MNRAS **000**, 1–9 (2017) Preprint 7 March 2017 Compiled using MNRAS LATEX style file v3.0

Chemical Tagging of Clusters in the GALAH Survey: Two Populations and New Members in the Pleiades

Janez Kos,[1⋆] Joss Bland-Hawthorn,[1] Ken Freeman,[2] Gayandhi M. De Silva,[3,1] Martin Asplund,[2] Sanjib Sharma,[1]
[1] *Sydney Institute for Astronomy, School of Physics, A28, The University of Sydney, NSW, 2006, Australia*
[2] *Research School of Astronomy & Astrophysics, Australian National University, ACT 2611, Australia*
[3] *Australian Astronomical Observatory, North Ryde, NSW 2133, Australia*

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

The technique of chemical tagging uses the elemental abundances of stellar atmospheres to 'reconstruct' homogeneous star clusters that have long since dissolved. The stellar abundance space $\mathcal{C}$([Fe/H], [X_i/Fe]) has up to 30 elements or dimensions – although many of these dimensions are not independent – in the GALAH spectroscopic survey, which aims for one million stars based at the Anglo-Australian Telescope. How to find clustering reliably in a noisy high-dimensional space is a difficult problem that remains largely unsolved. Here we explore *t-distributed stochastic neighbour embedding* (t-SNE) which identifies an optimal mapping of a high-dimensional space into fewer dimensions while conserving the original clustering which can be identified by eye in a two-dimensional map. We show that this method is a reliable tool for chemical tagging as it can: (i) resolve clustering in the $\mathcal{C}$-space alone, (ii) recover known open and globular clusters with high efficiency and low pollution, (iii) relate field stars to known clusters. t-SNE also provides a useful visualization of a high-dimensional space. We demonstrate the method on a dataset of 13 abundances measured in the spectra of 187,000 stars by the GALAH survey. We recover most of the 9 observed clusters (6 globular and 3 open clusters) in $\mathcal{C}$-space with minimal pollution from field stars and low number of outliers. One association – the Pleiades – appears to be two distinct, co-moving chemical groups reminiscent of a merged binary cluster. With the aid of chemical tagging, we also identify two Pleiades supercluster members (which we confirm kinematically), one as far as 6° away from the cluster centre.

Key words: methods: data analysis — stars: abundances — open clusters and associations — open clusters and associations: individual (Pleiades)

1 INTRODUCTION

The use of chemical tagging in galactic archaeology was first proposed by Freeman & Bland-Hawthorn (2002). They suggested that the abundances of elements in stars can be used as unique signatures over their lifetime to 'reconstruct' stellar groups that have long since dissolved. Theoretical arguments indicate that chemical homogeneity (with the exception of light elements) is guaranteed in open clusters up to 10^5 $M_\odot$ and in globular clusters up to a limit of 10^7 $M_\odot$ (Bland-Hawthorn et al. 2010a). In practice, the degree of homogeneity may depend on the initial abundance spread in the collapsing cloud (Feng & Krumholz 2014). But, to date, essentially all open clusters appear to be chemically homogeneous to better than 0.1 dex (De Silva et al. 2006; Sestito et al. 2007; Bovy 2016). Both young and ancient (up to ∼9 Gy) open clusters appear to be chemically homogeneous (De Silva et al. 2006, 2007) indicating that pollution from the interstellar medium does not wipe out this information. For chemical tagging to be feasible for field stars, a large amount of high quality data has to be collected, i.e. on the order of 10^6 observed stars and ∼30 measured elements (Bland-Hawthorn & Freeman 2004; Ting et al. 2015). Indeed, these are the design goals of the GALAH[1] survey on the HERMES instrument at the Anglo-Australian Telescope (Barden et al. 2010; De Silva et al. 2015; Martell et al. 2017). This requirement can be much lower for 'soft' chem-

⋆ E-mail: janez.kos@sydney.edu.au

[1] GALAH survey webpage is `https://galah-survey.org`

ical tagging if there is additional information (e.g. kinematics, location) to associate the stars.

To chemically tag stars, one has to search for clustering in the $\mathcal{C}$-space, i.e. a N-dimensional space determined by the measured number of elemental abundances. Strictly speaking, these dimensions are unlikely to be independent, e.g. iron-peak elements are strongly coupled. In the GALAH survey, there are up to 30 elements that can be measured for each star but in our study we will concentrate on a smaller number ($N = 13$) of elements with well determined abundances.

How are we to find substructures in a high dimensional space? The human brain is excellent at detecting clustering in three or fewer dimensions, but falls short in more dimensions. Most work to date has focussed on finding clusters in the original N-dimensional space. For example, Hogg et al. (2016) searched for chemical groups in the APOGEE data by the k-means algorithm and showed that some clusters correspond to groups in the phase space. Blanco-Cuaresma et al. (2015) utilized PCA to distinguish between known clusters in a sandbox study. Bland-Hawthorn et al. (2010b) use a density-based hierarchical clustering algorithm and introduce the S-statistic to show that clustering exists in a simulated dwarf galaxy. Mitschang et al. (2014); Quillen et al. (2015) used a probabilistic approach to resolve chemical groups in a blind chemical tagging study, but was unable to make any conclusions about the their origin. The simplest method of all is often used: by analysing the abundances and some relations between them, Martell et al. (2016) were able to visually identify halo stars that originate in globular clusters and de Silva et al. (2011) related Hyades supercluster members to one chemical group. A more advanced algorithm that also provides the visualisation was used by Jofre et al. (2016) who applied a method of evolutionary trees to stellar abundances and produced phylogenetic tree for 21 solar twins and the Sun.

One of the most successful methods in recent years exploits the huge computational power now available in desktop computers. The so-called t-distributed stochastic neighbour embedding (t-SNE) algorithm is a remarkable technique for reducing the dimensions of a problem (van der Maaten & Hinton 2008). Each high-dimensional data point is embedded into usually two dimensional 'visualization' space where 'similar' points are kept together and 'dissimilar' points are moved apart. Once the problem is reduced into two dimensions, the clustering can be identified by eye. We find that this method is highly effective in identifying known and unknown cluster members. The method has some limitations: (i) it is a black box that is difficult to tune or control; and (ii) the abundance measurement errors are not used in the present application. But the method is extraordinarily powerful as demonstrated in recent papers, for example, to efficiently identify peculiar stars and stellar populations in large surveys with a high level of completeness (Matijevič & The RAVE Collaboration 2015; Lochner et al. 2016; Valentini et al. 2016; Traven et al. 2016).

We describe our data in Section 2 and the method in Section 3. In Section 4 we explore the efficiency of our method on 9 clusters and in Section 5 we present a more detailed analysis of the Pleiades cluster. In Section 6 we discuss the implications of this work and future development of the field.

2 THE DATA

The data set analysed here has been drawn from three programs: the GALAH pilot program, the K2 follow-up survey, and the main GALAH survey (De Silva et al. 2015; Martell et al. 2017). Different programs have different selection functions, but share the same observing procedures, reduction pipeline, and analysis pipeline (Kos et al. 2017). Stars from all programs are analysed together so the stellar parameters and the abundances are comparable. All stars used in this paper have abundances measured by The Cannon (Ness et al. 2015) and trained on observed benchmark stars and some K2 stars with known seismic gravities (Stello et al. 2016) and abundances measured by the SME (Valenti & Piskunov 1996; Piskunov & Valenti 2016).

The complete data set consists of 187,640 stars, mostly dwarfs, observed before Jan 1 2016. 15,601 stars have unreliable stellar parameters and were excluded from the study. As this is the first time The Cannon has been used with GALAH data and the first internal release of abundances, the uncertainties of the measured abundances have not been determined yet. A map of the observed fields on the celestial sphere is given in Martell et al. (2016).

Our data-set consists of abundances of 13 elements (Na, Mg, Al, Si, K, Ca, Sc, Ti, Cr, Fe, Ni, Cu, Ba) measured by The Cannon as well as stellar parameters ($T_{\rm eff}$, $\log g$, [Fe/H], and radial velocity) measured by fitting synthetic models of stellar atmospheres to one dimensional GALAH spectra (Kos et al. 2017). Abundances for all 13 elements were measured for each of 187,640 stars.

Proper motions from UCAC4 are also available for all stars and parallaxes from Gaia for some. For most stars we calculated the photometric distances following (Zwitter et al. 2010).

3 T-DISTRIBUTED STOCHASTIC NEIGHBOUR EMBEDDING

t-distributed stochastic neighbour embedding (t-SNE) is an algorithm from a family of manifold learning algorithms. It can project high-dimensional data into lower-dimensional (for practical reasons usually two-dimensional) space, preserving the intrinsic structure (van der Maaten & Hinton 2008). t-SNE has seen extensive use in data science and made a break into astronomy as a classification algorithm (Matijevič & The RAVE Collaboration 2015; Lochner et al. 2016; Valentini et al. 2016; Traven et al. 2016), same as other manifold learning algorithms (e.g. Vanderplas & Connolly 2009; Daniel et al. 2011; Bu et al. 2014). We extend its use as a pure manifold learning algorithm to find structure in a 13-dimensional $\mathcal{C}$-space.

t-SNE's input is a set of N high-dimensional objects $\mathbf{x}_1, \ldots, \mathbf{x}_N$. In our case each $\mathbf{x}_i$ will be a collection of 13 abundances for a single star:

$$\mathbf{x}_i = \left(\left[\frac{\rm Na}{\rm Fe}\right]_i, \left[\frac{\rm Mg}{\rm Fe}\right]_i, \ldots, \left[\frac{\rm Ba}{\rm Fe}\right]_i \right) \quad (1)$$

Following van der Maaten & Hinton (2008), we first calculate similarities $\mathbf{p}_{ij}$ of the input set:

$$\mathbf{p}_{ij} = \frac{\mathbf{p}_{i|j} + \mathbf{p}_{j|i}}{2N}, \quad \mathbf{p}_{j|i} = \frac{\exp\left(-||\mathbf{x}_j - \mathbf{x}_j||^2/2\sigma_i^2\right)}{\sum_{k\neq i} \exp\left(-||\mathbf{x}_j - \mathbf{x}_k||^2/2\sigma_i^2\right)} \quad (2)$$

Table 1. Clusters with observed members in the GALAH and K2 surveys.

Cluster	N of stars	Type	Notes
47 Tuc	90	GC	Membership from Tucholke (1992).
M30	4	GC	
M67	113	OC	Membership from Geller et al. (2015). 404 spectra of 113 unique stars.
NGC288	14	GC	
NGC362	27	GC	
NGC1851	7	GC	
NGC2516	3	OC	Membership from Jeffries et al. (2001).
Pleiades	27	OC	
ω Cen	230	GC	246 spectra of 230 unique stars.

GC=globular cluster
OC=open cluster

where σ_i is a parameter that depends on the perplexity and local density of the data-set. We aim to produce a lower dimensional map with objects $\mathbf{y}_1, \ldots, \mathbf{y}_N$ with similarities:

$$\mathbf{q}_{ij} = \frac{(1 + ||\mathbf{y}_i - \mathbf{y}_j||^2)^{-1}}{\sum_{k \neq i}(1 + ||\mathbf{y}_k - \mathbf{y}_i||^2)^{-1}} \quad (3)$$

To find the optimal mapping where $\mathbf{q}_{ij}$ reflects $\mathbf{p}_{ij}$ as well as possible, we minimize the Kullback-Leibler divergence:

$$KL(P||Q) = \sum_{i \neq j} \mathbf{p}_{ij} \log \frac{\mathbf{p}_{ij}}{\mathbf{q}_{ij}}. \quad (4)$$

Kullback-Leibler divergence is a non-convex function that is minimised by gradient descend initialised randomly. Different runs of t-SNE, even when same parameters are used, can therefore result in a different mapping. The algorithm is usually run several times and the mapping with lowest Kullback-Leibler divergence is used.

The scale of the map produced by t-SNE is irrelevant. Only the relative relations between objects and groups on the map hold any information. We will therefore refrain from plotting the coordinate system on the maps at all.

The above algorithm has time dependence of $\mathcal{O}(N^2)$, because we have to calculate similarities for every pair of objects. This is impractical for most applications, so we employ the Barnes-Hut algorithm to calculate sparse similarities in $\mathcal{O}(N \log(N))$ time (van der Maaten 2013). With such optimization we can analyse our whole dataset on an average desktop computer in $\lesssim 1$ hour.

4 RECOVERING KNOWN CLUSTERS IN THE CHEMICAL SPACE

The GALAH survey targeted only a small number of individual clusters as part of the pilot survey. Other clusters were observed because their stars happen to fall into the magnitude range of the main survey. A list of the observed clusters is given in Table 1. Clusters without membership catalogues were analysed by us; membership was determined by applying cuts in radial velocity, positions, and in some instances, through their proper motions.

Figure 1 shows the abundances for all 13 elements in the observed clusters. Notice that the scatter for some elements is consistently high, regardless the cluster. Around half of the scatter is statistical noise. We demonstrate that by measuring the uncertainties of the abundances from repeated observations of field stars and M67. Only measurement made from spectra of SNR$\geq$ 45 per pixel were used in order to distinguish between repeats done with the purpose of quality estimate and repeats done to boost the SNR of some lower quality data. The rest of the scatter is systematic arising from the abundance determination pipeline being sensitive to temperature variation or dwarf–giant distinction.

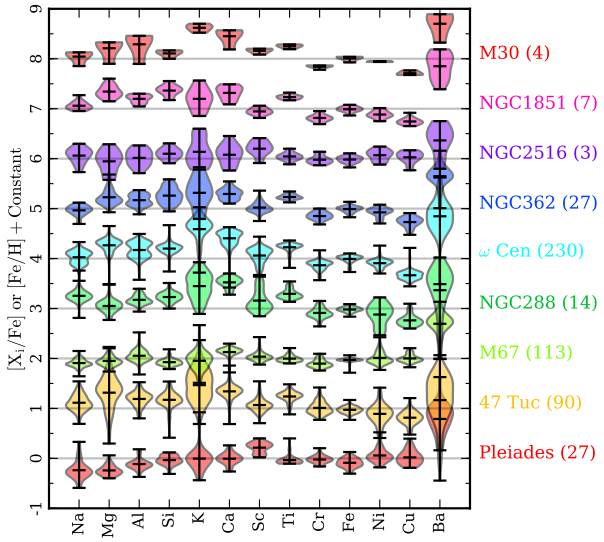

Figure 1. Abundances of 13 elements in 9 studied clusters. A violin plot represents the distribution of measured abundances for all stars that we identified as cluster members. Number of all stars in each cluster is given next to the cluster name. Note how some elements have consistently more scattered distribution.

The scatter in the Ba and K abundances is highest. Elements like Fe, Ti, and Cr have lower uncertainties. It is therefore not fair to treat elements with different uncertainties as equally important dimensions in the $\mathcal{C}$-space. Before we use the abundances in t-SNE, we standardize them so that the distribution of abundances of every element has a zero median and a standard deviation of a unity. Standardization is done once for the complete data-set. Then we change the standard deviation of the standardized set based on the weights that are proportional to the scatter we observe in clusters. Elements with more scatter will have a narrower distribution, so the distances in those dimensions will always be damped and will not carry as much importance as those for less scattered elements. Because it is hard to quantitatively determine the weight for each element, we will distribute the elements into 4 groups. Ba and K have by far the highest scatter, so they will be given a weight of 0.25. Mg and Ca also have high scatter, so they will have weights equal to 0.5. Fe, Ti, Cr, and Cu have the smallest scatter and will have a weight of 2.0, and the rest of the elements will have a weight of 1.0. Weights, uncertainties measured from the repeated observations, scatter in clusters, and related weights are collected in Table 2. Weights are a way to

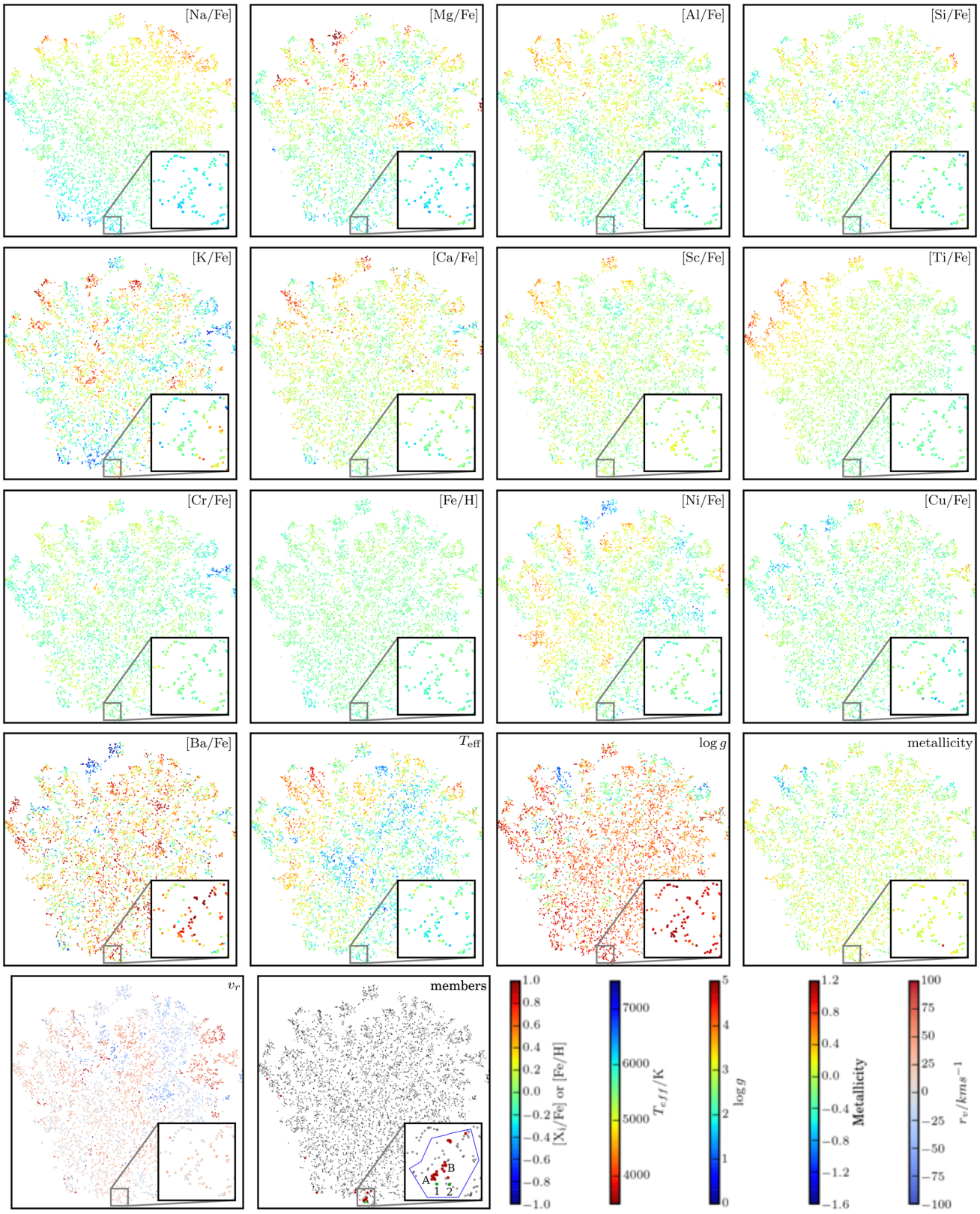

Figure 2. t-SNE projection of 9408 stars in a 40° radius around Pleiades. Abundances of 13 elements used to create the projection are colour-coded. T_{eff}, $\log g$, metallicity and radial velocity colour-codes are also plotted. The panel labeled "members" shows the stars that belong to the cluster in red and field stars in grey. Two stars marked in green and numbered 1 and 2 are newly discovered Pleiades members discussed in Section 5. 17 out of 27 Pleiades stars lie in the two tight groups in the bottom of the map marked A and B. The blue polygon marks the Pleiades' chemical group (see Figure 5)

Table 2. Uncertainties measured from 1579 repeated observations in the whole GALAH sample and 377 repeated observations of M67 stars compared to scatter observed in Figure 1, and weights assigned to each element. Uncertainties and scatters are expressed as standard deviations.

Element	Uncertainty from all repeats	Uncertainty from M67 repeats	Scatter in all clusters	Weight
	dex	dex	dex	
Na	0.063	0.063	0.122	1.0
Mg	0.078	0.079	0.162	0.5
Al	0.066	0.063	0.129	1.0
Si	0.053	0.053	0.106	1.0
K	0.099	0.129	0.228	0.25
Ca	0.065	0.056	0.131	0.5
Sc	0.050	0.054	0.115	1.0
Ti	0.044	0.048	0.071	2.0
Cr	0.047	0.049	0.081	2.0
Fe	0.024	0.021	0.060	2.0
Ni	0.056	0.057	0.112	1.0
Cu	0.049	0.036	0.095	2.0
Ba	0.114	0.135	0.230	0.25

implement uncertainties into the t-SNE, as in our case the uncertainties of individual measurements have not been estimated. Without these weights, there would be less groups in the t-SNE map and the stars from known clusters would end up scattered over a larger area.

We use the weighted abundances to produce a t-SNE projection for a region around Pleiades (Figure 2) and other clusters (Apendix A). In case we have more than one measurement for a star, like for most M67 stars and some ω Cen stars, we first calculated the average abundances for each star and used those in the t-SNE. Stars with repeated observations are therefore only plotted once in the t-SNE maps. No other information is used in the projection, even-though other stellar parameters are displayed in the color-coded t-SNE maps. One can pick out many groups in the t-SNE map, some more pronounced than the others. Groups associated with each cluster are marked and we leave a detailed analysis of other pronounced groups for a different study.

Figure 3 shows only the members on the t-SNE maps for 8 remaining clusters. Out of all nine clusters, we claim that t-SNE gives good results for all but two of them. In NGC2516 the membership of three stars is not completely certain, so we can not base our conclusions on this cluster. Note that we only cover an edge of the cluster in one of the observed fields, so a low number of members is expected. We did not find any big groups in the $\mathcal{C}$-space for 47 Tuc, so the chemical tagging of this cluster was unsuccessful. It is well known that 47 Tuc is not a chemically homogeneous cluster (e.g. Thygesen et al. 2014). We indeed observe an unusual scatter in light elements that is higher than for other clusters.

5 CHEMICAL POPULATIONS AND NEW MEMBERS OF PLEIADES

Pleiades is a young (Brandt & Huang 2015) cluster for which we expect to find some members well away from the centre of the cluster, yet close enough that we can focus only on a small region around the cluster (Kroupa & Boily 2002). The tidal radius of Pleiades is $\sim 6^\circ$ (Adams et al. 2001), and we do not expect to find any members at the distances much larger than this, considering the number of observed stars. Even-though we focus our effort into a 40° radius around the Pleiades to demonstrate the method on a larger number of stars. Pleiades are in one of the most northern fields GALAH explored, in one of the K2 fields, so the 40° radius region includes mostly K2 fields, a few pilot survey fields and some regular fields at $\delta < +10^\circ$ that have been observed and most of the region has no observations at all. The Pleiades members were identified by us by making cuts in the position, radial velocity, and proper motions. This way we identified 27 members. We choose Pleiades for this experiment, because they are the only young cluster with distinct kinematics we can use to verify new member candidates.

After making the t-SNE map we can see in Figure 2 that most Pleiades stars fall into two closely related and tight groups with not many polluting field stars. The two groups have slightly different abundances of [Sc/Fe], [Ba/Fe], and [Fe/H] (Figure 4), thus a small separation between the two groups. Stars from both groups are within a 0.5 dex in $\log g$ and within 1000 K in $T_{\rm eff}$ where stars from group A are in average hoter than stars from group B. It is hard to decide whether the separation is real or artificially induced, as we do not have good enough estimates of the abundance uncertainties for individual stars yet. There are similar observations in the literature (Gebran & Monier 2008) matching our large scatter in [Sc/Fe] and [Ba/Fe] abundances, but not confirming two separate groups, which could only be due to low number of observed stars. Binary clusters do exist (Slesnick et al. 2002) and it can be speculated that Pleiades might be a merged cluster, based on two separate chemical groups. However any decisive conclusions will require more data and more careful analysis of abundances. In any case we show that features like split chemical groups can be picked out by t-SNE while still conserving the hierarchy and putting both groups close together. We define the Pleiades chemical group by combining all small groups with at least one known Pleiades member that are close to the main group. This decision is arbitrary but conservative. The chemical group is marked with a blue polygon in Figure 2.

We expect some polluting field stars to be in the same chemical group, because 13 measured abundances is not enough to completely isolate the cluster (Ting et al. 2012; Mitschang et al. 2012; Ting et al. 2015). Therefore we can not claim that all the stars in the chemical group are Pleiades members. For clusters with adequate kinematics information, however, 13 abundances are enough, as we can use independent dimensions: radial velocity, amplitude of the proper motion and direction of the proper motion. We also have photometric information that we combine into a single parameter: the distance. These 4 additional dimensions are enough to select from the polluting stars only those with the kinematics and distances that match Pleiades'. This leaves us with two stars that we claim are candidates for Pleiades members. The step of reducing ~ 30 polluting stars into 2 member candidates is illustrated in Figure 5.

There are actually 2 more stars in the whole 40° radius region that match the Pleiades kinematics. They are both $> 20^\circ$ away from the cluster and do not fall near the

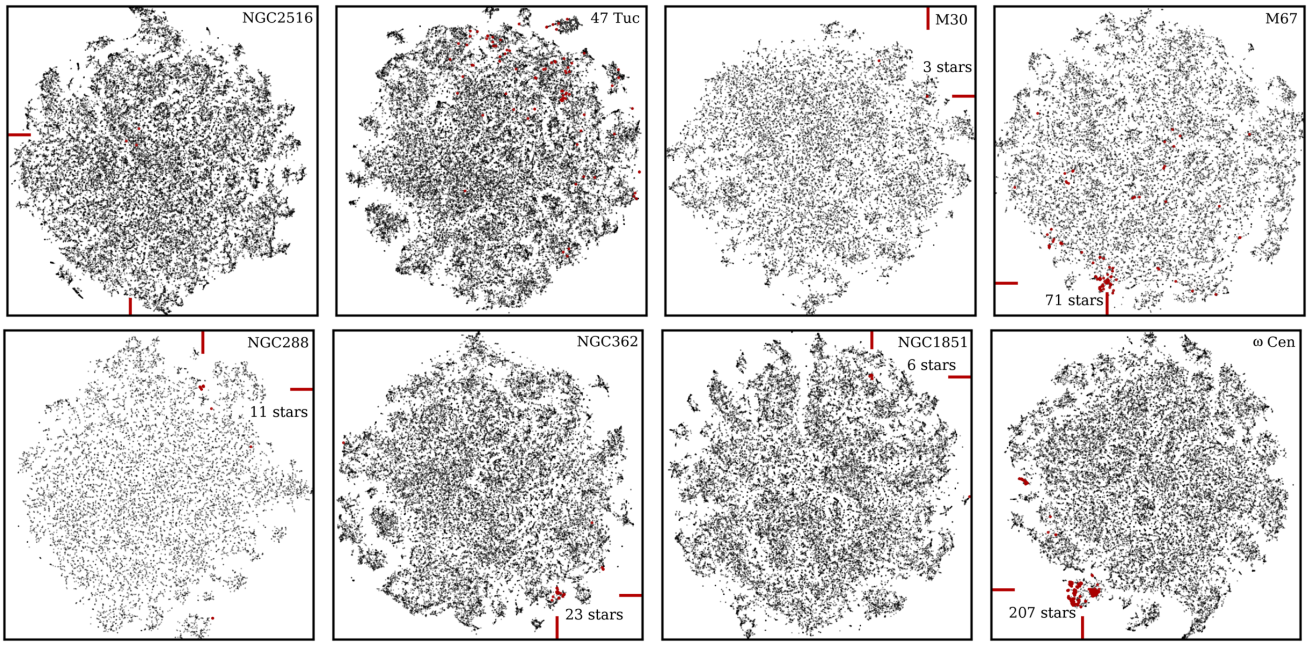

Figure 3. t-SNE projections for regions around the 8 remaining clusters. Known members are marked in red and the number in some panels tells the number of the stars in the main group, as the points often overlap. Top-left, to bottom-right the following clusters are shown: NGC2516 (41,106 stars in a 30° radius), 47 Tuc (44,037 stars in a 35° radius), M30 (20,254 stars in a 35° radius), M67 (25,648 stars in a 45° radius), NGC288 (11,535 stars in a 45° radius), NGC362 (41,578 stars in a 35° radius), NGC1851 (33,882 stars in a 35° radius), and ω Cen (33,281 stars in a 30° radius). Red dashes on the edge of each panel point toward the position of the main group of members.

Pleiades' chemical group in the t-SNE map. This means that after reducing the number of stars from $\sim$ 9400 and $\sim$ 2 coincidental stars to $\sim$ 30 stars by chemically tagging the cluster, we expect to find $\sim$ 0.0064 stars that by chance have the same kinematics as the Pleiades and fall into the cluster's chemical group. We found two, which are therefore Pleiades members with a high degree of certainty. It must be noted, that star number 2 is a known supercluster candidate (Mermilliod et al. 1997) that escaped our cluster membership determination for being too far from the cluster centre. Star number 1, however, has no relation to Pleiades in the literature.

6 DISCUSSION

We show that t-SNE is an appropriate algorithm to search for clustering in the $\mathcal{C}$ space by demonstrating its performance on all clusters we observed in GALAH. With reasonable exceptions the method performs well, which we further demonstrate by chemically tagging Pleiades members and discovering new members without previous knowledge of their existence.

Maybe even more important conclusion is that extremely precise abundances are not always needed to successfully chemically tag cluster members or maybe even field stars. It turns out that the ability to find structures in the $\mathcal{C}$ space is more valuable. Precise abundances help reducing number of outliers and prominence of the chemical groups, but we show that the groups exist and can be isolated even when chemical homogeneity is only in the order of 0.1 dex. This is best shown by being able to match the two new Pleiades members to the cluster, even when the abundances do not always agree with the abundances of Pleiades. Even with large discrepancy in [Cu/Fe] the two stars still lie close to the Pleiades manifold and are therefore correctly identified as members.

A nice demonstration of the t-SNE superiority is the hirearchy seen in some clusters where more than one population is found. Different populations form different groups, but they all compose one larger group that includes the majority of the cluster members. In a two-dimensional map one can easily see and correctly interpret these structures. This is very hard to do in a higher-dimensional space without a good visualization of all dimensions. Build-in hirearchy also saves us from tagging dwarfs and giants separatelly, like some studies in the literature do. It can be seen from the $\log g$ panel in Figure 5 and even better from the figures in the Appendix A that giants are mostly separated from predominantly dwarf populated t-SNE maps. t-SNE knows nothing about $\log g$ and the result is purely a consequence of abundances being dependent on gravity. This can be either an abundance pipeline issue or a result of different stellar populations observed.

Adopting the Pleiades parallax of $\varpi = 4.45 \pm 0.49$ mas (Gaia Collaboration et al. 2016), median proper motion of our known Pleiades members ($\mu = 48.96$ mas y^{-1}), and proper motion of the newly discovered members, we can calculate, that the stars have been scattered out of the cluster $7.9 \pm^{\infty}_{4.6}$ My (star 1) and 0.68 ± 0.05 My (star 2) ago at a velocity significantly larger than the escape velocity of the cluster. Considering where on the HR diagram the two stars lie, they are indeed good candidates to be ejected from the cluster due to their low mass.

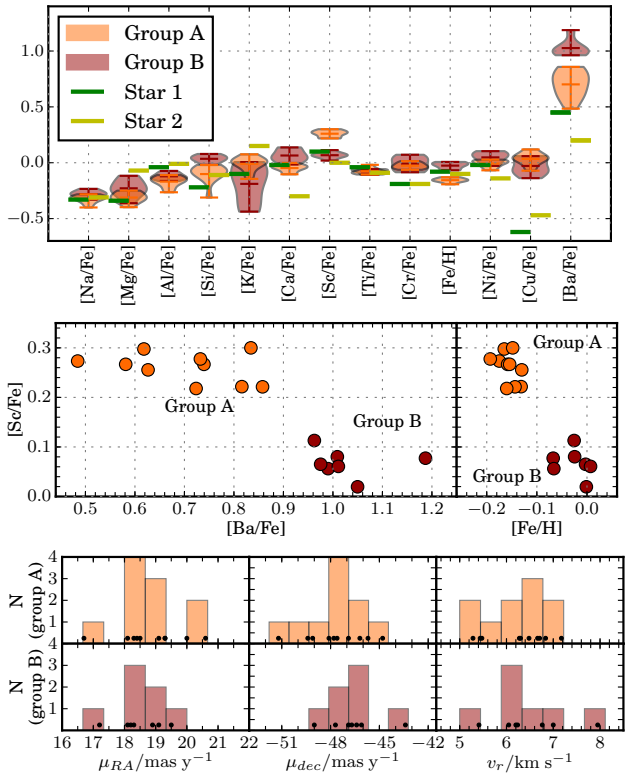

Figure 4. Top: The abundances for all 13 elements in groups marked A and B in Figure 2 and abundances for two new members. Middle: Groups A and B are separated in [Sc/Fe], [Ba/Fe], and [Fe/H] abundances. Abundances for each star are plotted. Bottom: Kinematics for each group.

One can notice that maps in Figure 3 show much more structure than we analysed in this paper. We explored other chemical groups and observed some regularities and patterns when kinematics and positions on the sky were inspected. There are, however, some polluting stars in these groups as well and decisive conclusions are hard to make. We leave the topic of pure blind chemical tagging of field stars for another paper. Blind chemical tagging will also be much easier on a set of 30 abundances and 300,000 stars soon to be produced by the GALAH collaboration. More observed elements mean much less pollution of chemical groups, so we might soon be able to find long-lost relationships between field stars for the first time with a good reliability. We also expect to find many more distinct clusters than we can see in the presented t-SNE maps (Bland-Hawthorn & Sharma 2016).

ACKNOWLEDGEMENTS

JK is supported by a Discovery Project grant from the Australian Research Council (DP150104667) awarded to J. Bland-Hawthorn and T. Bedding.

REFERENCES

Adams J. D., Stauffer J. R., Monet D. G., Skrutskie M. F., Beichman C. A., 2001, AJ, 121, 2053
Barden S. C., et al., 2010, in Ground-based and Airborne Instrumentation for Astronomy III. p. 773509, doi:10.1117/12.856103
Blanco-Cuaresma S., et al., 2015, A&A, 577, A47
Bland-Hawthorn J., Freeman K. C., 2004, Publ. Astron. Soc. Australia, 21, 110
Bland-Hawthorn J., Sharma S., 2016, Astronomische Nachrichten, 337, 894
Bland-Hawthorn J., Krumholz M. R., Freeman K., 2010a, ApJ, 713, 166
Bland-Hawthorn J., Karlsson T., Sharma S., Krumholz M., Silk J., 2010b, ApJ, 721, 582
Bovy J., 2016, ApJ, 817, 49
Brandt T. D., Huang C. X., 2015, ApJ, 807, 58
Bu Y., Chen F., Pan J., 2014, New Astron., 28, 35
Daniel S. F., Connolly A., Schneider J., Vanderplas J., Xiong L., 2011, AJ, 142, 203
De Silva G. M., Sneden C., Paulson D. B., Asplund M., Bland-Hawthorn J., Bessell M. S., Freeman K. C., 2006, AJ, 131, 455
De Silva G. M., Freeman K. C., Asplund M., Bland-Hawthorn J., Bessell M. S., Collet R., 2007, AJ, 133, 1161
De Silva G. M., et al., 2015, MNRAS, 449, 2604
Feng Y., Krumholz M. R., 2014, Nature, 513, 523
Freeman K., Bland-Hawthorn J., 2002, ARA&A, 40, 487
Gaia Collaboration Brown A. G. A., Vallenari A., Prusti T., de Bruijne J., Mignard F., Drimmel R., co-authors ., 2016, preprint, (arXiv:1609.04172)
Gebran M., Monier R., 2008, A&A, 483, 567
Geller A. M., Latham D. W., Mathieu R. D., 2015, AJ, 150, 97
Hogg D. W., et al., 2016, ApJ, 833, 262
Jeffries R. D., Thurston M. R., Hambly N. C., 2001, A&A, 375, 863
Jofre P., Das P., Bertranpetit J., Foley R., 2016, preprint, (arXiv:1611.02575)
Kos J., et al., 2017, MNRAS, 464, 1259
Kroupa P., Boily C. M., 2002, MNRAS, 336, 1188
Lochner M., McEwen J. D., Peiris H. V., Lahav O., Winter M. K., 2016, ApJS, 225, 31
Martell S. L., et al., 2016, ApJ, 825, 146
Martell S. L., et al., 2017, MNRAS, 465, 3203
Matijevič G., The RAVE Collaboration 2015, Proceedings IAU Symposium No. 317, 11, 336
Mermilliod J.-C., Bratschi P., Mayor M., 1997, A&A, 320, 74
Mitschang A. W., De Silva G. M., Zucker D. B., 2012, MNRAS, 422, 3527
Mitschang A. W., De Silva G., Zucker D. B., Anguiano B., Bensby T., Feltzing S., 2014, MNRAS, 438, 2753
Ness M., Hogg D. W., Rix H.-W., Ho A. Y. Q., Zasowski G., 2015, ApJ, 808, 16
Piskunov N., Valenti J. A., 2016, preprint, (arXiv:1606.06073)
Quillen A. C., Anguiano B., De Silva G., Freeman K., Zucker D. B., Minchev I., Bland-Hawthorn J., 2015, MNRAS, 450, 2354
Sestito P., Randich S., Bragaglia A., 2007, A&A, 465, 185
Slesnick C. L., Hillenbrand L. A., Massey P., 2002, ApJ, 576, 880
Stello D., et al., 2016, preprint, (arXiv:1611.09852)
Thygesen A. O., et al., 2014, A&A, 572, A108
Ting Y.-S., Freeman K. C., Kobayashi C., De Silva G. M., Bland-Hawthorn J., 2012, MNRAS, 421, 1231
Ting Y.-S., Conroy C., Goodman A., 2015, ApJ, 807, 104
Traven G., et al., 2016, preprint, (arXiv:1612.02242)
Tucholke H.-J., 1992, A&AS, 93, 293
Valenti J. A., Piskunov N., 1996, A&AS, 118, 595
Valentini M., et al., 2016, preprint, (arXiv:1609.03826)
Vanderplas J., Connolly A., 2009, AJ, 138, 1365
Zwitter T., et al., 2010, A&A, 522, A54

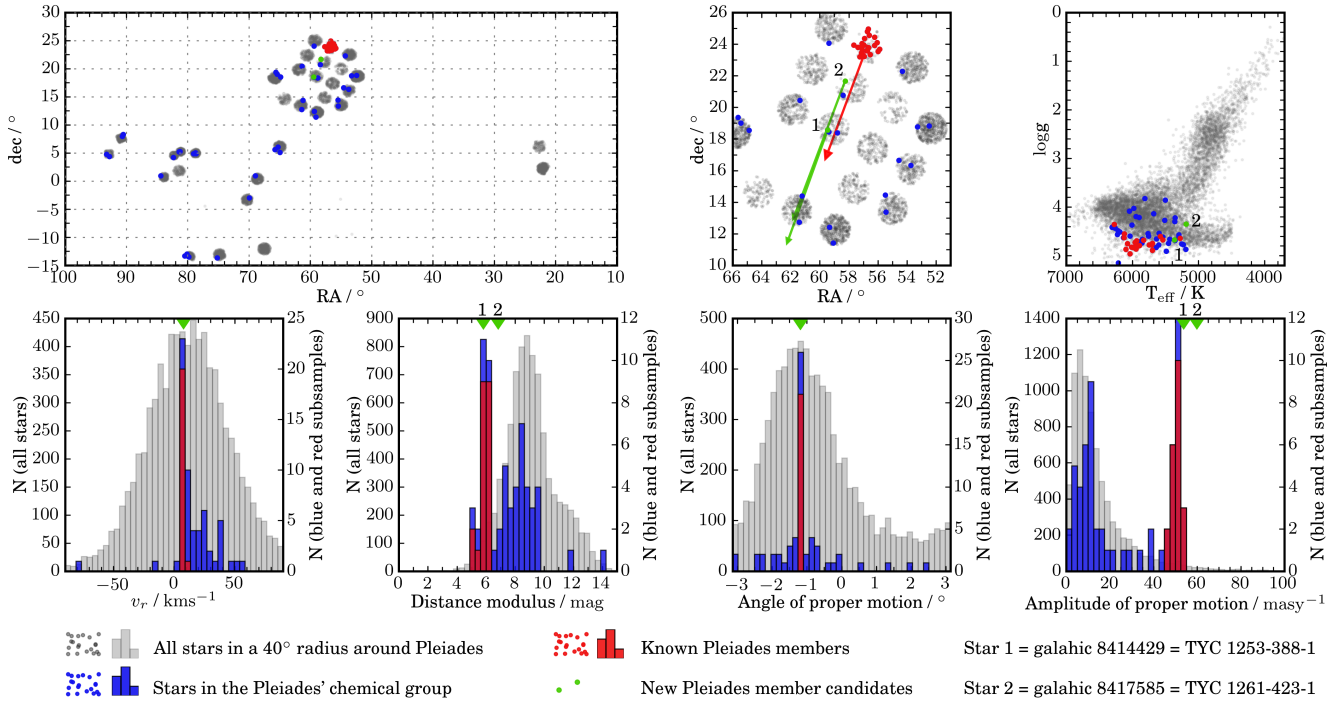

Figure 5. Position of the analysed stars on the sky (top-left and top-middle panels), on the HR diagram (top-right panel) and on the radial velocity, distance, and proper motion histograms (bottom row). In grey are plotted all the analysed stars in the 40° radius around Pleiades. In blue are the stars that belong into the Pleiades chemical group (inside the blue polygon in Figure 2). In red are known Pleiades members also marked with red symbols in Figure 2. Green are two new Pleiades stars that we discovered by chemical tagging.

de Silva G. M., Freeman K. C., Bland-Hawthorn J., Asplund M., Williams M., Holmberg J., 2011, MNRAS, 415, 563
van der Maaten L., 2013, preprint, (arXiv:1301.3342)
van der Maaten L., Hinton G., 2008, The Journal of Machine Learning Research, 9, 85

APPENDIX A: T-SNE PROJECTIONS OF THE REMAINING CLUSTERS

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