Galaxy Nurseries: Crowdsourced analysis of slitless spectroscopic data

Hugh Dickinson,[1] Claudia Scarlata,[1] Lucy Fortson,[1] Micaela Bagley,[1] Vihang Mehta,[1] John Phillips,[1] Ivano Baronchelli,[2] Sophia Dai,[3] Nimish Hathi,[4] Alaina Henry,[4] Matthew Malkan,[5] Marc Rafelski,[4, 6] Harry Teplitz,[2] Anita Zanella,[7] and Chris Lintott[8]

[1] School of Physics and Astronomy, Tate Laboratory of Physics, 116 Church St. SE, Minneapolis, MN 55455, United States of America
[2] IPAC, Mail Code 314-6, Caltech, 1200 E. California Blvd., Pasadena, CA 91125, United States of America
[3] National Astronomical Observatories Chinese Academy of Sciences: Chaoyang-qu, Beijing, China
[4] Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, United States of America
[5] Physics and Astronomy Department, University of California, Los Angeles, CA 90095-1547, USA
[6] Department of Physics & Astronomy, Johns Hopkins University, Baltimore, MD 21218, USA
[7] European Southern Observatory, Karl Schwarzschild Straße 2, 85748 Garching, Germany
[8] Oxford Astrophysics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford, OX1 3RH, UK

THE GALAXY NURSERIES PROJECT

The *Galaxy Nurseries*[1] project was designed to enable crowdsourced analysis of slitless spectroscopic data by volunteers using the *Zooniverse*[2] online interface. The dataset was obtained by the WFC3 Infrared Spectroscopic Parallel (WISP) Survey collaboration (Atek et al. 2010) and comprises NIR grism (G102 and G141) and direct imaging of 432 fields. The scientific goals of WISP Survey require reliable identification of emission lines (e.g. Masters et al. 2014; Atek et al. 2014). Spectral contamination by overlapping signals from multiple sources or diffraction orders as well as residual artifacts of the data reduction can significantly complicate the automatic detection and identification of emission lines. Visual verification of automatically detected features has proved essential to obtain a pure sample of emission lines.

In *Galaxy Nurseries* verification of putative emission lines was delegated to citizen-scientist volunteers. Data were presented as "subject images" using the fixed format illustrated in the top panels of Figure 1. For each target, volunteers were provided with the two-dimensional spectrum (B), the corresponding one-dimensional extraction (A) and its direct image (C). Volunteers were instructed to evaluate only the feature identified by the green crosshairs and to decide whether it was a genuine emission line or more likely an artifact, providing a Boolean Valued (i.e. "Real" or "Spurious") label. Following its launch, *Galaxy Nurseries* was completed in only 40 days, gathering 414,360 classifications from 3003 volunteers for 27,333 putative emission lines. At least 15 classifications were obtained for each subject image. For reference, it took approximately 4.5 months for the full sample of lines to be visually inspected by two members of the WISP Survey Science Team (WSST).

Corresponding author: Hugh Dickinson
hdickins@umn.edu

[1] www.zooniverse.org/projects/hughdickinson/galaxy-nurseries
[2] www.zooniverse.org

RESULTS

Volunteer responses for each subject were aggregated to compute f_{Real}, the fraction of volunteers who classified the corresponding emission line as "Real". To evaluate the accuracy of volunteer classifications, their aggregated responses were compared with independent assessments provided by members of the WSST[3]. The distribution of f_{Real} for all subjects is shown in the *bottom-left* panel of Figure 1 and is subdivided to identify emission lines that were verified (blue) and vetoed (orange) by the WSST. The relative small number of lines that were validated by the WSST underlines the need for visual inspection of the full sample. Overall, there is a broad agreement between the WSST and volunteers classifications. The distribution of f_{Real} peaks at 0.85 for the sample of verified lines, indicating that the majority of volunteers agree on the reality of these lines. Similarly, the distribution of f_{Real} peaks at 0.1 for the sample of vetoed lines. It is clear, however, that the agreement is not perfect: there is a large number of lines which were vetoed by the WSST and would be classified as real by most of the volunteers (about one third of the objects with $f_{\text{Real}} > 0.7$ have been vetoed by the WSST). The opposite is not true: only a minority of lines with low f_{Real} were in fact real according to the WSST. This result suggests that, with a minimal level of training, volunteers are successful in securely identifying artifacts in the data.

The result is reinforced by a more quantitative analysis that accounts also for the emission line signal-to-noise ratio (S/N). The remaining *middle* panels of Figure 1 illustrate how the purity ($P_S = TP/(TP + FP)$)[4] and completeness ($C_S = TP/(TP+TN)$) change as a function of $f_{\text{Real,threshold}}$ and S/N. These results show that the degree of categorical separation provided by volunteer classifications may be adequate in certain situations. If purity is paramount, then for $S/N \gtrsim 5$, choosing $f_{\text{Real,threshold}} \sim 0.7$ yields a sample purity $P_S \gtrsim 0.8$. The corresponding sample completeness is $C_S \sim 0.6$. Conversely, a more complete ($C_S \sim 0.9$) albeit impure ($P_S \sim 0.6$) sample is feasible for $S/N \gtrsim 10$ if $f_{\text{Real,threshold}} \sim 0.5$.

The results of *Galaxy Nurseries* demonstrate the feasibility of identifying genuine emission lines in slitless spectra by citizen scientists. We recognise that robust scientific analyses typically require samples with higher purity and completeness than raw volunteer classifications provide. Nonetheless, choosing optimal values for $f_{\text{Real,threshold}}$ allows a large fraction of spurious lines to be vetoed, substantially reducing the timescale for subsequent professional analysis of the remaining potential lines.

REFERENCES

Atek, H., Malkan, M., McCarthy, P., et al. 2010, ApJ, 723, 104, doi: 10.1088/0004-637X/723/1/104

Atek, H., Kneib, J.-P., Pacifici, C., et al. 2014, The Astrophysical Journal, 789, 96

Masters, D., McCarthy, P., Siana, B., et al. 2014, The Astrophysical Journal, 785, 153

---

[3] Full results of the WSST evaluation will be presented in Bagley et al. 2019 (in prep.)

[4] The labels T(rue)/F(alse) are assigned to lines verified/vetoed by the WSST, while lines for which f_{Real} greater/smaller than $f_{\text{Real,threshold}}$ are labeled as P(ositive)/N(egative).
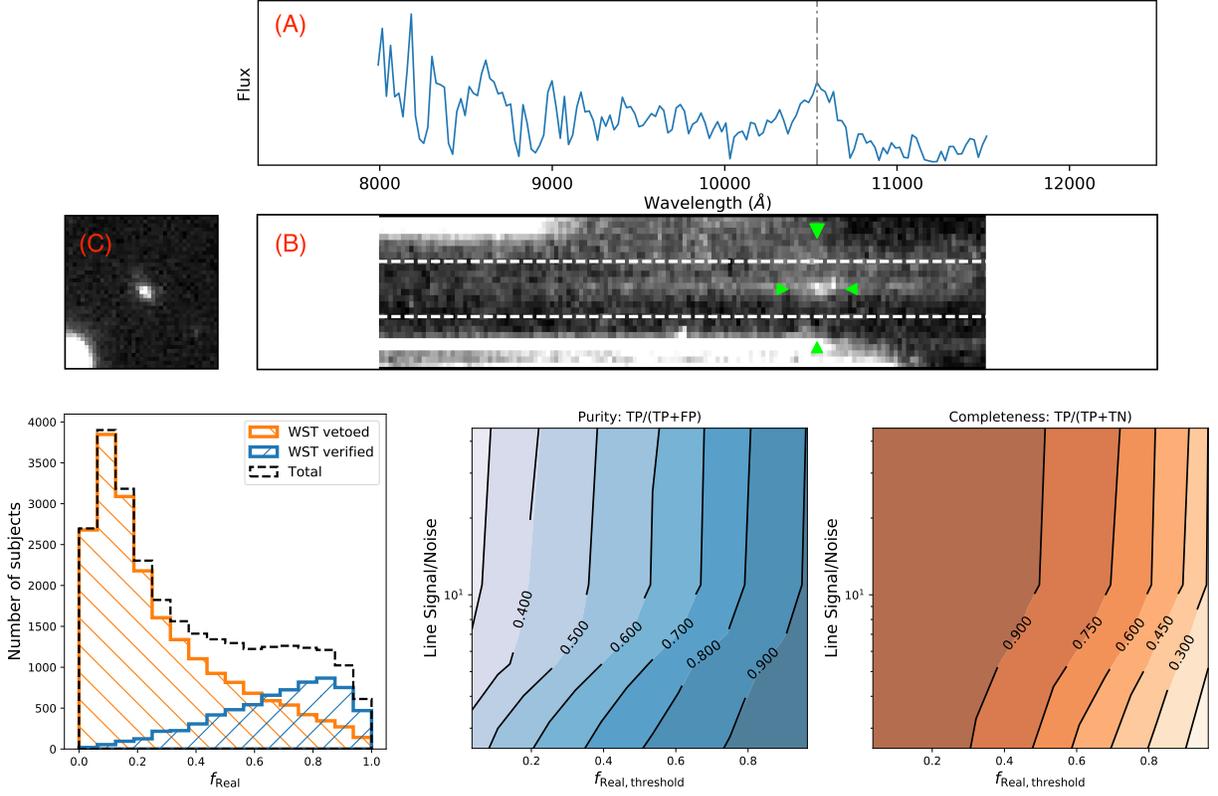
Figure 1. *Top:* Typical fixed-format subject image used by volunteers to classify putative emission lines as real or spurious. *Lower left:* The distribution of $f_{\mathrm{Real,threshold}}$ for the full sample (*black, dashed*), the sub-sample of WSST-verified lines (*blue*) and the sub-sample of WSST-vetoed lines (*orange*). Sample purity ($P_S = TP/(TP + FP)$, *lower centre*) and completeness ($C_S = TP/(TP + TN)$, *lower right*) realised using volunteer classifications as a function of threshold subject vote fraction $f_{\mathrm{Real,threshold}}$ and line signal-to-noise ratio.