proof of concept. By transferring this technique to other local galaxies and photometric datasets we plan to apply the method to upcoming large surveys such as GAIA, LSST and VISTA.

References

Bomans, D. J., Becker, A., & Weis, K. 2015, in IAU Symposium, Vol. 307, IAU Symposium, 414–419

Massey, P., Neugent, K. F., & Morrell, N. 2015, ArXiv e-prints

Massey, P., Neugent, K. F., Morrell, N., & Hillier, D. J. 2014, ApJ, 788, 83

Mauerhan, J. C., Van Dyk, S. D., & Morris, P. W. 2011, AJ, 142, 40

Shara, M. M., Bibby, J. L., Zurek, D., et al. 2013, AJ, 146, 162

Skiff, B. A. 2013, VizieR Online Data Catalog, 1, 2023

John Eldridge: Have you thought about using theoretical spectra as a training set?

Alexander Becker: Yes, and there is an interesting poster, but a problem of putting the distance of the LMC and making it like the observations.

Phil Massey: Neugent and I have new spectral types for $\sim$ 1700 stars in M31 and M33. We're on the verge of publishing them, and they may be usefull in your very interesting work.

Claus Leitherer: Is there anything special with your outliers? E.g., are they in a crowded region, or are they at the faint limit, or is there contamination by nebular lines?

Alexander Becker: No, they are not in problematic regions like 30 Dor or at the faint limit.