North-PHASE: studying periodicity, hot spots, accretion stability, and early evolution in young stars in the Northern hemisphere

A. Sicilia-Aguilar [1]★† R. S. Kahar,[1] M. E. Pelayo-Baldárrago,[1,2,3] V. Roccatagliata,[4,5] D. Froebrich [6] F. J. Galindo-Guil [7] J. Campbell-White,[8] J. S. Kim [9] I. Mendigutía [10] L. Schlueter,[1] P. S. Teixeira,[1] S. Matsumura,[1] M. Fang,[11] A. Scholz,[12] P. Ábrahám,[13,14,15] A. Frasca [16] A. Garufi,[5] C. Herbert [6] Á. Kóspál[14,15,17] and C. F. Manara [8]

Affiliations are listed at the end of the paper

ABSTRACT

We present the overview and first results from the North-PHASE Legacy Survey, which follows six young clusters for five years, using the 2 deg^2 FoV of the JAST80 telescope from the Javalambre Observatory (Spain). North-PHASE investigates stellar variability on time-scales from days to years for thousands of young stars distributed over entire clusters. This allows us to find new YSO, characterize accretion, and study inner disc evolution within the cluster context. Each region (Tr 37, Cep OB3, IC 5070, IC 348, NGC 2264, and NGC 1333) is observed in six filters (SDSS griz, *u* band, and J0660, which covers Hα), detecting cluster members as well as field variable stars. Tr 37 is used to prove feasibility and optimize the variability analysis techniques. In Tr 37, variability reveals 50 new YSO, most of them proper motion outliers. North-PHASE independently confirms the youth of astrometric members, efficiently distinguishes accreting and non-accreting stars, reveals the extent of the cluster populations along Tr37/IC 1396 bright rims, and detects variability resulting from rotation, dips, and irregular bursts. The proper motion outliers unveil a more complex star formation history than inferred from *Gaia* alone, and variability highlights previously hidden proper motion deviations in the surrounding clouds. We also find that non-YSO variables identified by North-PHASE cover a different variability parameter space and include long-period variables, eclipsing binaries, RR Lyr, and δ Scuti stars. These early results also emphasize the power of variability to complete the picture of star formation where it is missed by astrometry.

Key words: techniques: photometric – surveys – stars: formation – stars: variables: general – stars: variables: T Tauri, Herbig Ae/Be – open clusters and associations: individual: Tr 37, NGC 2264, NGC 1333, Cep OB3, IC 5070, IC 348.

1 INTRODUCTION

Young stars are born in clusters and acquire most of their mass during the protostellar phase (Lada & Lada 2003). Due to angular momentum conservation, stars emerge surrounded by protoplanetary discs. For a few-Myr, the discs accrete onto the stars and eventually form planetary systems, their properties shaped by accretion. Both the angular momentum evolution of the system and the matter transport through the planet-forming region depend on the accretion process, channelled onto the star via the magnetosphere for most low-mass stars (Koenigl 1991; Hartmann, Herczeg & Calvet 2016). The star–disc interaction affects the stellar rotation, and the link between rotation and stellar activity (Bouvier et al. 1995) also means that accretion can influence the system's habitability. The accretion rate evolution also determines the types of planets formed (e.g. Matsumura, Brasser & Ida 2021) and, since the magnetospheric

disc truncation (Donati et al. 2019) may limit planet migration (Lin, Bodenheimer & Richardson 1996), the inner disc evolution is crucial to understand the orbital distribution of planets. Despite its importance, the main limitation to study the star–disc connection is that its spatial scales range from a few stellar radii (the size of the magnetospheres, e.g. Bouvier et al. 2007; Johnstone et al. 2014), to fractions of au in the inner planet-forming region, which are beyond imaging reach, especially for solar analogues, most of which are very faint.

Time resolution is a powerful tool to investigate tiny structures. In fact, *'using time to map space'* is the only feasible method to study small-scale processes and structures for statistically significant numbers of sources. Young T Tauri stars (TTS) are variable by definition (Joy 1945), so variability is one of the most reliable ways to find young stellar objects (YSO; Briceño et al. 2001). Variability in YSO is caused by accretion, occultations by circumstellar material (from the disc, accretion columns, or disc winds), and hot and cold spots on the stellars surface (Herbst et al. 1994; Bozhinova et al. 2016; Fischer et al. 2023; Herbert, Froebrich & Scholz 2023). The innermost disc structure and star–disc alignment can also be probed by track-

★ E-mail: a.siciliaaguilar@dundee.ac.uk
† Based on observations made with the JAST80 telescope at the Observatorio Astrofísico de Javalambre, in Teruel, owned, managed and operated by the Centro de Estudios de Física del Cosmos de Aragón.

ing extinction events and their periodicity (Roccatagliata et al. 2011).

The variability time-scales reveal the location of the physical processes that produce it (e.g. Lightfoot 1989; Eiroa et al. 2002; Hamilton et al. 2005; Sicilia-Aguilar et al. 2020). Structures at the stellar surface and up to the disc corotation radius vary due to stellar rotation on time-scales of days (Bouvier et al. 2007; Cody & Hillenbrand 2014; Rebull et al. 2020; Froebrich et al. 2021), while the inner disc time-scales depend on the Keplerian period at each radius, ranging from days to years (e.g. Fischer et al. 2023). Accreting stars show stronger variations and even outbursts (Froebrich et al. 2023), but activity, spots, and even occultations by circumstellar matter also happen in stars that have ceased to accrete (Scholz et al. 2019). Thus variability reveals the entire YSO population, unveiling the cluster structure independently of the astrometric properties of its members, and including objects missed by *Gaia*.

Most YSO studies so far concentrate on small areas due to field of view (FoV) limitations, but *Gaia* shows that even well-known clusters extend far beyond their classical limits (Meingast & Alves 2019; Röser & Schilbach 2019; Meingast, Alves & Rottensteiner 2021; Mendigutía et al. 2022; Tarricq et al. 2022) and often have multiple populations (Roccatagliata et al. 2020; Teixeira, Scholz & Alves 2020; Pelayo-Baldárrago et al. 2023; Szilágyi et al. 2023). While the periphery of young clusters is largely unexplored, star formation is found to be more complex than previously thought (Wright et al. 2022). The cluster outskirts are key to understand star formation (Smilgys & Bonnell 2017; Mendigutía et al. 2022) due to their different initial conditions, and their evolution and interaction with other systems (Jerabkova et al. 2021). The star–cluster connection is also critical for the evolution of planetary systems. The cluster environment (e.g. sparse versus dense regions, UV irradiation) affects the initial mass function (Hsu et al. 2013), the sizes of the discs (de Juan Ovelar et al. 2012; Eisner et al. 2018), the disc lifetime (Fang et al. 2013), and the planetary system properties (Kruijssen, Longmore & Chevance 2020). Finding astrometric outliers among clusters is also important to fully characterize disc evolution (Bertini, Roccatagliata & Kim 2023). Indeed, early interactions also played a role for the young Solar System (Kenyon & Bromley 2004; Mamajek et al. 2015; de la Fuente Marcos & de la Fuente Marcos 2018; Bailer-Jones 2022), so exploring a variety of complete young clusters is the best way to put the Solar System in context.

Here, we present the initial results of the North-PHASE project, using the first four months of data and one of the clusters, Tr 37, as a test bench to define and construct the tools needed for variability tracking, and investigating what variability surveys can add to our knowledge, even in very well-known regions. Section 2 describes the North-PHASE project. The data acquisition, calibration, and variability analysis tools are described in Section 3. The initial results on the benchmark cluster are presented in Section 4. The conclusions of this first part of the project, including a summary of results, overview of the legacy values, and future plans, are given in Section 5.

2 THE NORTH-PHASE SURVEY

The 'Periodicity, Hot Spots, Accretion Stability and Early Evolution in young stars in the Northern hemisphere' survey (North-PHASE) is a legacy survey aimed at *'using time to map space'* in young stars via variability. The project targets six clusters with stars aged 1–5 Myr (Tr 37, NGC 2264, NGC 1333, IC 348, IC 5070, and Cep OB3), containing altogether several thousands of YSO, using the

Javalambre Observatory 83 cm JAST80 telescope in Spain.[1] The clusters were selected to ensure a range of ages (from 1 to 5 Myr), a diversity of star forming histories (low-mass versus high-mass star forming regions) and for being visible from Javalambre for at least 7 months per year. North-PHASE started in 2023 February, and will continue until 2028 January, producing five years of time-resolved data that will be updated periodically. Our aim is to perform a large-scale, large-FoV, systematic survey to find YSO, explore their variability, investigate accretion over time and its dependence on stellar mass, obtain statistics on accretion outbursts, and their role in star and planet formation, and study the evolution of stars and discs within the cluster context.

The first advantage of North-PHASE is the 2 deg^2 FoV of the T80Cam camera,[2] which allows us to cover each cluster in one shot. The large FoV and variable cadence (being sensitive to variability from days to years) are critical to identify YSO in an unbiased way, including the clusters' outskirts beyond the 'classical' cluster boundaries. This provides an advantage in extension, depth, and spatial resolution compared to similar surveys with smaller FoV (such as HOYS; Froebrich et al. 2018, 2024). Fig. 1 shows the large FoV, plus a zoom in a high-extinction region to reveal the depth of the images. The pixel scale is 0.55 arcsec pixel^{-1}. In addition, by using five of the broad-band filters within the J-PLUS filter system[3] (*g*,*r*,*i*,*z* Sloan SDSS filters), the *u*-band filter (uJAVA), and the J0660 narrow-band filter (which covers from ∼6510–6690 Å, including the Hα line), we can reliably distinguish the processes happening in stars and discs. The six filters also give us an advantage compared to other variability surveys such as the Zwicky Transient Facility (ZTF; Bellm et al. 2019). The observations are planned with variable cadence, with the main cadence being one to three days during periods from 10 d to a month, and observations blocks spaced by a couple of months, depending on visibility, to target the key variability time-scales in young stars (e.g. Fischer et al. 2023). Details of the pointings are given in Table 1.

North-PHASE aims to achieve a complete census of YSO aged 1–5 Myr, down to 0.15–0.3 M$_\odot$, classifying variability types and mechanisms as well as measuring stellar parameters. The 5 yr of observations will allow us to measure how accretion and rotation are linked through this age interval for stars with different masses. Changes in accretion on these time-scales provide information on the stability of the star–disc connection, and the role of perturbations by planets or companions in the mass transport through the inner disc, and the variability of the magnetic field. Thanks to the *u*-band and Hα filters, we will trace accretion rates and how they vary. Combining the information on stellar rotation, spot-like variability, and extinction events, we will determine magnetosphere sizes and explore the connection between the star and the inner planet-forming regions and explore longer term changes in accretion related to changes in the inner disc and magnetic cycles in young stars.

In addition, North-PHASE will also provide very deep combined images and detect other types of variable field stars, so its legacy value expands beyond the star formation community to general stellar variability, stellar properties, and the requirements to use the JAST80 in Galactic observations. The combination of long and short cadences renders it well-suited for pinpointing fast-moving sources, such as objects within our solar system (Cortés-Contreras et al. 2019;

---

[1] https://oajweb.cefca.es/telescopes/jast80
[2] https://oajweb.cefca.es/telescopes/t80cam
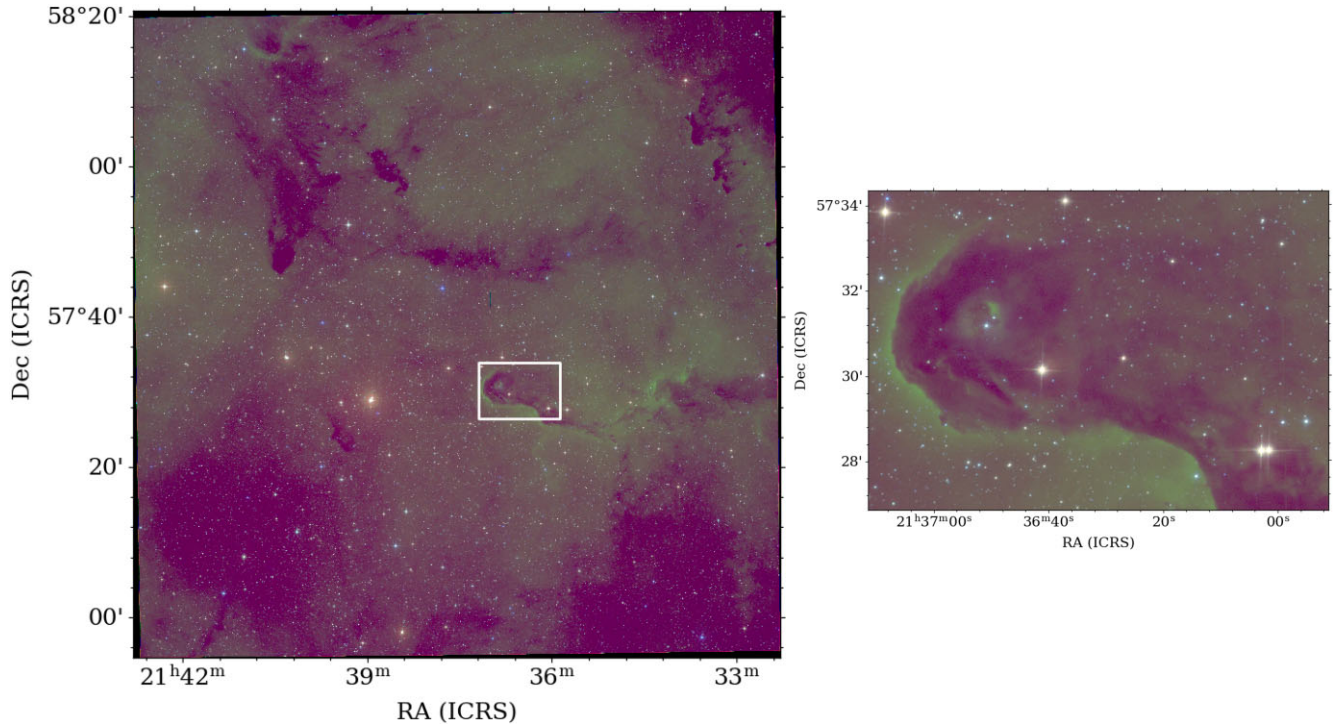[3] https://www.j-plus.es/survey/instrumentation

Figure 1. A view of Tr 37 according to North-PHASE in full (left) and zoomed around the white box surrounding the IC1396A globule (right). Red corresponds to zSDSS, green corresponds to J0660 (Hα), and blue corresponds to gSDSS. All images, shown in log scale, are combination of nine individual observations.

Table 1. Summary of the observations and current results for North-PHASE. Central coordinates, distances (estimated from *Gaia*) and ages (see references below) are given, plus filters and exposure times (in s) for each cluster as well as a summary including number of nights analysed, number of stars detected at least six times in the rSDSS filter, and variable stars of any type identified (following the methodology of Section 3.4, which requires variability in at least four indices and $g < 19$ mag). The uJAVA filter is only done on dark nights (about 1 out of every 5). Cep OB3 does not contain enough data to look into variability yet. The references for the cluster ages are [a]Richert et al. (2018), [b]Froebrich et al. (2021), [c]Pelayo-Baldárrago et al. (2023), and [d]Getman et al. (2012).

Cluster	RA	Dec	D	Age	t_g	t_r	t_i	t_z	t_u	t_{J0660}	Nr.	Nr. Stars	Dates	Nr. Varia. (four
	(ICRS)	(ICRS)	(pc)	(Myr)	(s)	(s)	(s)	(s)	(s)	(s)	Obs.	(rSDSS)	(JD, d)	indices)
IC 348	03:43:39	+ 32:17:53	313	2–4[a]	7	7	4	7	75	40	51	5 211	2459979–2460025	102
NGC 1333	03:30:50	+ 31:19:41	292	1–2[a]	7	7	4	7	75	40	54	4 983	2459979–2460025	59
NGC 2264	06:40:51	+ 09:23:20	730	1–3[a]	30	30	15	30	300	150	73	24 567	2459981–2460046	593
IC 5070	20:52:05	+ 44:06:39	800	1–3[b]	30	30	15	30	300	150	39	30 376	2460039–67	317
Tr 37	21:37:26	+ 57:38:50	925[c]	1–4[c]	30	30	15	30	300	150	31	89 680	2460040–67	894
Cep OB3	22:58:11	+ 62:41:06	825	2[d]	30	30	15	30	300	150	9	32 625	2459979–81	–

Racero et al. 2022). Finally, the large FoV and diversity of pointings as well as the number of visits per year (69) will also allow us to quantify the effects of satellite constellations trails on data sets such as ours, which has been done so far mostly using simulations (Hainaut & Williams 2020; Bassa, Hainaut & Galadí-Enríquez 2022).

3 OBSERVATIONS AND DATA CALIBRATION

3.1 Observations

For this first analysis of the North-PHASE programme, we studied the data acquired between the start of the programme in 2023 February and May. The observations consist of three exposures per night for filters *g*, *r*, *i*, *z*, and J0660 (Hα), acquired with variable cadence. On dark nights (about 20 per cent of the time), a further set

of three observations is recorded using the near-UV uJAVA filter. Due to a telescope glitch, some data sets had four observations. Exposure times depend on the filter, and are halved for the nearby regions IC 348 and NGC 1333 (see Table 1). All clusters are observed following a similar pattern but, due to their coordinates and visibility, at the time of this analysis there is still a very substantial difference between the number of data points available for each region, with NGC 2264 so far having over 70 data sets in filters grizHα, while this number is reduced to 9 for Cep OB3.

The three nightly repetitions are meant to achieve a deeper photometric coverage by combining the data. Used individually, the nearly simultaneous images allow us to test the data and relative calibration self-consistency, and may also reveal in the future very short time-scale variability in certain targets. In this work, we have explored the result of using the data points independently versus averaging them per day. Further details are discussed in Section 3.4.

3.2 Relative photometry calibration

The raw data products delivered to North-PHASE are similar to those provided for the J-PLUS survey,[4] including photometric catalogues and basic processed images. To produce these catalogues, the data are first reduced using the J-PLUS pipeline[5] (Cenarro et al. 2019; Bonoli et al. 2021). This includes the basic calibration of the images, including merging the eight CCDs of the T80Cam detector and performing standard bias, flat fielding, cosmic ray removal (van Dokkum 2001), distortion correction, and coordinate calibration. The pipeline then creates the processed catalogues with various types of photometry. Since the pipeline was designed for the detection of galaxies, SEXTRACTOR (Bertin & Arnouts 1996; Bertin et al. 2002; Bertin 2006, 2011) is first used to derive photometry, but additional aperture photometry is also given for apertures in the range 1–20 arcsec. The pipeline also provides an absolute calibration based on Pan-STARRS objects in the field and performed against the 6 arcsec aperture. Although useful for a first-look, the pipeline calibration is optimized for extragalactic fields with very low extinction, and it was found to have offsets as large as 0.2 mag from night-to-night in our clusters.

For variability studies, a robust relative correction that allows us to compare the magnitude variations from night-to-night is more important than an absolute flux calibration, as it already allows us to identify and classify the variable stars. Therefore, the main aim at this stage was to optimize the relative photometry calibration. We first selected the best aperture, 3 arcsec (nominally 2.8 arcsec), which, after inspection of the images and comparison of results, was found to offer the best compromise between light collection and avoiding overlap in crowded regions or near nebulosity. We then built customized PYTHON routines for the relative calibration and variability index calculation. Cross-matching between catalogues was done using the Starlink Tables Infrastructure Library Tool Set (STILTS), which improves code efficiency. STILTS was called within the same PYTHON program used to calibrate. In general, we use a radius of 0.5 arcsec and the TMATCHN subroutine for cross-matching of objects. The same procedure was used to match the North-PHASE catalogues to the *Gaia* DR3 survey in the analysis of Tr 37 (see Section 4). The present calibrations were done against the first observation in each set, which were good nights in terms of seeing and transparency. At this stage, this allows us to get the most of the still-limited data sets available, with the plan set to derive deeper calibrations based on many combined images (as done in Froebrich et al. 2018; Evitts et al. 2020) as the survey advances.

The relative calibration for individual nights follows from a sigma-clipping procedure (similar to Sicilia-Aguilar et al. 2004, 2008, 2013). For each filter and date, we extracted the 3 arcsec aperture photometry data, and performed a cross-match with the first observation. The data were clipped to remove dubious points and outliers, requiring a minimum of seven observations to be considered as a candidate for calibration. Further clipping was used to remove variable objects as well as those too faint or too close to the saturation limit, comparing with the magnitudes in the reference night.

We explored whether there could be differences in the calibration for fainter or brighter stars, but found none. Colour terms are negligible, as expected, since each filter is compared with itself. Therefore, the relative calibration was then derived as the average and standard deviation of the magnitude difference with respect to the reference line, for the clipped stars. The final relative magnitudes were derived by adding the corresponding nightly calibration offset to the 3 arcsec photometry. As a final verification, light curves of non-variable objects were examined, revealing no significant correlations. This was further checked in the analysis of Tr 37 (Section 4).

The final relatively calibrated catalogues were cleaned, removing spurious objects from cloud rims and bleeding patterns by requesting sources to be observed in at least five independent images. This may miss extreme variables or faint outbursting stars but, as any other very faint object, we will not be able to study those in detail until more data have been gathered. Future deep combined images will also help us to obtain a detailed catalogue of verified detections for these very faint targets. The resulting calibrated tables and variability indices are available via a dedicated OAJ catalogues web portal specifically adapted by CEFCA to the North-PHASE products presented here. It offers a user-friendly environment that permits a direct access to the database by means of Astronomical Data Query Language (ADQL) queries and based on the Virtual Observatory (VO) Table Access Protocol (TAP), among other services.[6,7]

As for the dynamical range of the survey, for the distant clusters we find that stars brighter than *Gaia* $g = 12.5$ mag tend to suffer from saturation, while those fainter than *Gaia* $g = 19$ mag are often much affected by noise in filters rSDSS, iSDSS, zSDSS, and J0660 (or after >17.5 mag, for gSDSS). Objects beyond both ends are excluded from the following analysis. The uJAVA filter is compromised at an even brighter stage, but we will use J0660 to study accretion variability, with uJAVA required to calibrate accretion at less-regular intervals. We also note that these limits refer to individual observations with the exposure times reflected in Table 1. Very faint objects are still detectable by combining the images, even though this reduces the number of time samples. As for addressing saturation, among intermediate-mass stars, we have modified the sequence from the second year of the survey on by including one short exposure out of each set, which strongly reduces the extent of the problem.

3.3 A flux calibration for Tr 37

Although the quality of the absolute calibration does not impair the identification of variable stars, for the purpose of validation and to establish the methodology to follow for the rest of the survey, we also performed an absolute flux calibration for Tr 37. An absolute calibration is essential for placing the sources in the HR diagram to measure stellar properties, as well as for deriving absolute accretion rates. The derivation of absolute stellar properties requires good understanding of the variability due to different phenomena (accretion, spots, extinction by circumstellar and interstellar material).

Since the relative calibration is based on the observations taken on JD 2460039.67 d, we also performed the absolute calibration on that data set. To calibrate the broad-band filters, we used data from Pan-STARRS1 (Chambers et al. 2016; Flewelling et al. 2020; Magnier et al. 2020), extracted with PYTHON.ASTROQUERY.[8] We found the filters to be very similar, and colour terms are only significant for zSDSS, for which a colour term on (r-i), chosen because of the lower level of noise for the red population, substantially improves the fit. For the narrow-band Hα filter J0660, we used data from the INT/WFC Photometric H-Alpha Survey of the Northern Galactic

---

Table 2. Calibration results for Tr37, relative to the master night JD 2460039.7 (Section 3.3). The calibration is given as a zero-point for sources without colour terms, while zSDSS has significant colour terms and thus a more complex transformation. The number of sources used in the fit is also given.

Filter	Correction (mag)	Nr. Sources
gSDSS	1.633 ± 0.014	5546
rSDSS	1.816 ± 0.015	9299
iSDSS	0.862 ± 0.021	10 377
zSDSS	$1.147(\pm 0.001) + 0.1052(\pm 0.0015) \times$ (r-i)	12 320
J0660	1.411 ± 0.039	9466

Plane (IPHAS) survey[9] (Barentsen et al. 2014). IPHAS also includes data on rSDSS and iSDSS, which we compared with Pan-STARRS1. In particular, for rSDSS, which has negligible colour terms, the correction agrees down to few per cent. The correction for iSDSS includes non-negligible colour terms, so it is not directly comparable. As a rule, we take the corrections derived from Pan-STARRS1 because it is also being used for the general Javalambre calibration, it is less noisy, the colour terms are lower, and they include all the four main broad-band filters.

The calibration of the uJAVA filter is critical to determine the accretion rate (Gullbring et al. 1998). For Tr 37, we used the Johnsons *U*-band data from Sicilia-Aguilar, Henning & Hartmann (2010), which had been acquired with the Calar Alto 3.5 m telescope using the Large Area Imager for Calar Alto (LAICA) camera.[10] Since the uJAVA from the North-PHASE survey and the Johnsons *U* filter used by LAICA are significantly different, the colour terms are large and we found that the currently available data sets did not have enough depth to guarantee sufficient sources for calibration. We therefore leave the calibration of uJAVA aside for now, to be improved with further knowledge of the photometry of individual stars as well as absolute calibrations obtained with the JAST80 telescope.

The results of the calibration are presented in Table 2. Saturation for Tr 37 members occurs for $r < 13.3$ mag, and the noise level tends to make it very hard to determine variability for objects fainter than $r = 19.3$ mag. For Tr 37 members with extinction $A_V \sim 1.4$ mag (see Pelayo-Baldárrago et al. 2023, from now on PB23), this ranges from massive T Tauri stars (spectral types F-G) down to M3-M4 types, or a mass of about $0.3\,M_\odot$, although gSDSS is more limited ($\sim 0.4\,M_\odot$). This is slightly better than the planned final detection limit of $0.3\,M_\odot$ at 900 pc and $A_V \sim 1.5$ mag. While the sensitivity in the lower mass end will increase with combined observations in the future, the fact that most A-type stars are already at least partly saturated is an issue, and it will be mitigated in the future by halving the exposure time of one out of each of the three repetitions.

3.4 Variability analysis and variability indices

Variability indices offer a quick way to detect different types of variable stars based solely on relatively calibrated data. The core of the North-PHASE programme is detecting stellar variability and using it to classify stars. The first step in this process is to create the light curves for each source, which will be later analysed via variability indices. The data for each source were combined using STILTS, incorporated into the PYTHON programs used to calculate the variability indices. The data were further cleaned, removing objects

that have not been detected a minimum number of times (set to 5) to remove contaminants such as cloud structures as well as mistaken detections that appear on the bleed tracks of saturated objects. All objects were also assigned a set of 'working coordinates' which, for each band, corresponded to the first data set where the star was detected. The data sets are then considered on a band-by-band basis, or matched to another band.

We explored several variability indices and filter combinations, using the known members in Tr 37 to test them and to set the limits of what should be considered as variable while minimizing inclusion of non-variables (more details are given in Section 4). The main indices we considered were the Stetson indices (defined for pairs of filters, Stetson 1996) and a modification of the simple variability indices from Sicilia-Aguilar et al. (2004).

There are various ways to define the Stetson index (see Stetson 1996). The efficiency of the index to distinguish variables is similar, although the thresholds obviously differ. For each pair of filters a and b, the Stetson index S_{a-b} can be calculated as

$$S_{a-b} = \frac{1}{\sqrt{N(N-1)}} \sum_{i=0}^{N} \text{sign}(P(a-b)_i)\sqrt{|P(a-b)_i|}, \quad (1)$$

where

$$P(a-b)_i = \frac{(m_{a,i} - \overline{m}_a)}{\sigma_{a,i}} \frac{(m_{b,i} - \overline{m}_b)}{\sigma_{b,i}}. \quad (2)$$

Here, $m_{x,i}$ represents the i-th magnitude in filter x, $\overline{m}_x$ is the average magnitude for the same filter, $\sigma_{x,i}$ is the uncertainty in the i-th measurement, and N is the number of observation pairs. There is no rule as to how the filters can be paired, so we tried all combinations of the broad-band filters except uJAVA. We explored pairing individual observations, that is, joining each one of the (typically 3) repetitions of each filter with the corresponding repetitions of the other filter separately, for each one of the nights, as well as using the average of the nightly repetitions. In this case, the uncertainty is taken to be the standard deviation of each set of measurements. We found that the second method offered a better estimate of the uncertainties, and was thus adopted as the standard for the survey's Stetson indices.

The simple variability indices defined by Sicilia-Aguilar et al. (2004) were based on counting the number of data points that exceeded the uncertainty threshold for a given magnitude as 'variable', and using the percentage of 'variable' points versus the total to assess the variability of a source. For North-PHASE, we introduced a modification, estimating the Gaussian probability of each measurement, assuming that the measured values are normally distributed around the mean with σ equivalent to the typical uncertainty for each magnitude range. This variant of the index has the advantage that it will give more weight to objects that have very significant deviations on a small number of nights, which were missed by simple counting and could affect rapid, irregular variables (e.g. Sicilia-Aguilar et al. 2008). The improved SA04 indices also allow us to better detect sources that have sustained, lower level variability, such as those with small and/or cold stellar spots and rotational modulation. To estimate the magnitude-dependent typical uncertainty for each night (σ), we followed Sicilia-Aguilar et al. (2004), binning the nightly data set in magnitude intervals, and estimating the typical spread compared to the average magnitude. A 3-degree polynomial function was then fitted through the bins to derive the 'uncertainty curve' as a function of magnitude. Each individual magnitude measurement was then compared to the uncertainty curve for its night/data set, so that the total deviation for the magnitude in filter a for the i-th night

$(m_{a,i})$ is

$$\Delta_{a,i} = \frac{|m_{a,i} - \overline{m}_a|}{\sqrt{2} \times \sigma_{\text{fit},a,i}(m_{a,i})}, \qquad (3)$$

where $\overline{m}_a$ is the average magnitude for that star, and the $\sigma_{\text{fit},a,i}$ function gives the expected (fitted) uncertainty for band a on night i, applied to the magnitude $m_{a,i}$. The $\sqrt{2}$ factor is simply the normalization, so that the deviation for each night will be given by the error function applied to $\Delta a, i$, and the final variability index is the normalized sum of all the deviations, i.e.

$$SA04_a = \frac{1}{(N-1)} \sum_{i=0}^{N} erf(\Delta_{a,i}), \qquad (4)$$

where N is the total number of observations for that star in a band ($N>5$ due to our requirements). The calculation was done for all broad-bands (griz) plus Hα (J0660). We note that the SA04 indices have a slight magnitude dependence, which is normally caused by the polynomial fit overestimating σ at the brightest end. If any, this tends to make bright variable stars slightly harder to identify, an issue that we will explore in more detail in the future, with more observations.

As a first test, Fig. 2 shows the difference in index distribution for young stars compared to all detected objects in the field for Tr 37 and NGC 2264, even when a single index is used. For Tr 37, we selected the YSO from PB23, while for NGC 2264, we selected those that are listed as cluster members by the SIMBAD data base. Although both clusters have different ages and different variability levels as well as a different number of YSO, the distributions of indices for YSO versus the field population are very different. We also observe that YSO tend to show a bimodal distribution in the SA04 indices, with objects peaking at ∼0.3 (no variability) and ∼0.7. As we will see in Section 4.2, this peak is a key in the variability behaviour of YSO and can be explored to distinguish them from other, non-young variables. Conversely, we find that even if only one index is considered, the distribution of parallaxes for the variable sources is clearly peaked at the clusters distances (see Fig. 3), a further confirmation that the variability indices are suitable for identifying young stars.

Table 1 shows the initial results regarding numbers of nights available for each cluster within this preliminary data set, as well as the number of stars detected (requiring observations in at least five nights to call it a detection), and the number of variable stars.

4 A CASE STUDY: TR 37

The Tr 37 cluster (Loren, Peters & Vanden Bout 1975; Platais, Kozhurina-Platais & van Leeuwen 1998; Contreras et al. 2002) is one of the best-studied young-star forming regions. It has been targeted by stellar variability studies (with a far smaller scope than North-PHASE, e.g. Sicilia-Aguilar et al. 2004; Morales-Calderón et al. 2009; Meng et al. 2019), near-infrared 2MASS and *Spitzer* surveys (Reach et al. 2004; Froebrich et al. 2005; Sicilia-Aguilar et al. 2006b; Morales-Calderón et al. 2009; Rebull et al. 2013), X-ray (Getman et al. 2007; Mercer et al. 2009; Getman et al. 2012), spectroscopy (Sicilia-Aguilar et al. 2005, 2006a; Nakano et al. 2012; Sicilia-Aguilar et al. 2013; Pelayo-Baldárrago et al. 2023), far-IR Herschel data (Sicilia-Aguilar et al. 2015), and *Gaia* (PB23). Hα photometry (Barentsen et al. 2011) and near-UV data (Sicilia-Aguilar et al. 2010) have been used to measure accretion rates.

This large collection of data have been used to constrain the properties of hundreds of YSO (e.g. the total number of members and likely members from the literature is 1536; PB23), making it an ideal place to optimize our variable-detecting algorithms. Of the 1872

members and member candidates listed by PB23, including their new astrometrically identified ones, 1489 are within the North-PHASE field.[11] Of those, 626 are detected in all filters within the established good-quality limits, and 315 among them have good *Gaia* DR3 data (defined as those with a Re-normalized Unit Weight Error, RUWE, below 1.4 for a reliable astrometric solution;[12] Roccatagliata et al. 2020; Lindegren et al. 2021). This last point is the most restrictive one and, since a large majority of the objects are faint, is satisfied by only 16 878 objects out of 37 882 sources with at least five observations per filter (excluding uJAVA). The following sections explore the best ways to identify variable stars and discuss the properties of the new variable YSO, exploring for this purpose all objects with reliable *Gaia* astrometry and ancillary information regarding the presence of protoplanetary discs, accretion, and activity. For the time range considered, the observations of Tr 37 span 47 d, and we are excluding the uJAVA filter (with too few data sets) from the analysis.

4.1 Examining the accuracy of the variable identification

We first tested the capabilities of North-PHASE for detecting young and variable stars, which requires a uniform sample of members with negligible contamination. To make sure that our analysis was not misled by objects of dubious membership, we first filtered a group of high-confidence members from the original list from PB23. In addition to requesting RUWE<1.4, parallax limits between 0.6 and 1.6 mas were also enforced. This resulted in 321 objects, of which 71 had been identified exclusively by astrometry, among which 315 have data in all North-PHASE bands. This group composes our low-contamination sample to test the variability criteria. We note that most bona-fide members (confirmed by spectroscopy and IR excesses) do not fall in this category due to the stringent limits, but the main aim with this part is to test the methodology. Studying the variability of the known members as a whole will come as a second objective, once the criteria are well-established.

Similarly, we extracted a collection of probable non-members from the initial potential member list, composed of those with a good astrometric solution but parallax outwith the member range. This is used as a control group for variability purposes, since they are in general stars that share many characteristics with the high-probability members in terms of magnitudes and colours (and hence were at some point suggested to be cluster members), but are most likely unrelated background or foreground objects. There are a total of 134 sources in this class.

We first checked the correlation between indices for YSO and sources that have been rejected as cluster members. The results are shown in Appendix A, Fig. A1. Besides showing a clear difference in variability indices between members and non-members, it also reveals a strong correlation between different Stetson and SA04 indices, as expected if they accurately trace variability. The correlation is not linear, which is good because it means each index is more sensitive to different types of variability.

We further compared the distribution of objects in various classes, including cluster members confirmed by astrometry versus those from the literature, the probable non-members (as defined by PB23), and the field population. The analysis is shown in Fig. 4. Looking at the variability indices, the new astrometric members from PB23 share

---

[11]Some YSO in the extended IC 1396 clouds (Barentsen et al. 2011; Nakano et al. 2012; Rebull et al. 2013) fall out of the North-PHASE FoV.

[12]In the subsequent sections, 'good astrometric data' corresponds to this RUWE<1.4 criterion.
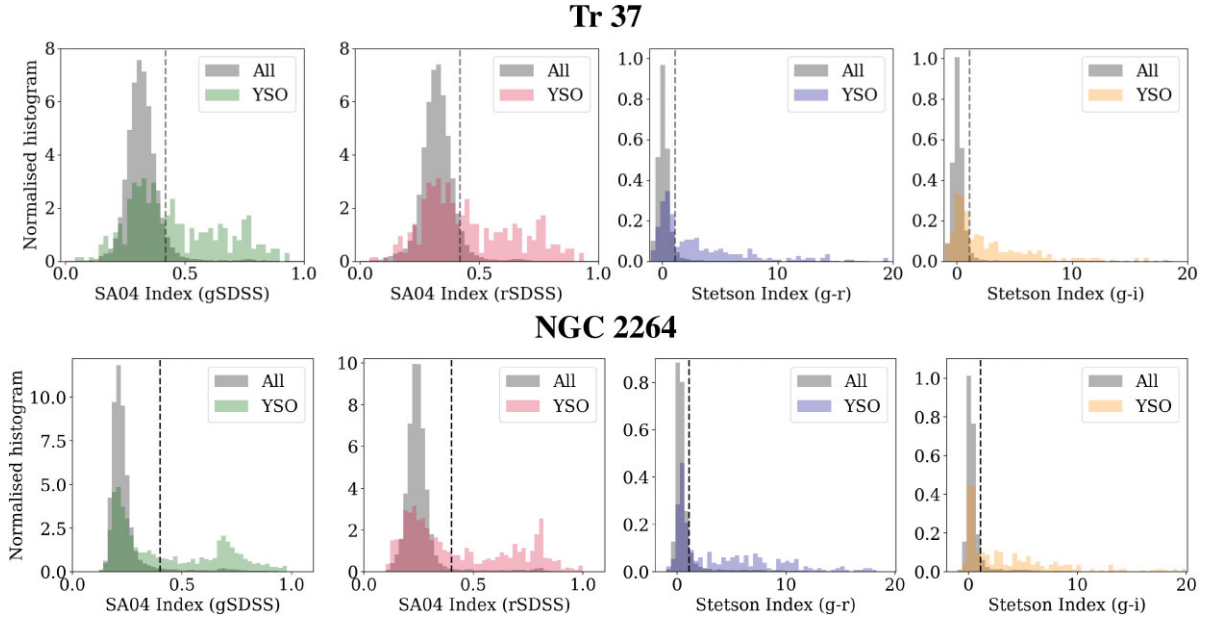
Tr 37

NGC 2264

Figure 2. Normalized histograms showing how the SA04 and Stetson indices behave differently for known members versus the general field stars. Top: results for Tr 37. Bottom: results for NGC 2264. The YSO were taken from PB23 for Tr 37 (see also Section 4), while for NGC 2264 we use those listed as cluster members by SIMBAD. In both cases, they are compared to the bulk of the population detected in each filter, including members and field stars. The SA04 indices tend to better characterize variables in the clusters even if the observations are partial, although the Stetson indices have the advantage that they are less magnitude-dependent. We observe differences in variability between clusters but, in general, the results for other clusters/other filters are very similar.

Tr 37

NGC 2264
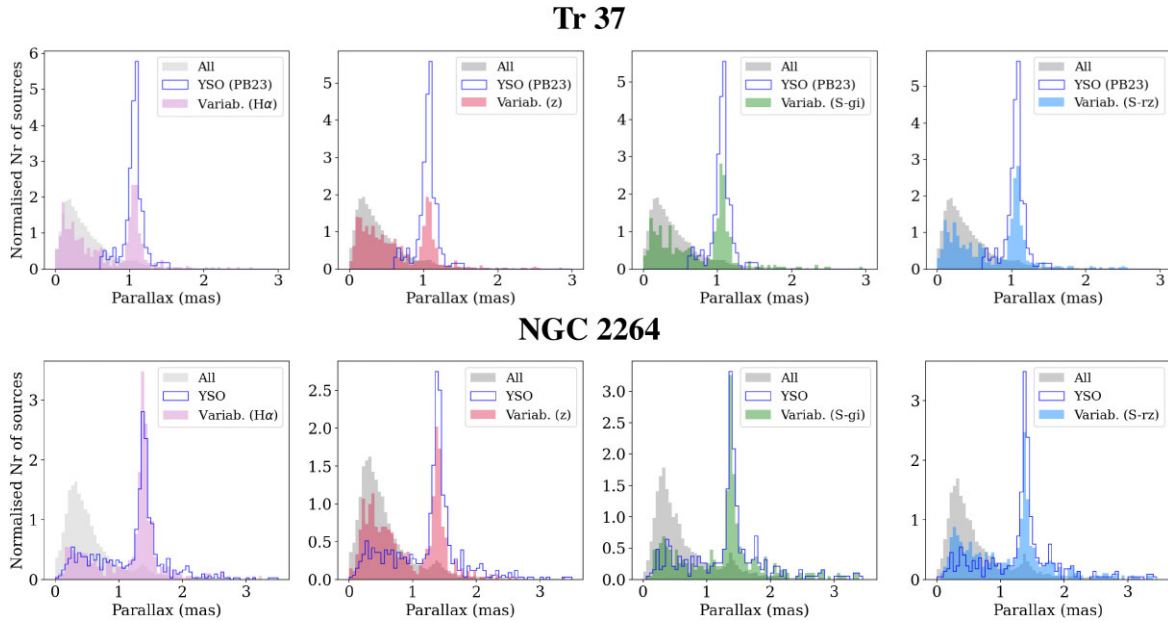
Figure 3. Normalized histograms showing the parallax distribution for stars classified as variable, using a single index, compared to the bulk of field stars, and to known YSO. We set the limits for variability at >0.42 for SA04 indices, and >1.1 for Stetson indices. The top panels show the results for Tr 37, using the YSO compiled from the literature by PB23. The bottom panels show the results for NGC 2264, with the YSO from SIMBAD. Note that, as shown, both YSO samples contain a non-negligible contamination but even so, for both clusters, the peaks at the cluster parallaxes of ~1.1 and 1.5 mas are evident.

the typical variability index distribution of the confirmed literature members and are clearly different from the probable non-members. This verifies a low contamination in the sample of astrometric candidates from PB23, as we do not see differences in the indices of the general collection of known members (identified by multiple methods, including spectroscopy, X-ray, and the presence of IR excesses) versus those identified by *Gaia*. A further insight in the

differences in variability for various types of objects is given in Section 4.3.

4.2 Identifying and classifying new variables

After verifying that variability is correlated with youth and membership, we moved onto identifying other young, variable stars in the
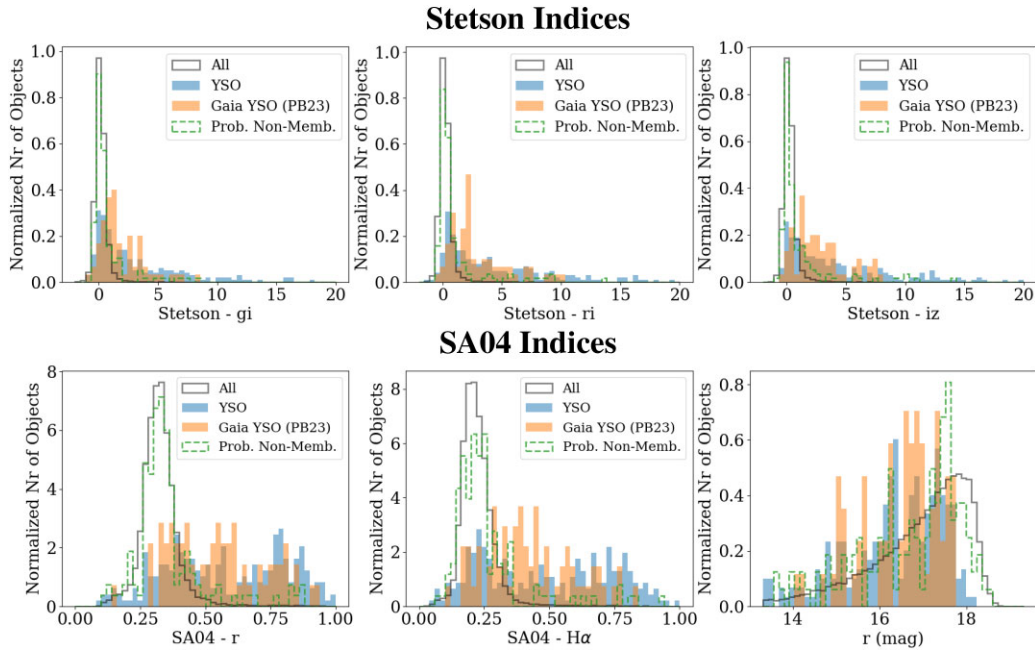
Figure 4. Normalized histograms for the variability indices for Tr 37, including the field population, the highly reliable members from the literature (YSO), the objects selected by PB23 as astrometric members, and the probable non-members. The differences between YSO and field stars were used to define the variability criteria. While the objects classified as probable non-members are very similar to the field stars, astrometric members are similar to those found by other means. The last panel of the SA04 indices block shows the rSDSS magnitude range.

region. We first constructed histograms of the parallaxes for objects identified as variables via individual indices. To define the criteria for variable stars, we started with the most restrictive selection. After eliminating sources that are close to the saturation or noise limits, we also selected those with good *Gaia* DR3 data (Gaia Collaboration 2016, 2023). By looking at the distribution of all objects, we observe a clear difference in behaviour for Stetson indices >2 and SA04 indices >0.5. This difference is further enhanced by requesting objects to be variable in at least two indices of each class and having a minimum of 20 observations in all filters, which allows us to use less stringent limits for similar results, namely Stetson>1.1 and SA04>0.42. The Hα filter displays a slightly different behaviour and is likely sensitive to higher degrees of variability, but since the difference is small, we are at present treating it in the same way as the rest until more data becomes available. We note that the requirement of excellent *Gaia* DR3 data is failed by a significant number of confirmed members (see PB23). Nevertheless, our aim here is to test the method, for which having reliable *Gaia* data is an invaluable asset. The requirements about the *Gaia* data will be relaxed in the future when more data allows us to determine stellar parameters and light curves, adding further information to distinguish YSO from other types of variables.

Fig. 5 examines the location in the HR diagram, parallaxes, and proper motions for known members compiled from the literature (PB23), as well as new variable stars found by North-PHASE. The initial collection of variable stars was further divided into candidate YSO when positioned between the 0.5 and 20 Myr isochrones in the colour–magnitude diagram (using the CMD isochrones; Bressan et al. 2012; Marigo et al. 2013) as well as parallax within 0.8–1.6 mas. The isochrone limits are chosen because they contain most of the confirmed Tr 37 members (PB23), and the parallax limits respond to the fact that the background variables and the cluster parallax peak merge at around 0.8 mas, and objects below this limit may have a similar chance to be cluster members or contaminants. We also mark

those that are variable but have parallaxes outwith the cluster range, with particular attention to those that fall in the same region of the colour–magnitude diagram, which we label 'HR contaminants'. Objects like these will be the bulk of the contaminants in the literature, since they are harder to distinguish from genuine YSO. Other variable stars include older systems below the isochrones, and background giants. To explore their variability indices, we consider the average index for each star, including all the available combinations, as well as the standard deviation per object.

Examining the different groups of variable stars reveals that the main sources of contamination are background giants and AGB stars, which are also often variable. Although giants and YSO tend to cover similar index ranges for both the Stetson and the SA04 indices, older variables (below the 20 Myr sequence) and 'HR contaminants' have a different index distribution. As a note, we highlight that PB23 identified some genuine YSO both below the 20 Myr isochrone as well as over the 0.5 Myr one, but also that contamination by field stars is stronger there. While keeping in mind that the selection at this stage is quite restrictive, further study of complete SEDs and accretion signatures will help us to find discrepant sources. A double-sided Kolmogorov–Smirnov test (KS test) gives a probability below 10^{-11} for YSO and either older variables or 'HR contaminants' to be drawn from the same index distribution, for both indices. The case for giants is significant for the Stetson average index (0.2 per cent), albeit only marginal for SA04 (6 per cent). Considering the standard deviation of the indices reveals a larger difference between YSO and giants, rejecting the null hypothesis at a level of 10^{-9} for the SA04 indices, and at 0.02 per cent for the Stetson average index.

Significant differences also arise when considering individual indices, as shown in Appendix A, Fig. A2. We also see that the second peak in the SA04 indices is more prominent for YSO in the red bands, and that YSO tend in general to have more extreme variability indices. Although processes such as accretion and extinction are expected to
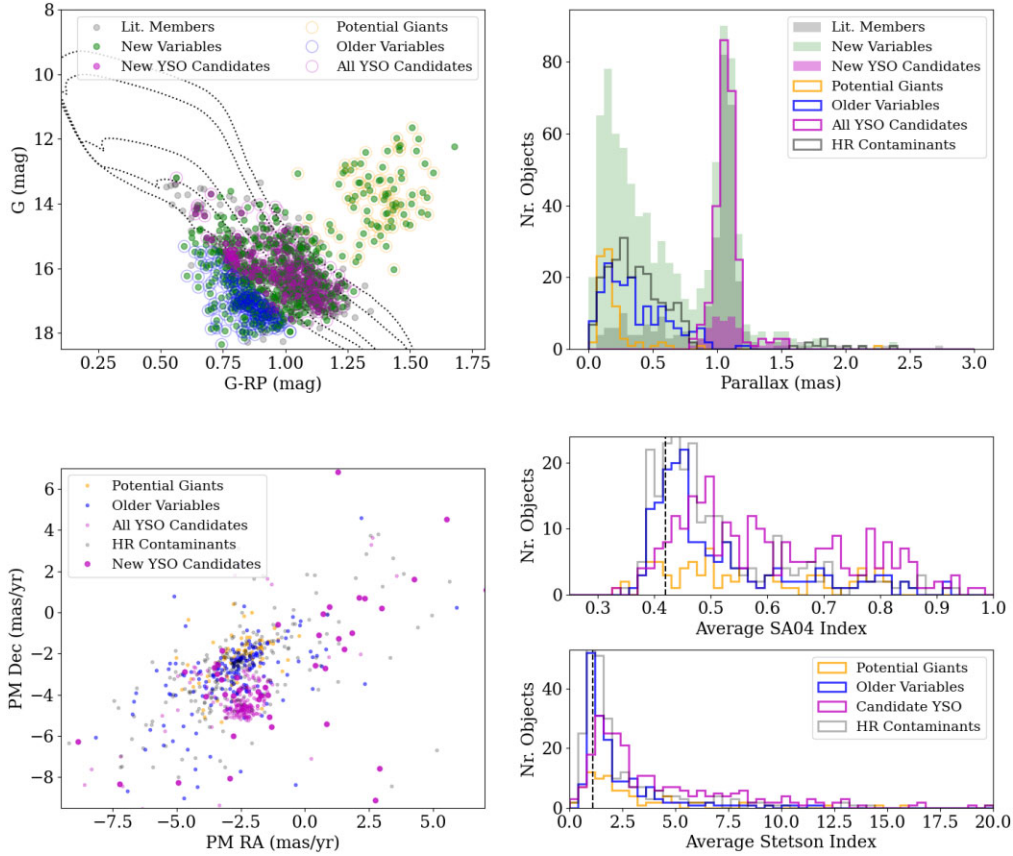
Figure 5. A study of the known and new variable stars. The top left panel show the colour–magnitude diagram for the *Gaia* filters with the CMD isochrones for ages 0.5, 1, 5, 10, and 20 Myr (corrected by cluster distance and reddening following PB23). Members and candidate members derived from the literature are shown in grey, while new variables are displayed in green. Among the new variables, we use the isochrones to select potential 'older' field variables (blue) as well as variable giants (yellow), and use a combination of isochrones and parallax to extract the candidate YSO (magenta). The 'HR Contaminants' have parallax outwith the cluster limits, but fall within the same isochrones than the YSO. The top right panel and bottom left panel show the parallax histogram for the different types of objects, and the proper motions in RA and Dec, respectively. The bottom right smaller panels shows two histograms for the average SA04 and Stetson indices, with the limits for variability shown as dashed lines. To highlight the astrometry of the newly discovered YSO, they are shown with filled histogram in the parallax plot and with larger symbols in the proper motion panel. Note that parallax and proper motions of field variable sources and giants are clearly different and that older variables and HR contaminants have significantly lower variability indices. The proper motion plot is zoomed in so that the detail of the substructure can be appreciated, there are a few high proper motion objects of all kinds beyond the limits we show. The Stetson index is also truncated to 20 for the same reasons. Fig. A2 presents further information regarding individual indices within each class.

produce stronger variations in the bluer bands, the fact that noise is higher in gSDSS for typically-red YSO may play a role in this. HR contaminants show narrower peaks at low index value, especially in the bluer bands. Although at this stage it is too early to define alternative criteria to separate young variables and the rest based on variability alone, the results are promising and suggest that a combination of indices can be used to assign YSO probabilities to objects for which further information and astrometric data are poor or unavailable.

After cleaning the new variable sample using the isochrones, we are left with 50 new members (to add to the 321 literature members that satisfy the strong selection criteria). For a cluster that has been so thoroughly studied as Tr 37, a member increase by 16 per cent is significant. Among the literature population, 146 stars fall within the variable class. Note that no selection has been done here regarding the parallax of the literature members, hence when clustering is considered (Section 4.3), we find higher rates of variability for clustered sources. Due to the detailed work of PB23 identifying new members by astrometry, we find that the YSO identified by variability tend to be consistent with the sparse subclusters identified

by PB23, or are high proper motion sources, similar to walk-away stars (Zhou et al. 2023) whose understanding is key to test the effects of environment on star formation. The dearth of new members identified within the Tr 37 main cluster proper motions indicates that the member search by PB23 was very efficient, which is also corroborated by the analysis of the variability of the new members (Fig. 5).

We also demonstrate the importance of variability to unveil complete YSO populations. In the same way that using *Gaia* astrometry addressed YSO typically missed by other surveys (e.g. intermediate-mass stars and stars without discs or accretion signatures; PB23), using variability further fills in the gap adding objects that do not share the same astrometry than the bulk of cluster members, such as proper motion outliers.

4.3 Variability and cluster structure

We analysed whether the subcluster structure of Tr 37 as stated by PB23 changes when considering the new variable YSO, as well as the clustering of objects depending on their variability. As shown in
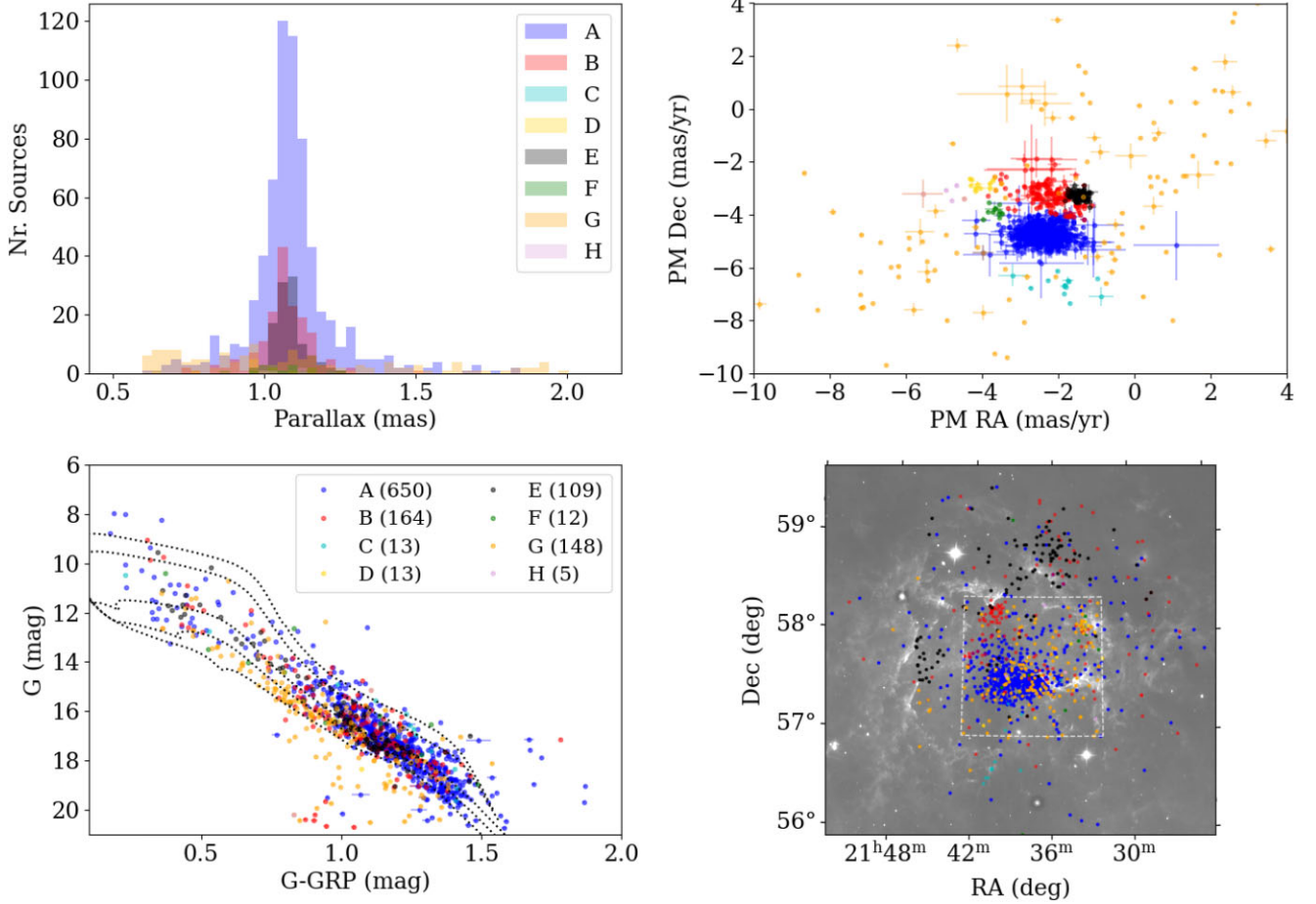
Figure 6. Subclusters in Tr 37, including the members from PB23 and the new variables from North-PHASE. We show the parallax distribution (top left), proper motion (top right), a colour–magnitude diagram (including CMD isochrones for ages 0.5, 1, 5, 10, and 20 Myr corrected by $A_V = 1.5$ mag, bottom left), and ICRS coordinates on the *WISE* mosaic for band W3 (as in PB23), with a white box to mark the North-PHASE FoV (bottom right). The colour scheme follows PB23 for subclusters A-G, adding subcluster H, a new group with low significance. Objects with a probability of at least 40 per cent of belonging to a subcluster are shown, to include those that have intermediate probabilities (hence the number of objects is different from those in Table 3). Note that the proper motion plot is zoomed in so that the detail of the substructure can be appreciated, there are a few high proper motion objects of all kinds beyond the limits we show.

Fig. 5, the new YSO span a rather broad region in the proper-motion plane, so they add valuable information to investigate distinct star-formation groups and episodes. We followed the methodology of Roccatagliata et al. (2020), based on Lindegren, Madsen & Dravins (2000) and also used by PB23. This involves a maximum-likelihood analysis of the tridimensional parameter space including parallax and proper motions in right ascension and declination. Subclusters in different parallax and proper motion ranges were tested against the whole population, and we considered as final subcluster structure the one leading to the higher likelihood. More details on the procedure are given in Appendix B.

For Tr 37, we did two tests. First, we ran the algorithm from PB23 including the new variable stars identified as cluster members, and then, independently, we used the same technique to examine the distribution of all variable stars in astrometric space, independently of whether they are cluster members. The first part helps us to see what the inclusion of variable stars detected over a large FoV can tell us about the global cluster structure. The second part is aimed at improving our understanding of how the astrometry can be used to separate different types of variables and whether YSO and field variables (as grouped by the algorithm) show differences in their variability indices.

4.3.1 Clustering analysis for all Tr 37 members

To study the combination of known and new variable YSO, as in PB23, we require the objects to have good astrometric data and imposed the same parallax cut (0.6–1.6 mas). Since our aim is to see the distribution of stars rather than to improve the subcluster parameters in the search of new members, we did not impose any magnitude or uncertainty restrictions, so that the algorithm was run over a total of 1094 stars, in contrast with the 541 high-quality ones selected by PB23. As in PB23, we performed several iterations with various grid parameters (see Appendix B for details), settling for a 43-point grid $\pm 3\sigma$ around the subclusters, and using the results from PB23, plus 3 more arbitrary subclusters, as initial trial. Non-significant subclusters are removed by the algorithm. As shown in Fig. 6 and Table 3, we recover all the six subclusters (A-F) identified by PB23, including the low-significance groups C and F, plus a similar extended population (G, which likely includes a higher degree of contamination than the rest), and a new subgroup, H, with only five members and thus low significance. The astrometric parameters of all the previously known subclusters are consistent, within the uncertainties. These results highlight the robustness of the cluster structure found by PB23, considering that most of the new YSO have

Table 3. Astrometric properties and number of objects for the subclusters identified among the entire YSO population of Tr 37, including stars identified by PB23 and new variables found by North-PHASE. The parallax (ϖ) and proper motions are given as the average and standard deviation for each group. The number of objects in each subgroup is restricted to those with a probability >80 per cent (thus not including objects that are consistent with more than one subcluster). For the PB23 results, we give the number of objects first identified among the members, plus the number of new members identified by their astrometric properties using the Mahalanobis distance (given in parenthesis) for the clusters used to identify new astrometric members (A,B,E,F only). Column 'Var. Fraction' lists the fraction of objects classified as variable versus the total number of objects with good quality North-PHASE data for each subcluster (in parentheses).

Subcluster	ϖ_{PB23} (mas)	$pm_{RA,PB23}$ (mas yr^{-1})	$pm_{Dec,PB23}$ (mas yr^{-1})	ϖ (mas)	pm_{RA} (mas yr^{-1})	pm_{Dec} (mas yr^{-1})	N_{PB23}	N_{NP}	Var. Fraction	Comments
A	1.101 ± 0.130	-2.432 ± 0.412	-4.719 ± 0.318	1.090 ± 0.144	-2.382 ± 0.413	-4.724 ± 0.304	418(+166)	625	0.68 (253)	–
B	1.098 ± 0.128	-2.251 ± 0.360	-3.131 ± 0.252	1.074 ± 0.093	-2.238 ± 0.433	-3.220 ± 0.341	33(+80)	121	0.84 (51)	–
C	1.002 ± 0.035	-1.835 ± 0.091	-6.603 ± 0.241	1.013 ± 0.041	-1.884 ± 0.421	-6.585 ± 0.222	7	10	0 (1)	–
D	0.963 ± 0.090	-3.876 ± 0.260	-2.954 ± 0.301	0.973 ± 0.151	-4.080 ± 0.218	-2.865 ± 0.148	9	10	0.86 (7)	–
E	1.075 ± 0.060	-1.374 ± 0.133	-3.424 ± 0.201	1.078 ± 0.046	-1.433 ± 0.110	-3.239 ± 0.101	14(+80)	87	0.75 (8)	–
F	1.092 ± 0.085	-3.632 ± 0.121	-3.883 ± 0.143	1.077 ± 0.042	-3.724 ± 0.152	-3.914 ± 0.131	7(+3)	6	1.0 (3)	–
G	0.997 ± 0.229	-1.909 ± 5.159	-3.124 ± 4.192	1.086 ± 0.379	-1.250 ± 5.648	-2.817 ± 4.496	64	139	0.55 (96)	Non-clustered
H	–	–	–	1.019 ± 0.078	-4.816 ± 0.109	-3.128 ± 0.238	–	3	0.67 (3)	Low sign.

significantly different proper motions compared to the previously known population, while also reinforcing the subclusters that had a lesser representation in their study. Since the member list from PB23 includes objects identified by astrometry (their 'new members') from subclusters A, B, E, and F, a relatively large increase in objects in those groups is expected independently of the variables. We also note that some stars have large probabilities of belonging to more than one group (e.g. ~40 per cent for two subclusters). Such objects are not counted as members of any in Table 3, which is capped to at least 80 per cent probability, but they are still considered as cluster members and plotted in Fig. 6.

The inclusion of the variables also strengthens the membership of several of the groups surrounding the main cluster, in particular, B, E, and F, and reveals further patterns within the extended population, G. Part of this extended population appears associated with the cloud rim in the north-west (see orange dots in the fourth panel of Fig. 6), but we do not find any particular distribution in proper motion for those objects. While the proportion of new variable YSO in the different subclusters is similar, we find a larger number of young variables among the extended population, G. Objects with discrepant proper motions are impossible to identify as cluster members via astrometry, a sign that variability studies are a key to fully bring into light cluster members in an independent way, and thus critical to study cluster interactions and the episodic nature of star formation.

Table 3 also contains information on the number of sources that were found to be variable, according to the criteria from Section 4.2. We find a very high percentage of variability among all subclusters with a significant number of members. Typically, at least 2/3 of objects are found to be variable in each subcluster, with the variability rate being higher for sources associated with the globules and bright rims (e.g. B, E), which are also expected to be younger. No difference is found between objects detected by astrometry (PB23 'new members') and those found by other means, a sign that contamination by proper motion and/or parallax interlopers is likely negligible. The extended population, G, also contains a significant, albeit lower, number of variables (55 per cent), suggesting that, even if contamination may be stronger within this group and some contaminants may be also variable stars (e.g. giants), there is still a significant number of YSO with anomalous proper motions in the cluster.

4.3.2 Clustering analysis for variable stars

For the second part of the analysis, we repeated the clustering analysis for all the variable stars with good *Gaia* data (769), without imposing distance or age constraints. One of the objectives here is to determine whether the astrometric clustering can also distinguish different types of variable stars. We ran the maximum-likelihood algorithm starting with the clusters defined by PB23 plus two further subclusters to account for the background variables. The results are presented in Fig. 7 and Table 4. We independently recover the selection based on the colour–magnitude diagram from Fig. 5, identifying the population of variable background giants and older main-sequence stars, as well as the basic structure of the region described by PB23. Subcluster E is not found, probably because many of its members are located out of the North-PHASE field, while those of the eastern rim are now set up as a variant of subcluster A (see below). The subclusters with few members (C, H) are also missing, fused with the now-more-extended population G'. The main cluster populations, including Tr 37 (see below) and the surrounding cloud population are identified but the latter, which was previously divided into two
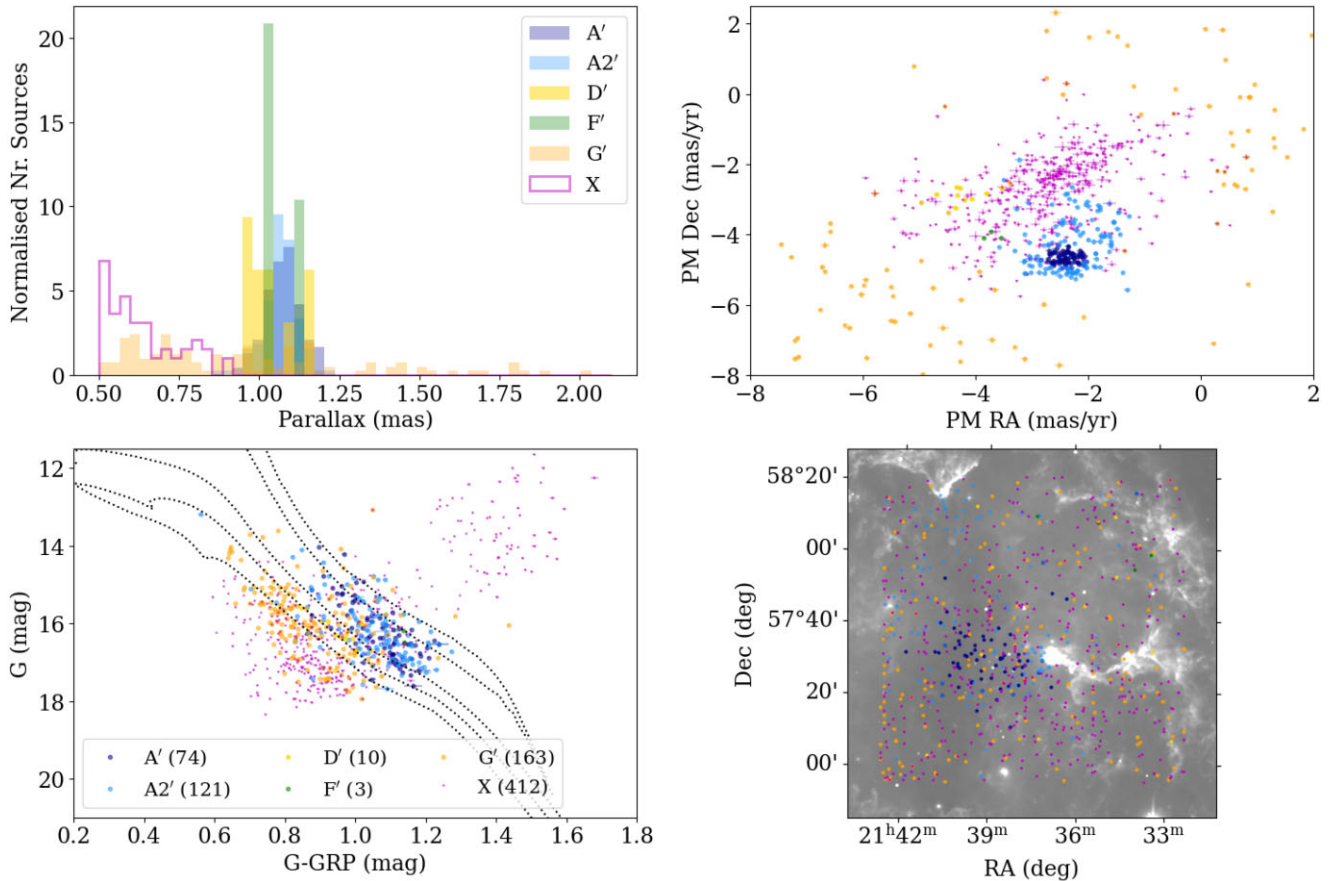
Figure 7. Clustering analysis for all the variable stars in the Tr 37 field (no cluster membership or YSO status required). As in Fig. 6, the top left panel shows the parallax distribution, top right panel shows the proper motion distribution (zoomed into the clustered parts), bottom left panel shows the colour–magnitude diagram with isochrones for 0.5, 1, 5, 10, and 20 Myr, and the bottom right panel shows the location of objects over the *WISE* W3 band image. We recover the main cluster (populations A' and A2', where the later includes some objects from previous populations B and F) and clearly identify background variables.

Table 4. Results of the clustering analysis for all the variable stars. The sublusters are named as variants of the subclusters from PB23, marked with a prime ('). The main cluster, A, splits in two when considering the variable sources only. Variants of subclusters D and F and of the extended population G are also recovered (D', F', G'), together with a pure background group (X). The background population is close to subcluster B in proper motion, but clearly distinct.

Subcluster	ϖ (mas)	pm_{RA} (mas yr^{-1})	pm_{Dec} (mas yr^{-1})	Nr. Sources	Comments
A'	1.066 ± 0.055	-2.392 ± 0.122	-4.645 ± 0.109	43	Similar to subcluster A
A2'	1.068 ± 0.050	-2.252 ± 0.539	-4.194 ± 0.725	99	Contains sources from subclusters A, B, and E.
D'	1.035 ± 0.070	-4.150 ± 0.332	-2.818 ± 0.130	8	Similar to subcluster D
F'	1.078 ± 0.050	-3.632 ± 0.089	-3.997 ± 0.090	2	Similar to subcluster F
G'	1.049 ± 0.781	-0.655 ± 6.831	-2.372 ± 6.215	145	Extended/non-clustered
X	0.256 ± 0.166	-2.738 ± 1.037	-2.652 ± 0.994	383	Background

main subclusters (E and F in PB23) is now merged together with the group A, which contains the majority of YSO.

The main subcluster A is identified, albeit divided in two groups, where A' is essentially identical to A and A2' is a variant that contains elements of population B and E as well. Interestingly, the A2' population is not co-located with the A' members, but rather associated with the IC 1396 cloud rims and presenting a larger proper motion spread. The inclusion of the northern globule (globule B in PB23) is logical, since this globule is related to the population B, which is now part of A2', and the same applies to the cloud rim to the east of Tr 37 (former population E). But a new result arises, that population A2' also contains the IC 1396 A globule. IC 1396 A was previously merged with the main Tr 37 cluster, which now appears

far more compact. Interestingly, radial velocity differences between IC 1396 A and Tr 37 have been known since long (Patel et al. 1995) and further confirmed more recently (Sicilia-Aguilar et al. 2019), but proper motions were not significantly different when considering the *Gaia* data for sources known to be physically related to the globule (Sicilia-Aguilar et al. 2019). By selecting variable stars only, we may be preferentially choosing the youngest targets, and this suggests a relation between the younger populations in the globules (Reach et al. 2004; Sicilia-Aguilar et al. 2006b; Getman et al. 2012) that was not revealed by *Gaia* data alone nor by individual stellar ages, due to noise. PB23 also identified a difference in age between clustered and non-clustered parts within some of the subclusters, in particular, the subcluster B, although the effect was not evident in others. The
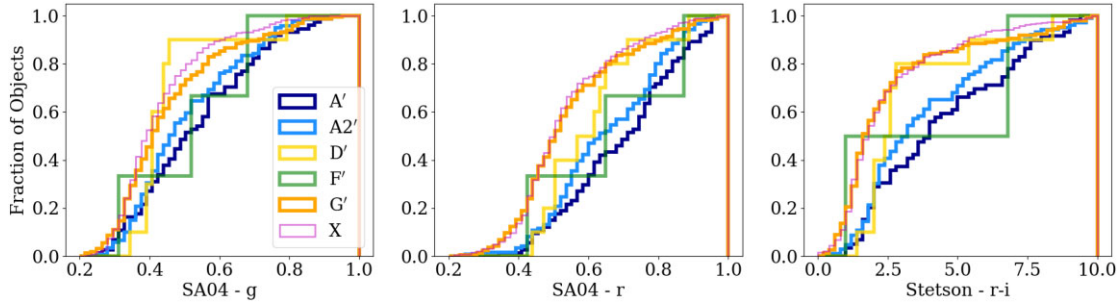
Figure 8. Cumulative distribution of the variability indices for objects classified as variable, according to their subclusters.

youth of the groups is also likely reflected in the high variability percentage found among those groups (Table 3). Since the current variability data are limited, this is a result that we will further explore within North-PHASE.

The extended population from PB23 now covers a larger parallax range and probably contains both high proper motion cluster members and field stars, most of which are in the background (subcluster G' in the figure). In addition, the clustering algorithm classifies the background variables in an independent, rather extended population at low parallax (X), clearly distinct from the YSO. The variables at intermediate parallaxes are also very spread in proper motions, as expected from a collection of field stars. From their positions in the colour–magnitude diagrams, they could also include some δ Scuti variables. A further group with only three objects is also identified, but it is not statistically significant and will not be further discussed.

As a final test, we investigated whether the variability properties are different between those 'variable only' subclusters. Fig. 8 shows distribution of the variability indices among the different groups. Objects consistent with cluster members (subclusters A', A2', D', F') are typically more variable and in a larger number of indices. The difference with the background population X is striking, and the extended population G' is also significantly different. A double-sided KS test reveals probabilities under 10^{-6} for all broad-band indices of group A' and X' to be drawn from the same population, and for the SA04 index for Hα the probability is 0.01. Comparing A' and G', the probability of both groups to share the same distributions of indices is always below 0.02, and much lower for the Stetson indices (which tend to track the most extreme variables). The difference between A' and G' is more striking than in the case of the members from A and G, probably because the parallax of G' is less well-constrained and thus includes more contaminants. But the results of this analysis confirm significant differences between the variability indices of young variables and variable field stars. These differences offer us a fundamental approach that we will explore in the future to classify objects for which astrometric data are not good or not available.

4.4 Variability and the presence of protoplanetary discs

Since protoplanetary discs are one of the key features of YSO related to many variability traits (e.g. Herbst et al. 1994), we looked at the variability of YSO with and without discs (inferred from their IR excesses) as well as at IR excesses among sources identified as variable by North-PHASE. The first part reveals the variability properties associated with the discs, tracking index performance regarding disc-related processes and what could be lost when new YSO are found by variability. The second part helps us to put in context all variables identified by North-PHASE.

4.4.1 Variability for sources with and without discs

The most detailed discs studies of Tr 37 were done using *Spitzer* (Sicilia-Aguilar et al. 2006b, 2013), so we used *Spitzer* data to explore discs versus variability for all the known members within the parallax limits imposed by PB23. Tr 37 was observed with *Spitzer* using both Infrared Array Camera (IRAC; Fazio et al. 2004), with wavelengths 3.6 to 8 µm, and the multiband imaging photometer (MIPS; Rieke et al. 2004) at 24 µm. The cluster is not entirely covered, but we find 284 sources that have both *Spitzer* observations and North-PHASE data, using a 2 arcsec radius to match (PB23).

Discs were identified by requesting IRAC [3.6]–[4.5]>0.2 mag, versus [5.8]–[8.0]>0.2. The first criterion identifies objects with a near-IR excess, but misses transitional and depleted discs. The second criterion is better suited at finding evolved discs, although it is also more susceptible to contamination. Using [3.6]–[4.5], we find a disc fraction of 25 per cent, while this raises to 48 per cent with [5.8]–[8.0]. These results are consistent with Sicilia-Aguilar et al. (2006b). We also find that 81 per cent of discs with [3.6]–[4.5] excess are variable (58/72), compared with 58 per cent (123/212) of those without discs. When [5.8]–[8.0] is considered, the fraction of variables changes to 70 per cent (disced, 95/135) versus 60 per cent (discless, 82/136). The first conclusion is that, even though we are still at the beginning of the survey, variability can detect about 2/3 of YSO both with and without discs. The second, that discs with excesses in the near-IR bands are more variable. This probably results from those discs being less evolved (Sicilia-Aguilar et al. 2006b) and having higher accretion rates (Sicilia-Aguilar et al. 2010), making them more prone to variability caused by hot spots, accretion, and circumstellar extinction events.

We also examined potential differences between the variability indices per among each group via histograms and double-sided KS tests. Fig. 9 shows some examples of the variability index distribution for objects with and without discs as defined above. Although disced sources tend to be more variable (especially, if they have near-IR excess, as mentioned before), there are discless ones with similarly high variability indices, and most YSO are variables. Table 5 shows that the differences in index distribution are statistically significant for all the SA04 indices (specially, for Hα) as well as for most of the Stetson indices. These differences also highlight the role of various types of indices to obtain a complete view of variability in YSO.

4.4.2 IR excess among variable stars

For the second part, we compared the 894 variable sources with the *Wide-field Infrared Survey Explorer* (*WISE*; Wright et al. 2010) as well as with *Spitzer*. A 2 arcsec radius finds 707 counterparts in *WISE*, while 309 stars were detected in at least two *Spitzer* bands. The
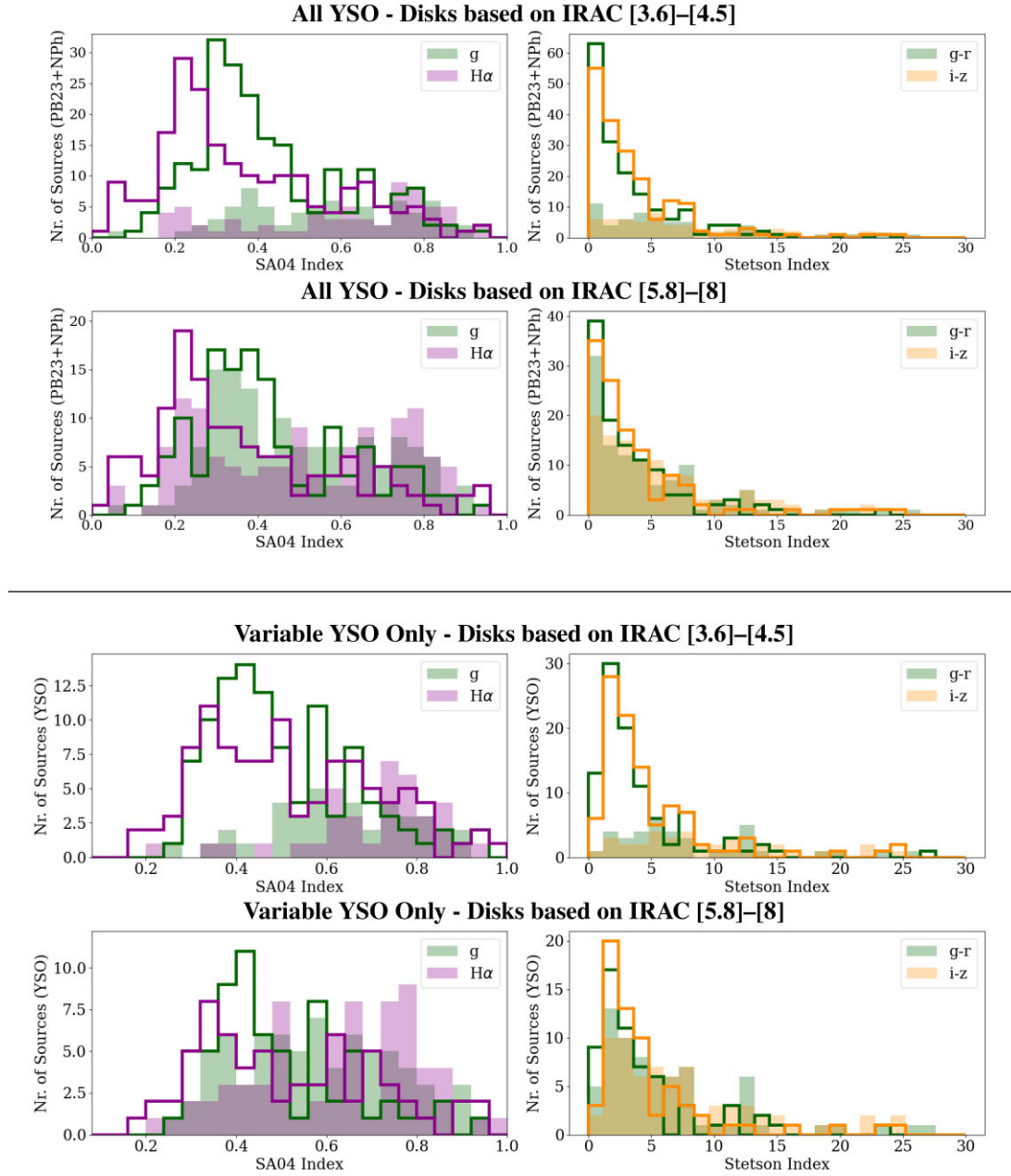
Figure 9. Variability index distribution for Tr 37 stars with (filled histogram) and without (step histogram) IR excess consistent with protoplanetary discs. The top two plots show the results for all the known YSO with *Spitzer* data, including those listed by PB23 and the new YSO candidates from North-PHASE. The bottom two panels correspond to those of the YSO that are found to be variable, including variable objects from PB23 and newly identified YSO candidates (based on the selection criteria from Section 4.2). Within each set, the *Spitzer* IRAC colours 3.6 μm–4.5 μm (top) and 5.8 μm–8 μm (bottom) are used to characterize the excess. Note that the Stetson index axis is truncated to value 30 so that the details at low indices are visible; there are very few objects beyond this limit and the KS tests (Table 5) include them all.

advantage of *WISE* is that it covers the entire region; the advantage of *Spitzer* is its better spatial resolution, making a clear difference in crowded areas and around nebulosity and allowing us to detect discs in more advanced stages of evolution. As shown by PB23, *WISE* W3 (12 μm) and W4 (22 μm) have too much contamination in Tr 37, so we can only identify discs using W1 (3.4 μm) and W2 (4.6 μm). This means that depleted and transition discs with holes, gaps, and/or small excesses may be missed, so disc fractions are a lower limit. We then examined the IR colours for the variable stars, subdividing them as in Section 4.2 and Fig. 5. We focused on the four groups of likely YSO, candidate giants, older sources, and 'HR contaminants',

and classify as excess sources those with colours >0.2 mag (PB23). Detailed figures are shown in Appendix A.

We find a substantial number of YSO variable sources with discs, although some discless sources have variability indices comparable to disced ones, strengthening the result from Section 4.4.1 that variability detects YSO with and without discs. For *WISE*, 63/291 sources have discs as seen in W1–W2 (disc fraction among variables 22 per cent). For *Spitzer*, 42/149 have [3.6]–[4.5] excesses (28 per cent), 76/149 have [5.8]–[8.0] excesses (51 per cent). Since MIPS did not reach the photospheric levels of Tr 37 objects later than A-type (Sicilia-Aguilar et al. 2006b), all sources found in this

Table 5. Results of the KS test comparing the variability indices for YSO with/without discs and accretion signatures. Discs are inferred using *Spitzer* data for IRAC, with colours [3.6]–[4.5] being sensitive to the innermost disc, while [5.8]–[8.0] can also detect transitional discs with holes and gaps and depleted discs (see Section 4.4). The columns for 'All' include all YSO sources from PB23 consistent with the cluster parallax, plus the new variable YSO from North-PHASE. The columns labelled 'Variable' include only those known YSO and YSO candidates that are variable. For accreting and non-accreting stars, we use the information from Sicilia-Aguilar et al. (2010) to determine the accretion status (see Section 4.5). Non-significant or marginally significant (*p*<0.05) differences are marked with ':'.

Index	KS p [3.6]–[4.5] (All)	KS p [5.8]–[8.0] (All)	KS p [3.6]–[4.5] (Variable)	KS p [5.8]–[8.0] (Variable)	KS p ($\dot{M}$)
SA04-g	6e-6	0.01	2e-5	8e-3	1e-4
SA04-r	4e-7	2e-3	4e-5	0.02	7e-3
SA04-i	4e-8	1e-4	1e-6	9e-4	1e-3
SA04-z	4e-9	2e-4	2e-7	9e-3	7e-4
SA04-Hα	2e-11	5e-5	1e-6	2e-2	5e-5
Stet. g-r	2e-7	0.1:	3e-5	0.05:	3e-4
Stet. g-i	3e-7	0.08:	5e-6	0.03	2e-3
Stet. g-z	2e-6	0.2:	3e-5	0.04	6e-4
Stet. r-i	9e-8	3e-4	6e-7	6e-3	1e-4
Stet. r-z	3e-8	3e-3	4e-6	9e-3	1e-5
Stet. i-z	5e-10	1e-3	5e-7	1e-3	3e-4

band have excesses. Although large variability indices are more frequent among disced sources, there are discless YSO with similarly large indices, even though the distribution of indices is clearly different among both groups. The disc fractions among variable YSO candidates are similar to the general disc fractions in the known YSO population, which is a further sign of variability efficiently picking both disced and discless YSO. Fig. 9 compares some of the variability indices for disced and discless sources, their differences highlighted by the KS tests in Table 5. We find that the distributions of all variability indices for objects with and without discs are significantly different, a further point that North-PHASE alone can add valuable information about the evolutionary status of YSO.

In contrast, there are essentially no 'older variables' nor 'HR contaminants' with excesses, but they are found for a few 'candidate giants', including 24 µm excess. Although IR excess among giant stars are common (e.g. Groenewegen 2012), we checked whether some could be very young stars or protostars, which often share the same region of the colour–magnitude diagram. Among 'candidate giants' with [3.6]–[4.5]>0.75 mag (the range of flat discs and protostars), we find several known TTS with massive discs, including 14–141, 11–2322, and 13–669 (the first two belong to the IC1396A globule population; Sicilia-Aguilar et al. 2006b), two sources with full discs observed with Herschel (21392541+5733202 and 21393104+5747140; Sicilia-Aguilar et al. 2015), a young X-ray source (21384282+5728547; Mercer et al. 2009), and two emission line stars (119 and 420 from Nakano et al. 2012). There is a further object with parallax 1.3 mas, which is likely a newly discovered YSO (at 21:37:43.87 +57:15:36.6) and three background variable sources with low parallax, which are likely variable giants. The number of YSO in this group is a sign that the isochrone cut is probably too restrictive, due to the compromise to limit contamination.

Among the eight 'candidate giants' with MIPS 24 µm excess, we recover source 420 from Nakano et al. (2012), but the rest are background sources. Three of them are listed as long-period variables by SIMBAD (Heinze et al. 2018; Gaia Collaboration 2023), one is a potential binary foreground star with parallax 2.2 mas, and the rest

are newly discovered background variables. This demonstrates that North-PHASE is also able to detect other variables, including long-period ones, a sign of the returns that it can provide to the community beyond star formation.

The conclusions are that our YSO selection criteria are very powerful to find real young stars, although some of them are lost to the 'giant candidates' and 'field stars' groups. The risk of missing sources to the 'field' is higher for dippers and objects with UXor variability (which are fainter and bluer in eclipse), but classification will improve as more variability data is collected, and considering the variability amplitude and light curves. YSO that fall over the isochrone into the 'giants' region tend to be very young and embedded and have massive discs and/or envelopes, so they can be further identified. The differences between the variability types within each class suggested in Section 4.2 hint that future study of this aspect will also aid the classification. The details are given in Table 6, which includes the summary of all the variable stars with IR counterparts as well as those that have a SIMBAD record about its class.

4.5 Detecting accretion and activity with North-PHASE

Accretion and chromospheric activity are key causes of YSO variability (e.g. Fischer et al. 2023). At the point of writing this introductory paper, stellar masses and luminosities, required to measure accretion, are not available yet. The number of uJAVA observations, which is key to determine absolute accretion rates (Gullbring et al. 1998), is still very small. Nevertheless, we can already explore whether the existing Hα and variability data correlate with the known accretion rates to test future capabilities.

For Tr 37, accretion rates or their limits have been measured from *U*-band photometry and spectroscopy for about a hundred objects (Sicilia-Aguilar et al. 2010), and via Hα photometry for about 150 sources (Barentsen et al. 2011), some of which are in common. Further Hα measurements from spectroscopy are available for an even larger sample, as it has played an important role in identifying a substantial number of members (Sicilia-Aguilar et al. 2004, 2005, 2006b; Nakano et al. 2012; Sicilia-Aguilar et al. 2013). There are many different proxies to derive accretion rates (e.g. Gullbring et al. 1998; Fang et al. 2009; Alcalá et al. 2014), not entirely deprived of systematic differences. Before calibrating the North-PHASE data, it is important to explore the agreement within existing measurements, which is not always evident (see Appendix C) and will need to be investigated in the future.

We first compared the variability indices for stars with and without known accretion. For this, we used the data from Sicilia-Aguilar et al. (2010), finding 82 cross-matches with the North-PHASE lists. The accretion data includes detections, upper limits, and objects for which accretion was ruled out using high-resolution Hα spectroscopy. Although correlations between indices and accretion rates are at most marginal (only indices involving gSDSS band seem to show some correlation, with Spearman *r* around 0.4 and probability *p* around 0.01), we observe that the indices for sources with and without accretion show clear differences (see Fig. 10). Table 5, containing the results of a double-sided KS test between the indices of non-accreting (35) and accreting (47) sources, demonstrates its significance. The differences show that the magnitude of the variability index can reveal information about the accretion status of the sources. Although the trend was already seen when checking the indices of disced sources (Fig. 9), we now clearly see that the second peak of the SA04 indices is strongly biased towards accreting sources, which explains its correlations with YSO previously mentioned.

Table 6. Summary of the variability types found in Tr 37. For details on the classification scheme, see Section 4.2 and Fig. 5. The excess sources are given relative to the total detected in each band. The classification corresponds to the SIMBAD naming, where we have included all types of YSO together (e.g. TTS, Emission Line, HAeBe, Orion type variable), plus eclipsing binaries (EB) and long-period variables (LPV). At this point, we do not make a distinction between candidates and confirmed types from the SIMBAD database. As 'Other', we include any other object listed as variable but under a different or unspecified category. Under 'Stars' we count any object classified as star by SIMBAD but for which no indication of variability has been reported. [a] Including pulsating and SB stars. [b] Including one pulsating star. [c] Including two RR Lyr and one pulsating star.

Tentative Class	Total Nr.	IR excess (W1-W2)	IR excess [3.6]−[4.5]	IR excess [5.8]−[8]	IR excess [3.6]−[24]	SIMBAD	YSO	EB	LPV	Other	Stars
YSO	304	59/268	40/146	74/146	37/37	220	176	6	1	2	35
Giants	98	10/95	12/20	12/18	9/11	54	12	–	38	3[a]	1
Field	183	1/99	8/66	17/25	–	13	1	6	–	4[b]	2
HRC	244	9/196	9/67	26/57	–	36	17	5	5	7[c]	2
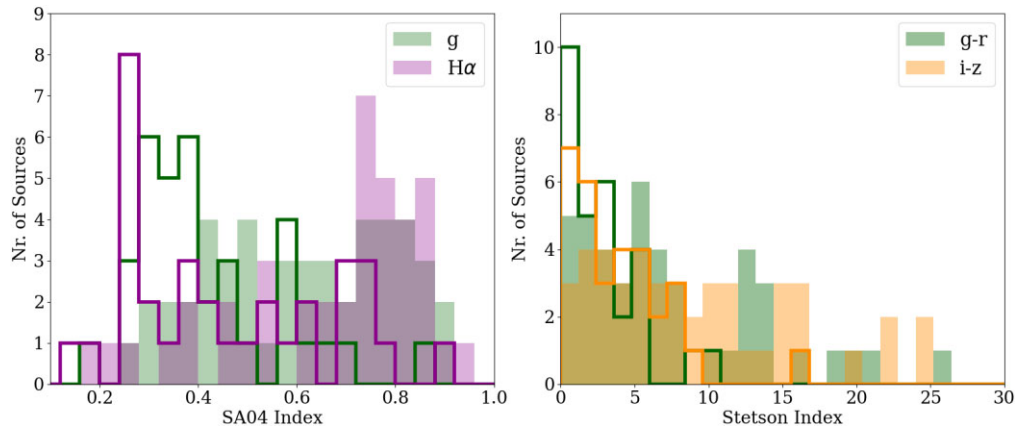
Figure 10. A comparison of the variability index distribution for stars with accretion (filled histogram) versus no accretion (step histogram), using the information on the accretion status from Sicilia-Aguilar et al. (2010). The left panel shows some examples of SA04 indices, while the right one shows examples with the Stetson indices. Note that the Stetson index x axis is truncated to value 30 so that the details at low indices are visible; there are very few objects beyond this limit and the KS tests (Table 5) include them all.

The results are in line with the previous discussions of the SA04 and Stetson indices. As expected, the Hα band shows the strongest differences among the SA04 indices. Although switching from an accreting to a non-accreting status is very rare (Sicilia-Aguilar et al. 2010; Sicilia-Aguilar et al. 2020), accretion rates for individual stars are typically variable by factors of few to tens, which may dilute the correlations between indices and accretion rate. It is thus of paramount importance to explore the results of the u-JAVA filter in the future, to be able to quantify accretion at the same time than variability.

We also used the J0660 filter to detect accreting and active stars via a colour–colour diagram, following Barentsen et al. (2011). We considered the average rSDSS, iSDSS, and J0660 magnitudes and standard deviations, exploring both field stars and YSO with estimated Hα equivalent width (EW; from Sicilia-Aguilar et al. 2005, 2006a; Barentsen et al. 2011; Nakano et al. 2012; Sicilia-Aguilar et al. 2013). For those, we classified them according to their published Hα EW as classical T Tauri stars (CTTS) or weak-lined T Tauri stars (WTTS) using a simple threshold of 10 Å (noting that this threshold may overestimate the CTTS in the lower mass range, while underestimating them towards the higher mass range). Only objects with at least five observations in all the involved filters were considered. The results are shown in Fig. 11. The young CTTS and WTTS stand out from the field stars and from each other, which validates the use of the North-PHASE photometry to distinguish YSO as well as accreting and non-accreting (in case of WTTS and YSO, chromospherically active) stars.
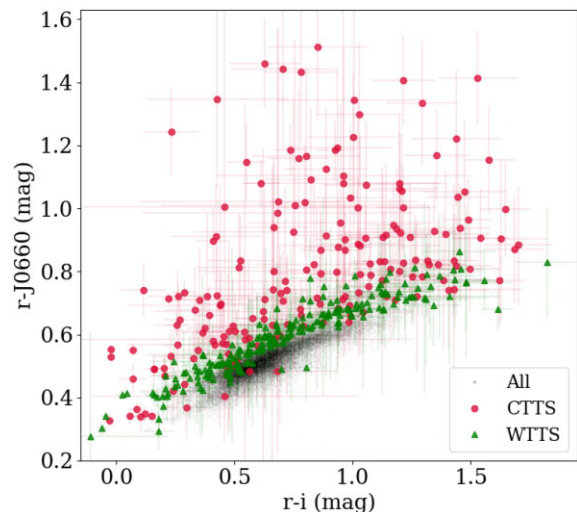
Figure 11. Colour–colour diagram involving average rSDSS, iSDSS, and J0660 colours, to explore Hα excesses. Small grey symbols show the field objects, red dots show known CTTS, and green triangles display known WTTS, classified based on existing spectroscopic and photometric Hα EW. The 'error bars' are not the photometric errors (which are much smaller), but the standard deviation of the magnitudes. The diagram shows how the r-J0660 colour distinguishes the CTTS from the WTTS and from the general population. Note that CTTS and WTTS have been corrected for the general cluster extinction, but the rest has not been modified.

Although a proper estimate of the Hα EW requires model photospheres (e.g. Barentsen et al. 2011), which is beyond our initial capabilities, we demonstrate that the process is feasible using the North-PHASE data. As before, we note that we are comparing non-simultaneous measurements, which will increase the spread. A rough estimate of the flux in the Hα line can be obtained from the J0660 flux, taking a scaled version of the rSDSS flux as the continuum. For this first attempt, we used the data acquired on the reference night. We used the filter zero point calibrations and FWHM from the Spanish Virtual Observatory[13] for the conversion, including the average cluster extinction from PB23 ($A_V = 1.4$ mag).

The resulting Hα EW was compared to EW measurements from spectra (Sicilia-Aguilar et al. 2005, 2006b; Nakano et al. 2012; Sicilia-Aguilar et al. 2013) and IPHAS photometry (Barentsen et al. 2011). To check if different calibrations could be an issue here, we made two tests: using the Pan-STARRS calibration for our rSDSS data that we are adopting for the survey (see Section 3.3), or using IPHAS to calibrate rSDSS (note that J0660 is always calibrated using IPHAS), but the results, despite systematics, are comparable. The procedure of estimating 'rough' EW is particularly inaccurate for late-type stars where the red part of the spectrum is very far from flat, which triggers a general offset in the measured EW, but we find a clear correlation between available EW measurements (Spearman $r \sim 0.3$ for spectroscopic results, or 0.5 comparing to photometric values, with essentially negligible probability that the results are not correlated). This shows that the North-PHASE J0660 photometric measurements can be used to estimate the strength of the Hα line, as well as to distinguish stars with and without emission, which adds to our criteria for the identification of YSO, even if the systematic differences will require an improved calibration (see Appendix C).

4.6 Detecting rotation and classifying light curves

At present, the North-PHASE data on Tr 37 is not complete enough to independently derive rotational periods. Nevertheless, we did some brief tests using sources for which information on periods and light curves was available from the *Transiting Exoplanet Survey Satellite* (*TESS*), to assess the feasibility of using North-PHASE data to detect rotational modulations and distinguish the causes of variability in objects in a wide range of magnitudes.

We focused on a few well-known objects classified as different types of variables using *TESS* data (see Kahar et al., in preparation; Schlueter et al., in preparation) from the sectors available for Tr 37 (16, 17, 56, and 57). The data were extracted using PYTHON TESSCUT and a 3 pixel window. For some stars, the presence of similar-magnitude, nearby ($\sim$20 arcsec) objects casts some doubts regarding the possible contamination, but we marked them accordingly and use them as a test of what North-PHASE can detect. The risk is higher for fainter targets.

The periods were derived from the *TESS* data using a Lomb-Scargle periodogram (VanderPlas 2018) followed by phase-folding to examine the resulting light curves. Note that, since most of the stars have multiple variability causes, ranging from rotationally modulated (hot and cold spots) to quasi-periodic (dips) and irregular (outburst), photometric periods of young stars are often not detectable in all data sets or during all epochs (Sicilia-Aguilar et al. 2023). The information from the *TESS* analysis is then used for a consistency test of the North-PHASE data. Phase-folding of the North-PHASE data allows us to check if it agrees with the known *TESS* period and infer other

potential periods (particularly helped by the densely sampled interval around JD = 2460039-47, although it will need to be confirmed with further data), while examining the light curves allows us to compare the typical features detected by *TESS* (e.g. dips, bursts, sinusoidal modulations, and irregular activity).

The selected sources are listed in Table 7, and a few examples are shown in Fig. 12. The naming convention follows Sicilia-Aguilar et al. (2005). Their magnitude range spans gSDSS$\sim$13–19 mag, which ranges from non-saturated data to the current noise limit. We find that all sources are labelled as variables in most filters, considering the SA04 indices although, as expected, those faint and red are often noisy in g, preventing the detection of variability at lower rates. For 13–277, also known as GM Cep and known to have variable accretion and dipper events (Sicilia-Aguilar et al. 2008; Semkov & Peneva 2012), we do not observe rotational modulation, but find a dip. Object 11–2131 is inconclusive, and we also fail to detect the 1 d period of 72–875 due to sampling or to the effect of nearby objects in the *TESS* data sets, but the remaining objects all display different levels of modulation. As seen in Fig. 12, the levels of the rotational modulation observed have peak-to-peak changes typically below 0.2 mag. As expected, the North-PHASE data sampling in intervals of 1–3 d allows us to verify the typical rotational periods in the range of 1–10 d (Bouvier et al. 1995). The conclusion from this limited experiment is that North-PHASE can detect stellar rotation for a large number of young stars, and distinguish between the different variability types in objects that may not be dominated by rotational modulation, or may present quasi-periodic variations.

5 CONCLUSIONS AND OUTLOOK

Using the first four months of data from the Javalambre North-PHASE Legacy Survey, we have explored and defined the criteria that should be applied in the identification of new young stars as well as other variable sources and performed a comprehensive study of one of the clusters, Tr 37. This initial part of the work allows us to demonstrate the power of the survey regarding the identification of YSO and other types of variables. The initial data set, albeit limited, allows us to recover known variable stars in Tr 37, as well as to identify 50 new YSO candidates. We also show that the variability indices used give us a very good handle to detect different types of variability, with the Stetson indices being more sensitive to strong variables (e.g. those suffering accretion bursts or dips) and the SA04 indices allowing us to detect milder but steadier variability, such as changes induced by cold or hot spots, as well as those related to accretion.

Using the well-known Tr 37 cluster as a test-bench, we find that variability is a very good tool to identify new cluster candidate members, being moreover complementary to other techniques, including classical searches for discs and accretion signatures, and astrometry. Therefore, getting an integral and global picture of star formation within clusters can only be done using larger fields and a combination of techniques to avoid biases in the membership collection.

Repeating the clustering analysis from PB23 including the variable stars strengthens the existing results, adding a significant number of younger, more variable stars in the surrounding clouds. A clustering analysis including only variable stars also reveals substructure that had been missing before, such as the splitting the originally monolithic main Tr 37 cluster into the central versus the globules population due to the higher degree of variability among objects in the surrounding globules. These differences, which had been washed out when using astrometry only, are a sign that variability can help us to disentangle the star formation history to a higher degree than with *Gaia* data alone. Being able to efficiently distinguish between types of variable stars using their astrometric properties

---

[13]http://svo2.cab.inta-csic.es/svo/theory/fps3/, see filters for the Observatorio Astronómico de Javalambre (OAJ).

Table 7. Comparing the variability detected by *TESS* and North-PHASE. We provide the calibrated, average North-PHASE gSDSS magnitude and standard deviation (a measurement of the spread). We show the best estimated period from *TESS* (from periodograms) and North-PHASE (as visually inferred from phase-folding of the limited data), if available. Note that, at present, the North-PHASE proposed periods are based on consistency with *TESS* and visual inspection of the phase-folded curves, being merely indicative. Sources with *TESS* data that may be compromised to some extent due to others nearby are marked with *. All sources are found to be variable in most filters, with Hα being compromised when close to nebulosity (e.g. for 14–125 and 73–71) and gSDSS being noisy for faint targets (e.g. 92–393 is marginal or noisy in indices involving g, but clearly variable in riz and Hα). The variability type is labelled as rotation ('Rot.'), irregular ('Irr.'), dipper ('Dip.'), burster ('Bur.'), and uncertain ('Unc.').

Name	gSDSS (mag)	P_{TESS} (d)	P_{NP} (d)	Variability (North-PHASE)	Type (*TESS*)	Type (North-PHASE)
13–277	13.58 ± 0.22	3.59	–	grizHα/All St	Rot./Dip.	Irr./Dip.
12–1091	16.89 ± 0.10	7.5:	3.6:	grizHα/All St	Rot.:/Dip.	Rot./Dip.
11–2037	16.91 ± 0.08	1.43	1.43	grizHα/All St	Rot.	Rot./Unc.
14–125	17.31 ± 0.14	9.9:*	9.5:	griz/All St	Rot./Irr.:	Rot.
73–71	17.90 ± 0.11	3.49	3.49	griz/All St	Rot.	Rot.
14–183	17.65 ± 0.16	6.44:*	6.4	grizHα/All St	Rot./Dip./Irr.	Rot./Dip.:
73–758	18.01 ± 0.09	1.10*	1.1	rizHα/All St	Rot./Bur.	Rot.
13–1048	18.01 ± 0.27	1.8:*	9.9:	grizHα/All St	Irr.	Rot./Dip.
11–2131	18.36 ± 0.25	–	5.58	grizHα/All St	–	Unc./Rot.:
21–998	18.93 ± 0.46	9.69*	9.7	grizHα/All St	Rot.:/Dip./Irr.	Rot.
72–875	19.14 ± 0.21	1.04:*	4.95:	grizHα/All St	Rot.	Rot.:
92–393	19.17 ± 0.17	1.40*	3.3:	rizHα/Red St	Rot.:	Rot.:
13–1238	19.23 ± 0.21	7.99:	7.1:	grizHα/All St	Unc.	Rot.
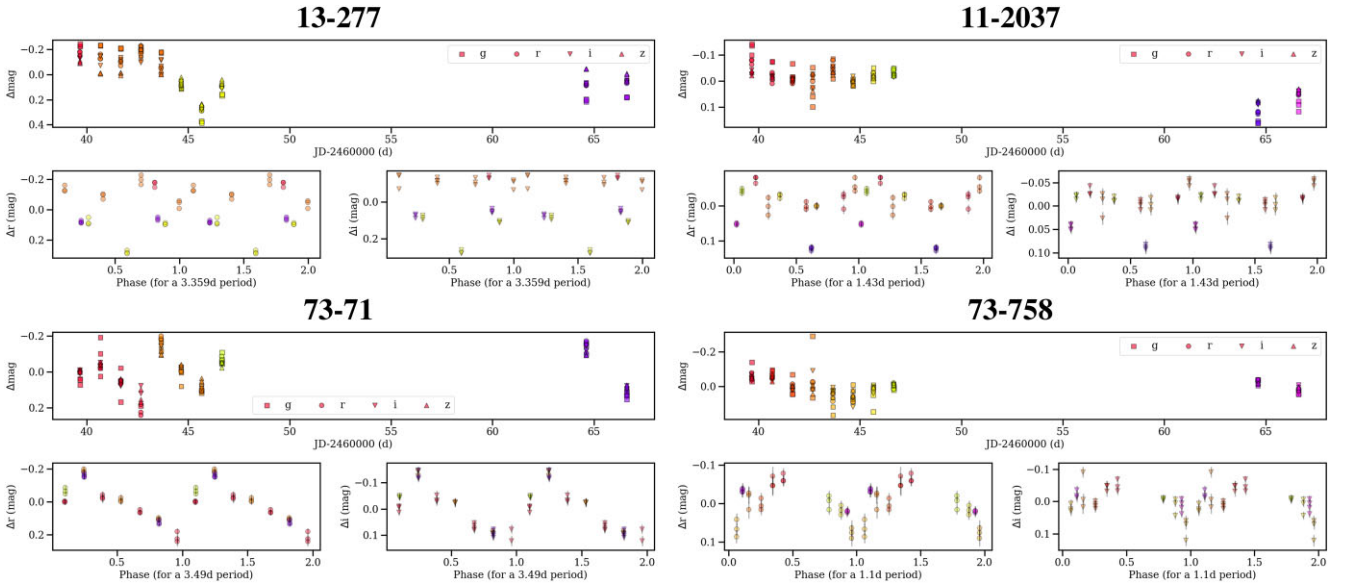
Figure 12. Some examples of rotational modulation observed by North-PHASE in Tr 37. For each target, the top panel shows the complete light curve relative to the average magnitude in each filter (note that dates are colour-coded, while filters are marked by different symbols). The two bottom panels show the phase-folded light curve for the best period for the rSDSS and iSDSS filters.

is also a sign that we are finding a diverse non-young, variable population.

Although the stellar properties are still beyond the reach of the survey, we are able to distinguish accreting objects as well as YSO with chromospheric activity. We also find significant differences between the variability indices for disced and discless objects, particularly striking in the Hα index. We observe that SA04 indices are more sensitive to small-scale variations, and there is also a clear difference between the performance of different bands, probably related to the diversity of variability causes (accretion, extinction, spots). Even if the strongest variability tends to be related to objects with massive protoplanetary discs as evidenced from their IR excesses, our combination of indices detects a similar fraction of variable YSO without discs in Tr 37. This evidences that our

methodology can effectively detect variations from spots in young objects, which is further supported by the confirmation of rotational modulations and other types of variability for stars with magnitudes gSDSS∼13–19 mag.

North-PHASE finds other types of variables as well, including pulsating stars, eclipsing binaries, and long-period variables. We also see that most of the objects that have been suggested to be cluster members but were revealed as contaminants tend to be background variable giants, which interestingly also show different variability trends. Our analysis shows that the variability indices for field stars and objects that are the main source of contamination (e.g. those that appear in the same region of the HR diagram but are not cluster members) are significantly different from those of YSO. Investigating such differences in variability will help in the future

to further distinguish YSO even when in the absence of further data or if astrometry is unreliable, as well as to assign YSO probabilities to new candidate sources based on variability alone.

From the results on Tr 37, we have now created a set of routines to identify young stars and examine their properties and can confirm that, so far, the North-PHASE survey exceeds its initial expectations, and will provide valuable data beyond the field of star formation.

DATA AVAILABILITY

The North-PHASE project is a Legacy Survey at the Javalambre Observatory. As such, the data are released after publication in the Javalambre Observatory site.[15] On the long term, the project will contribute to the Legacy data base at Javalambre and the Spanish Virtual Observatory, and all analysis tools, including customized PYTHON packages, will be released to the public via GitHub by the end of the project. In the meantime, the project team accepts reasonable requests of data and code and is open to collaboration.

REFERENCES

Alcalá J. M. et al., 2014, A&A, 561, A2
Astropy Collaboration, 2013, A&A, 558, A33
Astropy Collaboration, 2018, AJ, 156, 123
Bailer-Jones C. A. L., 2022, ApJ, 935, L9
Barentsen G. et al., 2011, MNRAS, 415, 103
Barentsen G. et al., 2014, MNRAS, 444, 3230
Bassa C. G., Hainaut O. R., Galadí-Enríquez D., 2022, A&A, 657, A75
Bellm E. C. et al., 2019, PASP, 131, 018002
Bertin E., 2006, in Gabriel C., Arviset C., Ponz D., Enrique S., eds, ASP Conf. Ser. Vol. 351, Astronomical Data Analysis Software and Systems XV. Astron. Soc. Pac., San Francisco, p. 112
Bertin E., 2011, in Evans I. N., Accomazzi A., Mink D. J., Rots A. H., eds, ASP Conf. Ser. Vol. 442, Astronomical Data Analysis Software and Systems XX. Astron. Soc. Pac., San Francisco, p. 435
Bertin E., Arnouts S., 1996, A&AS, 117, 393
Bertin E., Mellier Y., Radovich M., Missonnier G., Didelon P., Morin B., 2002, in Bohlender D. A., Durand D., Handley T. H., eds, ASP Conf. Ser. Vol. 281, Astronomical Data Analysis Software and Systems XI. Astron. Soc. Pac., San Francisco, p. 228
Bertini L., Roccatagliata V., Kim M., 2023, A&A, 671, L2
Bonoli S. et al., 2021, A&A, 653, A31
Bouvier J., Covino E., Kovo O., Martin E. L., Matthews J. M., Terranegra L., Beck S. C., 1995, A&A, 299, 89
Bouvier J., Alencar S. H. P., Harries T. J., Johns-Krull C. M., Romanova M. M., 2007, in Reipurth B., Jewitt D., Keil K., eds, Protostars and Planets V. Univ. Arizona Press, Tucson, p. 479

---

[14] http://www.astropy.org
[15] https://archive.cefca.es/catalogues/north_phase-paper1

Bozhinova I. et al., 2016, MNRAS, 463, 4459

Bressan A., Marigo P., Girardi L., Salasnich B., Dal Cero C., Rubele S., Nanni A., 2012, MNRAS, 427, 127

Briceño C. et al., 2001, Science, 291, 93

Cenarro A. J. et al., 2019, A&A, 622, A176

Chambers K. C. et al., 2016, preprint (arXiv:1612.05560)

Cody A. M., Hillenbrand L. A., 2014, ApJ, 796, 129

Contreras M. E., Sicilia-Aguilar A., Muzerolle J., Calvet N., Berlind P., Hartmann L., 2002, AJ, 124, 1585

Cortés-Contreras M. et al., 2019, MNRAS, 490, 3046

de Juan Ovelar M., Kruijssen J. M. D., Bressert E., Testi L., Bastian N., Cánovas H., 2012, A&A, 546, L1

de la Fuente Marcos R., de la Fuente Marcos C., 2018, Res. Notes Am. Astron. Soc., 2, 30

Donati J. F. et al., 2019, MNRAS, 483, L1

Eiroa C. et al., 2002, A&A, 384, 1038

Eisner J. A. et al., 2018, ApJ, 860, 77

Evitts J. J. et al., 2020, MNRAS, 493, 184

Fang M., van Boekel R., Wang W., Carmona A., Sicilia-Aguilar A., Henning T., 2009, A&A, 504, 461

Fang M., Kim J. S., van Boekel R., Sicilia-Aguilar A., Henning T., Flaherty K., 2013, ApJS, 207, 5

Fazio G. G. et al., 2004, ApJS, 154, 10

Fischer W. J., Hillenbrand L. A., Herczeg G. J., Johnstone D., Kospal A., Dunham M. M., 2023, in Inutsuka S., Aikawa Y., Muto T., Tomida K., Tamura M., eds, ASP Conf. Ser. Vol. 534, Protostars and Planets VII. Astron. Soc. Pac., San Francisco, p. 355

Flewelling H. A. et al., 2020, ApJS, 251, 7

Froebrich D., Scholz A., Eislöffel J., Murphy G. C., 2005, A&A, 432, 575

Froebrich D. et al., 2018, MNRAS, 478, 5091

Froebrich D. et al., 2021, MNRAS, 506, 5989

Froebrich D. et al., 2023, MNRAS, 520, 5413

Froebrich D. et al., 2024, MNRAS, 529, 1283

Gaia Collaboration, 2016, A&A, 595, A1

Gaia Collaboration, 2023, A&A, 674, A1

Getman K. V., Feigelson E. D., Garmire G., Broos P., Wang J., 2007, ApJ, 654, 316

Getman K. V., Feigelson E. D., Sicilia-Aguilar A., Broos P. S., Kuhn M. A., Garmire G. P., 2012, MNRAS, 426, 2917

Groenewegen M. A. T., 2012, A&A, 540, A32

Gullbring E., Hartmann L., Briceño C., Calvet N., 1998, ApJ, 492, 323

Hainaut O. R., Williams A. P., 2020, A&A, 636, A121

Hamilton C. M. et al., 2005, AJ, 130, 1896

Hartmann L., Herczeg G., Calvet N., 2016, ARA&A, 54, 135

Heinze A. N. et al., 2018, AJ, 156, 241

Herbert C., Froebrich D., Scholz A., 2023, MNRAS, 520, 5433

Herbst W., Herbst D. K., Grossman E. J., Weinstein D., 1994, AJ, 108, 1906

Hsu W.-H., Hartmann L., Allen L., Hernández J., Megeath S. T., Tobin J. J., Ingleby L., 2013, ApJ, 764, 114

Jerabkova T., Boffin H. M. J., Beccari G., de Marchi G., de Bruijne J. H. J., Prusti T., 2021, A&A, 647, A137

Johnstone C. P., Jardine M., Gregory S. G., Donati J. F., Hussain G., 2014, MNRAS, 437, 3202

Joy A. H., 1945, ApJ, 102, 168

Kenyon S. J., Bromley B. C., 2004, Nature, 432, 598

Koenigl A., 1991, ApJ, 370, L39

Kruijssen J. M. D., Longmore S. N., Chevance M., 2020, ApJ, 905, L18

Lada C. J., Lada E. A., 2003, ARA&A, 41, 57

Lightfoot J. F., 1989, MNRAS, 239, 665

Lin D. N. C., Bodenheimer P., Richardson D. C., 1996, Nature, 380, 606

Lindegren L., Madsen S., Dravins D., 2000, A&A, 356, 1119

Lindegren L. et al., 2021, A&A, 649, A2

Loren R. B., Peters W. L., Vanden Bout P. A., 1975, ApJ, 195, 75

Magnier E. A. et al., 2020, ApJS, 251, 6

Mamajek E. E., Barenfeld S. A., Ivanov V. D., Kniazev A. Y., Väisänen P., Beletsky Y., Boffin H. M. J., 2015, ApJ, 800, L17

Marigo P., Bressan A., Nanni A., Girardi L., Pumo M. L., 2013, MNRAS, 434, 488

Matsumura S., Brasser R., Ida S., 2021, A&A, 650, A116

Meingast S., Alves J., 2019, A&A, 621, L3

Meingast S., Alves J., Rottensteiner A., 2021, A&A, 645, A84

Mendigutía I., Oudmaijer R. D., Rigliaco E., Fairlamb J. R., Calvet N., Muzerolle J., Cunningham N., Lumsden S. L., 2015, MNRAS, 452, 2837

Mendigutía I., Solano E., Vioque M., Balaguer-Nuñez L., Ribas A., Huélamo N., Rodrigo C., 2022, A&A, 664, A66

Meng H. Y. A., Rieke G. H., Kim J. S., Sicilia-Aguilar A., Cross N. J. G., Esplin T., Rebull L. M., Hodapp K. W., 2019, ApJ, 878, 7

Mercer E. P., Miller J. M., Calvet N., Hartmann L., Hernandez J., Sicilia-Aguilar A., Gutermuth R., 2009, AJ, 138, 7

Morales-Calderón M. et al., 2009, ApJ, 702, 1507

Nakano M., Sugitani K., Watanabe M., Fukuda N., Ishihara D., Ueno M., 2012, AJ, 143, 61

Patel N. A., Goldsmith P. F., Snell R. L., Hezel T., Xie T., 1995, ApJ, 447, 721

Pelayo-Baldárrago M. E., Sicilia-Aguilar A., Fang M., Roccatagliata V., Kim J. S., García-Álvarez D., 2023, A&A, 669, A22

Platais I., Kozhurina-Platais V., van Leeuwen F., 1998, AJ, 116, 2423

Racero E. et al., 2022, A&A, 659, A38

Reach W. T. et al., 2004, ApJS, 154, 385

Rebull L. M. et al., 2013, AJ, 145, 15

Rebull L. M., Stauffer J. R., Cody A. M., Hillenbrand L. A., Bouvier J., Roggero N., David T. J., 2020, AJ, 159, 273

Richert A. J. W., Getman K. V., Feigelson E. D., Kuhn M. A., Broos P. S., Povich M. S., Bate M. R., Garmire G. P., 2018, MNRAS, 477, 5191

Rieke G. H. et al., 2004, ApJS, 154, 25

Roccatagliata V., Ratzka T., Henning T., Wolf S., Leinert C., Bouwman J., 2011, A&A, 534, A33

Roccatagliata V., Sacco G. G., Franciosini E., Rand ich S., 2018, A&A, 617, L4

Roccatagliata V., Franciosini E., Sacco G. G., Randich S., Sicilia-Aguilar A., 2020, A&A, 638, A85

Röser S., Schilbach E., 2019, A&A, 627, A4

Scholz A., Natta A., Bozhinova I., Petkova M., Relles H., Eislöffel J., 2019, MNRAS, 484, 4260

Semkov E. H., Peneva S. P., 2012, Ap&SS, 338, 95

Sicilia-Aguilar A., Hartmann L. W., Briceño C., Muzerolle J., Calvet N., 2004, AJ, 128, 805

Sicilia-Aguilar A., Hartmann L. W., Hernández J., Briceño C., Calvet N., 2005, AJ, 130, 188

Sicilia-Aguilar A., Hartmann L. W., Fürész G., Henning T., Dullemond C., Brandner W., 2006a, AJ, 132, 2135

Sicilia-Aguilar A. et al., 2006b, ApJ, 638, 897

Sicilia-Aguilar A. et al., 2008, ApJ, 673, 382

Sicilia-Aguilar A., Henning T., Hartmann L. W., 2010, ApJ, 710, 597

Sicilia-Aguilar A., Kim J. S., Sobolev A., Getman K., Henning T., Fang M., 2013, A&A, 559, A3

Sicilia-Aguilar A., Fang M., Roccatagliata V., Collier Cameron A., Kóspál Á., Henning T., Ábrahám P., Sipos N., 2015, A&A, 580, A82

Sicilia-Aguilar A. et al., 2016, Publ. Astron. Soc. Austr., 33, e059

Sicilia-Aguilar A., Patel N., Fang M., Roccatagliata V., Getman K., Goldsmith P., 2019, A&A, 622, A118

Sicilia-Aguilar A., Manara C. F., de Boer J., Benisty M., Pinilla P., Bouvier J., 2020, A&A, 633, A37

Sicilia-Aguilar A. et al., 2023, MNRAS, 526, 4885

Smilgys R., Bonnell I. A., 2017, MNRAS, 472, 4982

Spezzi L., De Marchi G., Panagia N., Sicilia-Aguilar A., Ercolano B., 2012, MNRAS, 421, 78

Stetson P. B., 1996, PASP, 108, 851

Szilágyi M., Kun M., Ábrahám P., Marton G., 2023, MNRAS, 520, 1390

Tarricq Y., Soubiran C., Casamiquela L., Castro-Ginard A., Olivares J., Miret-Roig N., Galli P. A. B., 2022, A&A, 659, A59

Taylor M. B., 2005, in Shopbell P., Britton M., Ebert R., eds, ASP Conf. Ser. Vol. 347, Astronomical Data Analysis Software and Systems XIV. Astron. Soc. Pac., San Francisco, p. 29

Teixeira P. S., Scholz A., Alves J., 2020, A&A, 642, A86

van Dokkum P. G., 2001, PASP, 113, 1420

VanderPlas J. T., 2018, ApJS, 236, 16

Wenger M. et al., 2000, A&AS, 143, 9

Wright E. L. et al., 2010, AJ, 140, 1868

Wright N. J., Goodwin S., Jeffries R. D., Kounkel M., Zari E., 2022, in Inutsuka S.-i., Aikawa Y., Muto T., Tomida K., Tamura M., eds, ASP Conf. Ser. Vol. 534, Protostars and Planets VII. Astron. Soc. Pac., San Francisco, p. 129

Zhou X., Herczeg G. J., Fang M., Wang X., Ren B. B., 2023, ApJ, 955, 11

APPENDIX A: SUPPLEMENTARY FIGURES

In this section, we include some of the figures that complete or clarify parts of the work. Fig. A1 shows the comparison between different types of variability indices, both for members and non-members, as discussed in Section 4.1. Fig. A2 presents the variability indices for objects in the variable groups as defined in Section 4.2. Figs A3 and A4 show the infrared colours versus the variability indices for objects in the classes discussed in Section 4.4.
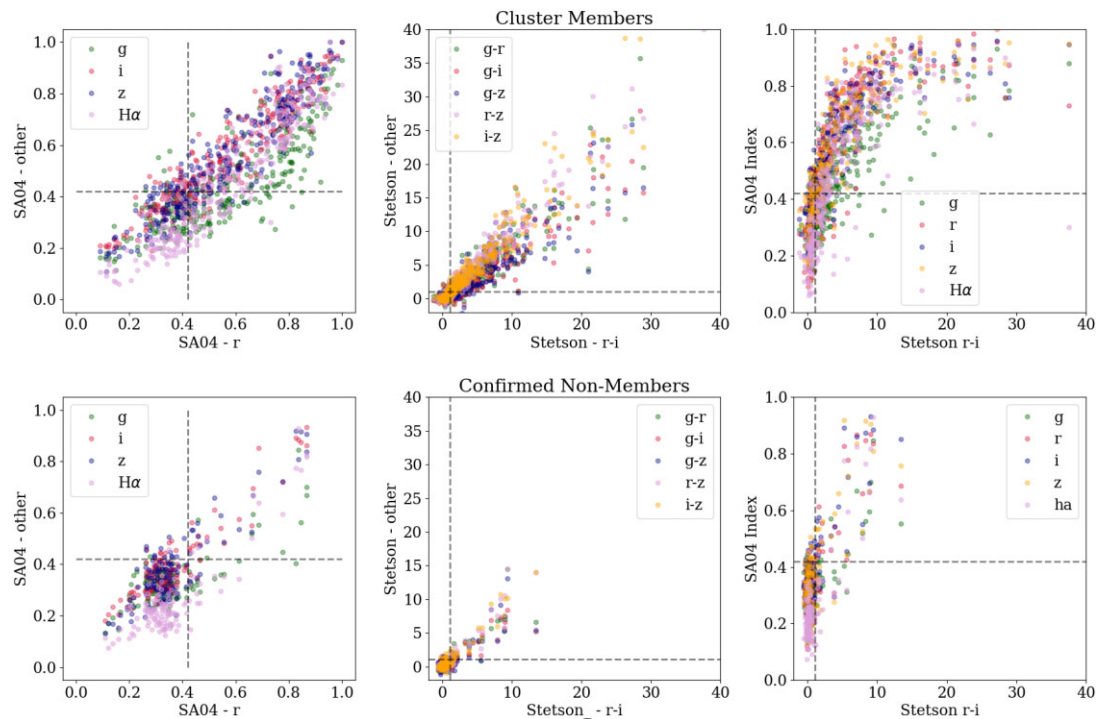
Figure A1. A comparison of various indices for confirmed members (to which we have also added the objects identified with *Gaia*; top panels) and the probably-non-member (bottom panel) from Table 1 in PB23. The dashed lines mark the limits beyond which we consider a star as variable (see Section 4.2). The variability indices are strongly correlated, and we find clear differences between objects known to be members versus those excluded from the membership class that share similar characteristics in terms of magnitude and colour. The number of variable objects is much higher among the highly reliable members, as expected, even though contaminants often include variable background giants.
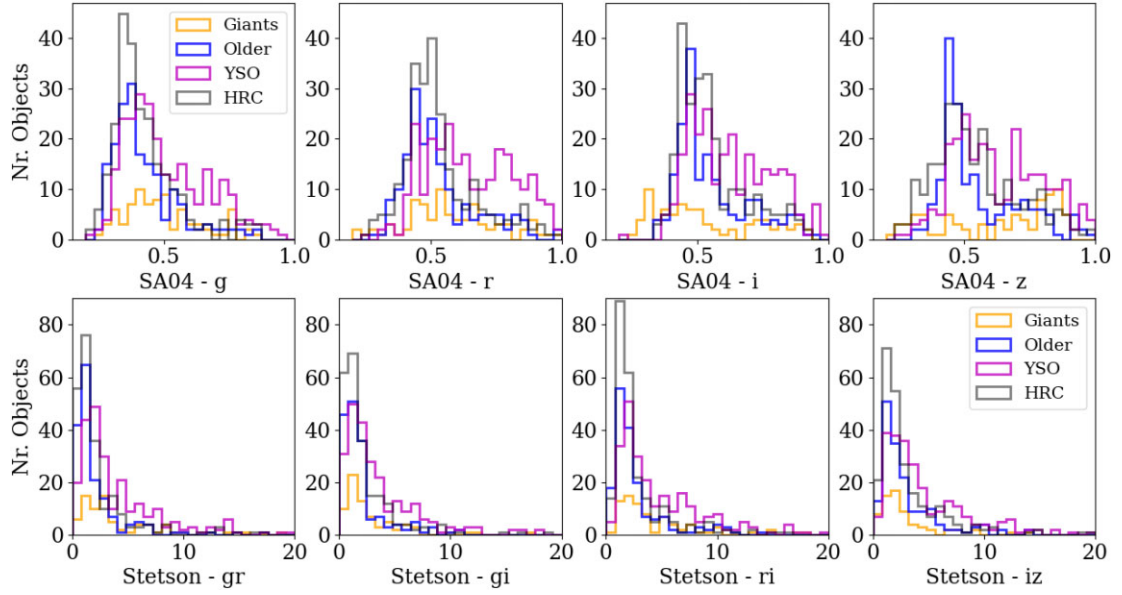
Figure A2. Individual variability indices for the different classes of variables as defined in Section 4.2. The groups refer to candidate giants, YSO, older variables, and HR contaminants. The differences between the various groups are evident.
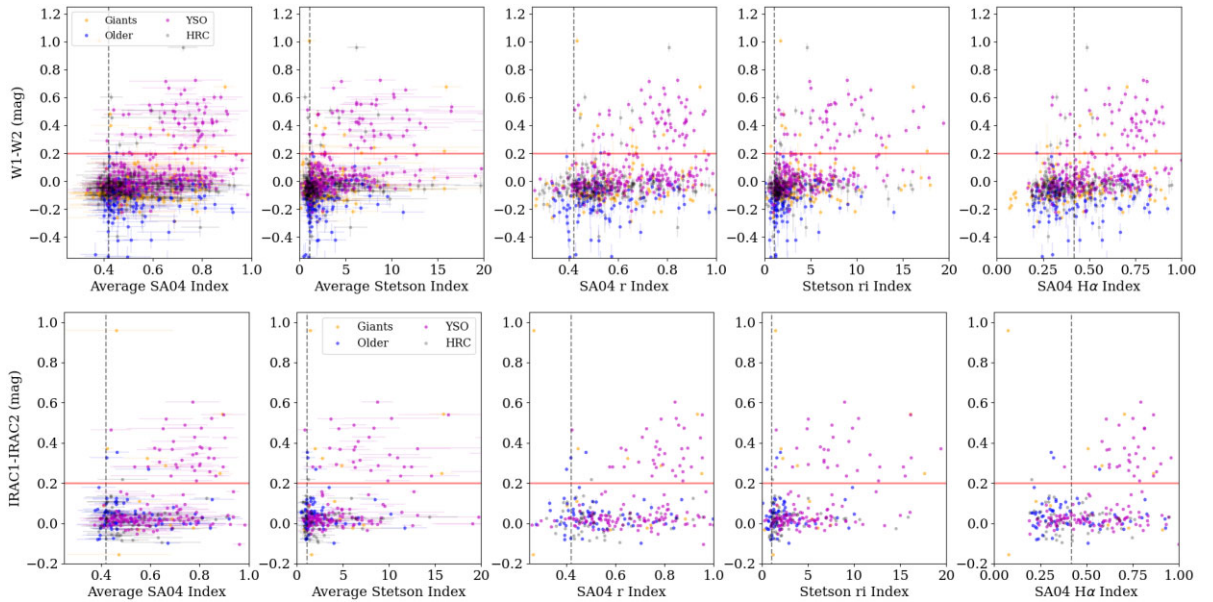

Figure A3. *WISE* W1-W2 (top) and *Spitzer* IRAC 1 – IRAC2 (bottom) colour for stars classified as variable. The type of variable star is the same than in Fig. 5, and thus uses the parallax and position in the colour–magnitude diagram to find candidate YSO, Giants, older field stars, and objects that fall between the 0.5–20 Myr isochrones but do not appear to be at the cluster parallax ('HR contaminants' or HRC). The average SA04 and Stetson indices are shown (the errorbars are the standard deviation of the different indices), as well as the SA04 *r* and Hα indices and the Stetson ri index. The vertical dashed line shows the limits considered for variability for each type of index. The red continuous horizontal line shows the lower limit for an object to be considered as having a disc (PB23). We note that, although many of the variable YSO candidate stars do have discs, not all the variable objects do. There is a trend for YSO with large variability indices to be discy, but a substantial number of objects with large variability have no near-IR excess.
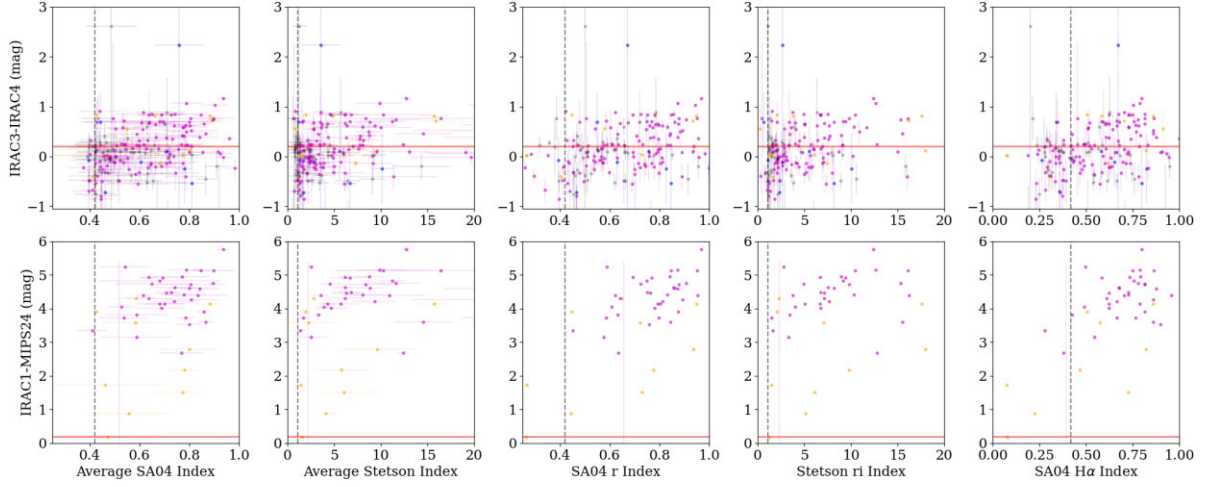
Figure A4. Same as Fig. A3 but for *Spitzer* IRAC bands 3 and 4 (top) and MIPS 24 μm (bottom). Note that the MIPS survey was only complete down to the photospheres of B-type stars at the distance of Tr 37, and hence all the objects detected in the bottom plots have infrared excesses, normally discs, and for that reason, they also tend to be highly variable.

APPENDIX B: THE MAXIMUM-LIKELIHOOD CLUSTERING PROCEDURE

The maximum-likelihood algorithm is based on the technique deployed by Lindegren et al. (2000) and used by Roccatagliata et al. (2018, 2020) to study the structure of Taurus, and by PB23 to study the structure of the IC 1396 region, to which Tr 37 belongs. This technique explores the multidimensional astrometric space (considering parallax and the two proper motions) for a distribution of subclusters, assigning individual stars to subclusters and calculating the maximum likelihood for each set of subclusters. The algorithm uses a customized PYTHON routine and starts with an initial subcluster distribution, taken as a 'guess' from visually inspecting the distribution of objects. This guess is subsequently modified by the algorithm (typically, within a parameter space of $\pm 3\sigma$) and the new likelihood is computed. The final, most significant subcluster structure is determined by maximizing the likelihood function (see equations below).

In this multidimensional space, each j-th subcluster is defined by seven parameters: the mean parallax (ϖ_j), the mean proper motion in right ascension, and declination ($\mu_{\alpha,j}$, $\mu_{\delta,j}$), their intrinsic dispersion for the stars in each subcluster ($\sigma_{\varpi,j}$, $\sigma_{\mu_\alpha,j}$, $\sigma_{\mu_\delta,j}$), and the fraction stars that belong to the subcluster (fs_j). Conversely, for each i-th star, we consider its astrometric parameters parallax (ϖ_i) and proper motion in right ascension and declination ($\mu_{\alpha,i}$, $\mu_{\delta,i}$), with their corresponding uncertainties as well as correlated uncertainties, ρ_i (note that the astrometric parameters are not independent). We then explore the 3D parameter space, where for the location of each star i it is represented by the vector $a_i = [\varpi_i, \mu_{\alpha,i}, \mu_{\delta,i}]$. The positions of the stars are contrasted with the locations of the subclusters, so the astrometric positions for each j-th subcluster are given by the vector $a_j = [\varpi_j, \mu_{\alpha,j}, \mu_{\delta,j}]$.

The probability of each i-th star belongs to the j-th subcluster is given by

$$P_{i,j} = fs_j \frac{L_{i,j}}{L_i}, \quad (B1)$$

where $L_{i,j}$ is the likelihood function for the i-th star belonging to the j-th subcluster, L_i is the total likelihood, and fs_j is the fraction stars of belonging to the j-th subcluster. The individual likelihood $L_{i,j}$ is

calculated as

$$L_{i,j} = (2\pi)^{2/3} |C_{i,j}|^{1/2} \exp\left[-\frac{1}{2}(a_i - a_j)^T C_{i,j}^{-1}(a_i - a_j)\right], \quad (B2)$$

and the sum of all individual likelihoods, multiplied by their corresponding fraction of stars, for the total of *n* subclusters, gives the total likelihood, i.e.

$$L_i = \sum_{j=1}^{n} fs_j L_{i,j}. \quad (B3)$$

Equation (B2) can be understood as a multidimensional Gaussian distance, where $C_{i,j}$ is the covariance matrix, $|C_{i,j}|$ is its determinant, thus

$$C_{i,j} = \begin{bmatrix} C_{ij,11} & C_{ij,12} & C_{ij,13} \\ C_{ij,21} & C_{ij,22} & C_{ij,23} \\ C_{ij,31} & C_{ij,32} & C_{ij,33} \end{bmatrix}, \quad (B4)$$

where the different terms of the matrix correspond to the various uncertainties, including correlated terms, i.e.

$$[C_{ij,11}] = \sigma^2_{\varpi,i} + \sigma^2_{\varpi,j}$$
$$[C_{ij,22}] = \sigma^2_{\mu_\alpha,i} + \sigma^2_{\mu_\alpha,j}$$
$$[C_{ij,33}] = \sigma^2_{\mu_\delta,i} + \sigma^2_{\mu_\delta,j}$$
$$[C_{ij,12}] = [C_{ij,21}] = \sigma_{\varpi,i} \times \sigma_{\mu_\alpha,i} \times \rho_i(\varpi, \mu_\alpha)$$
$$[C_{ij,13}] = [C_{ij,31}] = \sigma_{\varpi,i} \times \sigma_{\mu_\delta,i} \times \rho_i(\varpi, \mu_\delta)$$
$$[C_{ij,23}] = [C_{ij,32}] = \sigma_{\mu_\alpha,i} \times \sigma_{\mu_\delta,i} \times \rho_i(\mu_\alpha, \mu_\delta). \quad (B5)$$

In addition, a_i is the multidimensional vector for the i-th star, $a_i = [\varpi_i \mu_{\alpha,i} \mu_{\delta,i}]$, and a_j is the equivalent vector for the subcluster, $a_j = [\varpi_j \mu_{\alpha,j} \mu_{\delta,j}]$, and ^{T} indicates transposition. Therefore,

$$a_i - a_j = \begin{bmatrix} \varpi_i - \varpi_j \\ \mu_{\alpha,i} - \mu_{\alpha,j} \\ \mu_{\delta,i} - \mu_{\delta,j} \end{bmatrix} \quad (B6)$$

The total likelihood is then computed as the sum of all likelihoods for all the stars. In our case and for the total population, the maximum likelihood is 823.4.

APPENDIX C: A NOTE ON ACCRETION RATES FROM Hα FLUXES

Measuring accretion rates is in general complicated, since a very good knowledge of the source is required (e.g. Sicilia-Aguilar et al. 2010) and, moreover, numerous calibrations exist regarding the conversion between excess luminosity (in photometry or lines) and accretion luminosity (e.g. Fang et al. 2009; Alcalá et al. 2014). In general, the more indirect the method is, the higher the risk of miscalibrations (Mendigutía et al. 2015). Ahead of the North-PHASE complete survey, we explored the existing correspondence between accretion rates from U band and Hα, using the data for Tr 37. Fig. C1 shows a comparison between the values from Sicilia-Aguilar et al. (2010), obtained via U-band excess, and those from Barentsen et al. (2011), derived from Hα photometry, which reveals a systematic offset, the values from Hα being systematically higher by a factor of few, compared with those derived from U-band.

To check if the difference lay in the way the Hα EW is estimated from photometry versus the spectral value, we compared the Hα EW from Sicilia-Aguilar et al. (2006a); Nakano et al. (2012); Sicilia-Aguilar et al. (2013), measured from spectra, and those from photometry (Barentsen et al. 2011). This reveals no significant differences, albeit a large level of variability for individual objects, which had been also documented when comparing the spectra themselves (Sicilia-Aguilar et al. 2006a, 2013). The results are shown in Fig. C2. Nevertheless, the depth of Hα-based and U-excess-based measurements of the accretion rate is not comparable, so variable accretion together with detection limits may induce at least part of this behaviour.

A similar offset is so far found when estimating Hα EW from the North-PHASE photometry (see Fig. C3). In this case, a strong limitation lies in the lack of information on the individual spectral types and extinctions, which will be a particularly large problem for red stars (both due to extinction or to effective temperature). Since North-PHASE will offer constraints on the stellar properties once variability is well-determined, this issue will be addressed and further explored in the future.

Regarding the accretion rates, lower rates seem more consistent in general, especially because very high rates start being in conflict
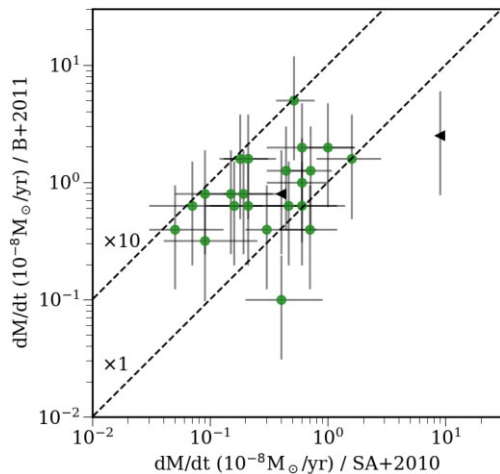

Figure C1. A comparison of the accretion rates as derived from u-band photometry (x-axis; Sicilia-Aguilar et al. 2010) and Hα photometry (y-axis; Barentsen et al. 2011). The generalized offset of accretion rates calculated from photometry and from U-band excesses reveals that we need to be cautious with the calibration.
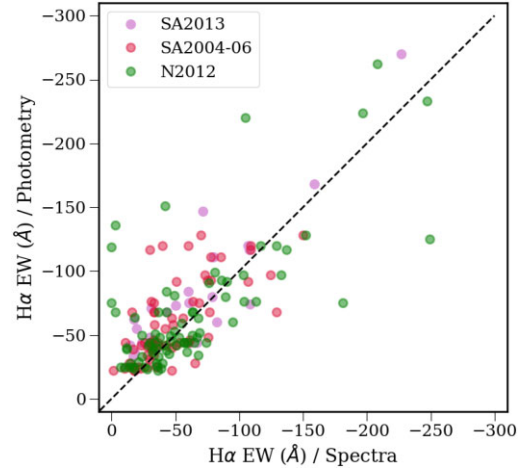

Figure C2. A comparison of Hα EW measured from spectra (x-axis; Sicilia-Aguilar et al. 2006a, 2013; Nakano et al. 2012) versus IPHAS photometry (y-axis Barentsen et al. 2011). The dashed line shows the identity result. Although there are significant differences between measurements of the same object taken on different dates, even if the same method is used, and there is a slight tendency for weak Hα to be missed by photometry, the agreement is reasonable.
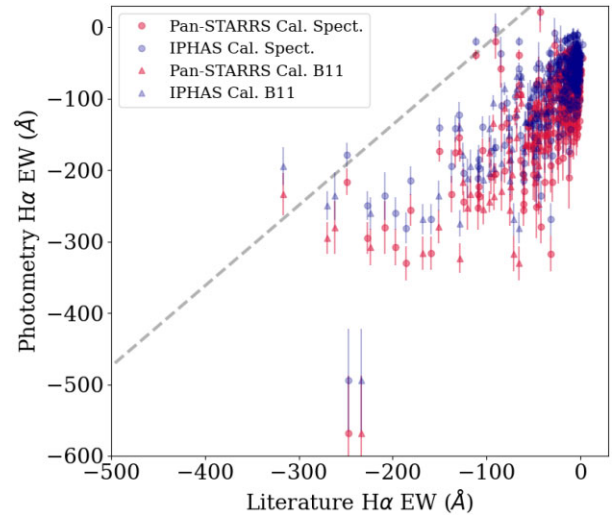

Figure C3. A comparison of the first-check Hα EW from North-PHASE. The x-axis contains measurements from the literature (see the text, circles correspond to spectral measurements, while triangles denote photometric estimates). In red, we show the results using Pan-STARRS data for calibration of the rSDSS filter, while in blue we show the same results but using rSDSS from IPHAS to calibrate our rSDSS band. The Hα data are always calibrated with IPHAS. Although this 'rough estimate' is clearly affected by systematics, we find a clear correlation which demonstrates that the North-PHASE J0660 photometry can be used to estimate the strength of the Hα line.

with disc masses and stellar ages (Spezzi et al. 2012; Sicilia-Aguilar et al. 2016). On the other hand, the Hα measurements from spectra and photometry appear to be very robust and, notwithstanding the source variability, indistinguishable. The reasons for the difference in accretion rates could be related to issues with the calibration, since line fluxes are always more uncertain than UV excesses to get accretion, and even more if they come from photometry. But, since the probed range of accretion rates is not too different from the typical factor-of-few variability observed in individual objects (and from

the typical uncertainties; Sicilia-Aguilar et al. 2010) there is also the possibility of selection effects, since Hα photometric measurements pick more extreme cases/objects thus favouring the detection of objects when they are at a higher accretion state. Although at this stage it is difficult to find a cause for these differences, the variability aspect of North-PHASE will offer a great opportunity to track the changes in Hα photometry and improve the calibration in the future.

[1]*SUPA, School of Science and Engineering, University of Dundee, Nethergate DD1 4HN, Dundee, UK*

[2]*Departamento de Fíica Teórica, Facutad de Ciencias, Universidad Autónoma de Madrid, Cantoblanco, E-28049 Madrid, Spain*

[3]*Universidad Nacional Mayor de San Marcos, Facultad de Ciencias Físicas, Unidad de Postgrado, Ciudad Universitaria, Lima 15801, Peru*

[4]*Alma Mater Studiorum, Università di Bologna, Dipartimento di Fisica e Astronomia (DIFA), Via Gobetti 93/2, I-40129 Bologna, Italy*

[5]*INAF – Osservatorio Astrofisico di Arcetri, Largo E. Fermi 5, I-50125 Firenze, Italy*

[6]*Centre for Astrophysics and Planetary Science, School of Physics and Astronomy, University of Kent, Canterbury CT2 7NH, UK*

[7]*Centro de Estudios de Física del Cosmos de Aragón (CEFCA), Plaza San Juan 1, E-44001 Teruel, Teruel, Spain*

[8]*European Southern Observatory, Karl-Schwarzschild-Strasse 2, D-85748 Garching bei München, Germany*

[9]*Steward Observatory, University of Arizona, 933 N. Cherry Ave., Tucson, AZ 85721-0065, USA*

[10]*Centro de Astrobiología (CAB), CSIC-INTA, Camino Bajo del Castillo s/n, Villanueva de la Cañada, E-28692 Madrid, Spain*

[11]*Purple Mountain Observatory, Chinese Academy of Sciences, 10 Yuanhua Road, Nanjing 210023, China*

[12]*SUPA, School of Physics and Astronomy, University of St Andrews, North Haugh, St Andrews KY16 9SS, Scotland, UK*

[13]*Institute for Astronomy (IfA), University of Vienna, Türkenschanzstrasse 17, A-1180 Vienna, Austria*

[14]*Konkoly Observatory, HUN-REN Research Centre for Astronomy and Earth Sciences, MTA Centre of Excellence, Konkoly Thege Miklós út 15-17, 1121 Budapest, Hungary*

[15]*Institute of Physics and Astronomy, ELTE Eötvös Loránd University, Pázmány Péter sétány 1/A, 1117 Budapest, Hungary*

[16]*INAF – Osservatorio Astrofisico di Catania, Via S. Sofia 78, I-95123 Catania, Italy*

[17]*Max Planck Institute for Astronomy, Königstuhl 17, D-69117 Heidelberg, Germany*

This paper has been typeset from a TEX/LATEX file prepared by the author.