=================
Referee report(s) on this version:

Report by referee 1:

The paper "Mapping Halpha -Excess Candidate Point Sources in the Southern Hemisphere Using S-PLUS Data" by Gutiérrez-Soto et al. presents the photometric selection of candidate H-alpha emitters via the color-excess they show in the r-band minus H-alpha narrow-band filter used by the S-PLUS project. Their results are compared with previous works, and validated via the identification of known objects in public databases. A method for object classification is also presented.

The results are interesting and deserve publication, providing some changes are applied.

I have a couple of relatively general suggestions:

1) Regarding the various classes of H-alpha emitters, a good summary is given by listing the relevant IPHAS results in the introduction. However, some more detailed results could be given for YSOs in IPHAS (Barentsen et al. 2011) or VPHAS+ (Kalari et al. 2015), as they show that the photometric excess can be used to determine accretion rates. Furthermore, in the bright range of magnitudes observed by IPHAS, it appeared that the a large number of H-alpha emitters were Classical Be stars (Corradi et al. 2008; Raddi et al. 2015). This result might be applicable to your list of candidates, but the comparison with SIMBAD does not seem to show that. It may be useful to look into this comparison by exploring the BeSS database (Neiner 2018; https://arxiv.org/abs/1811.05261)

WE would like to thank the refere for this thourough revision of our paper and the

Reply: ~~Thank you for~~ the valuable suggestions and ~~references~~ comments regarding other classes of Halpha emitters. We have incorporated a paragraph in the introduction to address the point~~s raised.~~ as suggested by the referee.

Regarding Be stars, our results align with expectations, as the Main Survey (MS) primarily targets regions far from the Galactic plane, where known Be stars are less prevalent. Our cross-match of the MS Halpha sources with SIMBAD confirmed the absence of known Be stars. Additionally, a cross-match with the BeSS database yielded no coincidences, further supporting this conclusion.

For the S-PLUS Galactic Disk Survey (GDS), which covers approximately 347.4 square degrees, we identified two Halpha sources coinciding with entries in SIMBAD, but none matched with BeSS. Extending this analysis to the entire MS and GDS catalog, we found one match in the MS. However, this object (r = 11.19) was excluded from our cleaned sample due to its magnitude.

While the absence of known Be stars in our sample is consistent with expectations, we acknowledge that some unidentified Be stars might still be present in our Halpha source list. Factors such as the limited GDS coverage and the non-uniform distribution of Be stars in the Galactic disk likely contribute to the low recovery rate. Moreover, previous studies (e.g., Corradi et al. 2008; Raddi et al. 2015) have identified Be stars in Halpha samples from surveys like IPHAS. However, the regions targeted in our study differ significantly, which likely explains the lack of detections.

write that for the first comment only

2) The adopted "Machine learning approaches" aim at providing a provisional photometric classification for the selected candidates. This result could be very useful for a targeted spectroscopic follow up. This goal is a bit lost in the small introduction given in section 5, while more details are spent on the method. Here, I would emphasize more the possible uses of photometric classification and a good reference might be the Corradi et al. (2008) paper, where different classes of H-alpha emitters appear to separate in the r-i, r-Ha diagram. In particular, it would be useful to draw tentative color-cuts in Fig. 18, which would allow a selection of candidate emitters without needing a classification algorithm.

Reply: We appreciate the referee's thoughtful comments. We have added text to the manuscript to emphasize that a preliminary classification of the Halpha sources was conducted using machine learning methods, with the goal of selecting candidates for targeted spectroscopic follow-up. We also agree that highlighting the potential uses of photometric classification is important. To address this, we have included tentative color cuts in the revised version of Fig. 9 (previously Fig. 8) to separate different classes of Halpha sources without relying on a classification algorithm. The manuscript text has been modified accordingly to reflect this addition.

Other minor comments:

- Section 2: what are SEX_FLAG_DET and CLASS_STAR? Please detail their usage

Reply: The SEX_FLAGS_DET parameter is a bit-flag from SExtractor indicating potential photometric issues. We selected objects with SEX_FLAGS_DET < 4 to ensure high-quality photometry. The CLASS_STAR parameter classifies sources as stars (1) or non-stars (0). We applied CLASS_STAR_r = 1 and CLASS_STAR_i = 1 to select sources with a high likelihood of being stars in both the r and i filters. We have added more details in the revised manuscript.

- Section 2: spatial variations of the PSF are modeled with a third degree polynomial. As a function of the XY position on CCD?

Reply: Yes, the spatial variations of the PSF were modeled using third-degree polynomials as a function of the pixel coordinates (X, Y) on the CCD. This method allows for accurate modeling of PSF variations across the field of view, which improves photometric precision. The text has been updated to reflect this clarification.

- Section 2: not many details are given on the photometric calibration of the survey, but issues with that are mentioned. Could you summarize how the flux calibration was performed?

Reply: To address the comment, we have added a detailed paragraph to the manuscript that clarifies the flux calibration process. This description includes the

use of synthetic photometry combined with reference catalogs to derive zero points (ZPs), the application of the stellar locus method to calibrate the u, J0410, and J0430 filters in regions lacking external calibration data, and the internal calibration steps to further refine the ZPs, improving accuracy to 0.01–0.02 mag. Additionally, we explain how the flux calibration is aligned with the Gaia system using average offsets derived from synthetic photometry. These procedures ensure consistent calibration across different regions of the sky, as described by Herpich et a. 2024.

- Survey strategy: are the r-, H-alpha, and i-band filters observed back to back? This information could be relevant for variable sources, e.g. RR Lyrae, eclipsing binaries, cataclysmic variables.

Reply: We appreciate valuable suggestion to clarify the observation sequence of the filters r, J0660 (Halpha), and i, as this information is indeed critical for studying variable sources such as RR Lyrae stars, eclipsing binaries, and cataclysmic variables.

In the survey, the filters are observed consecutively in the following sequence: u, J0378, J0395, J0410, J0430, J0515, g, r, i, z, J0660, J0861. The total observation time for each field is approximately 90 minutes, covering all 12 filters.

Each filter is observed with three exposures, and the average time per filter, including exposures, readout, and transitions, is approximately 7–8 minutes. Then, the filters r, J0660, and i are not observed back-to-back: There is a time gap of approximately 37–38 minutes between the r and J0660 observations due to the intervening filters. Similarly, there is a gap of around 15 minutes between the J0660 and i observations.

time candence and study as well as

This sequence and ~~timing~~ allow the detection of variable sources and the construction of light curves on time scales shorter than 30 minutes, which is suitable for studying short-period variability in sources like RR Lyrae stars and eclipsing binaries. We have added this information to the text accordingly.

I think this it wrong. we can study variable source on time scales longer than 30 minutes.

- Selection: the concentration of H-alpha emitters in specific areas of sky may be related to the presence of star forming regions and young clusters; thus, the higher numbers observed at specific galactic coordinates may not be significant, due to the patchy coverage of sky (Fig. 13 or Fig 14)

thank for this

Reply: We ~~appreciate~~ the referee~~'s~~ comment regarding the observed concentrations of Halpha excess sources in specific regions of the sky. We acknowledge that these concentrations may be influenced by both the presence of star-forming regions and the uneven sky coverage of the S-PLUS survey. In the Main Survey (MS), the distribution of Halpha excess sources is relatively uniform across Galactic longitude, reflecting the general distribution of stars in high-latitude regions with less star formation. In contrast, in the Galactic Disk Survey (GDS), the more pronounced peaks in the distribution are likely associated with star-forming regions, though they may also be influenced by the patchy sky coverage of the survey, as these peaks are not as significant when compared to the general stellar population.

We have also ~~included~~ added **a reference to Figure 6 in Herpich et al. (2024), which illustrates the sky coverage of the S-PLUS survey and provides** the **readers with additional context regarding the survey's footprint. The following text has been added to the manuscript:**

"It should be noted that the observed concentrations of Halpha excess sources in certain Galactic regions may be influenced by the uneven sky coverage of the S-PLUS survey (see Figure 6 in Herpich et al. 2024). In the MS, this effect may be more pronounced due to the lower star formation activity in high-latitude regions. For the GDS, the concentrations may be influenced by both the presence of star-forming regions and the patchy sky coverage of the survey. This limitation may lead to uneven sampling of the Galactic plane, potentially over- or under-representing the observed numbers in specific regions. Therefore, these peaks should be interpreted with this survey limitation in mind."

- Classification: I strongly suggest to provide more references/details and explanations for scikit-learn, silhouette-score, Davies-Bouldin Index, Kmeans, seaborn, F1 Macro Average (and other very technical details that I might forget), all of which may be unknown for readers that are not familiar with such methods and tools.

Reply: Done. We have added the requested explanations and references to improve the **clarity.** of the machine learning alrogith used in this work.

- Classification: a figure showing the relation among silhouette score and the davies-bouldin index as a function of n_neighbours and n_components would help the reader to understand how a choice for the best values of 50 and 2 was made.

Reply: In response to ~~your suggestion~~ this comment raised by the referee, **we have added a figure (Figure 15) showing the relationship between the silhouette score and the Davies-Bouldin index as a function of n_neighbors and n_components. The figure consists of two subplots: one for the silhouette score as a function of n_neighbors for each value of n_components, and another for the Davies-Bouldin index as a function of n_neighbors for each value of n_components. These plots illustrate how these validation indices vary as the parameters are adjusted, justifying the selection of the chosen values for n_neighbors and n_components. The requested figure is included in the revised manuscript.**