Astronomy & Astrophysics manuscript no. tsne_abunds
March 15, 2018

Dissecting stellar chemical abundance space with t-SNE

F. Anders[1,2], C. Chiappini[1,2], B. X. Santiago[3,2], G. Matijevič[1], A. B. Queiroz[3,2], M. Steinmetz[1]

[1] Leibniz-Institut für Astrophysik Potsdam (AIP), An der Sternwarte 16, 14482 Potsdam, Germany
e-mail: fanders@aip.de
[2] Laboratório Interinstitucional de e-Astronomia, - LIneA, Rua Gal. José Cristino 77, Rio de Janeiro, RJ - 20921-400, Brazil
[3] Instituto de Física, Universidade Federal do Rio Grande do Sul, Caixa Postal 15051, Porto Alegre, RS - 91501-970, Brazil

Received March 15, 2018; accepted ...

ABSTRACT

In the era of industrial Galactic astronomy and multi-object spectroscopic stellar surveys, the sample sizes and the number of available stellar chemical abundances have reached dimensions in which it has become difficult to process all the available information in an effective manner. In this paper we demonstrate the use of a dimensionality-reduction technique (t-distributed stochastic neighbour embedding; t-SNE) for analysing the stellar abundance-space distribution. While the non-parametric non-linear behaviour of this technique makes it difficult to estimate the significance of found abundance-space substructure, we show that our results depend little on parameter choices and are robust to abundance errors. By reanalysing the high-resolution high-signal-to-noise solar-neighbourhood HARPS-GTO sample with t-SNE, we find clearer chemical separations of the high- and low-[α/Fe] disc sequences, hints for multiple populations in the high-[α/Fe] population, and indications that the chemical evolution of the high-[α/Fe] metal-rich stars is connected with the super-metal-rich stars. We also identify a number of chemically peculiar stars, among them a high-confidence s-process-enhanced abundance pair (HD91345/HD126681) with very similar ages and v_X and v_Y velocities, which we suggest to have a common birth origin, possibly a merged dwarf galaxy. Our results demonstrate the potential of t-SNE and similar methods for chemical-tagging studies of large spectroscopic surveys.

Key words. Galaxy: general – Galaxy: abundances – Galaxy: disk – Galaxy: stellar content – Stars: abundances

1. Introduction

One of the major goals of modern Galactic astrophysics is to infer the formation history of our Milky Way. To achieve this aim it is necessary to obtain precise 6D stellar phase-space positions, detailed chemical abundance patterns, and reliable age estimates for large stellar samples. This chrono-chemo-kinematical map of the Galactic stellar populations can then be compared to predictions of various Milky-Way models, eventually unveiling the star-formation and dynamical history of our Galaxy.

Massive spectroscopic observing campaigns such as RAVE (Steinmetz et al. 2006), SEGUE (Yanny et al. 2009), the Gaia-ESO survey (Gilmore 2012), LAMOST (Deng et al. 2012), APOGEE (Majewski et al. 2017), and GALAH (Martell et al. 2017) have in the past decade increased both the volume coverage and the statistical sample sizes by more than two orders of magnitude, to $5\cdot10^6$ stars distributed from the solar vicinity to the far side of the Galactic bulge and the outer halo. In spite of this recent conquest of the Milky Way in terms of number of spectroscopically analysed stars, detailed multi-abundance chemo-kinematical studies of the immediate solar vicinity (Edvardsson et al. 1993; Fuhrmann 1998, 2011; Fuhrmann et al. 2017; Adibekyan et al. 2012; Bensby et al. 2014; Nissen 2015, 2016; Delgado Mena et al. 2017, e.g.) remain at least equally important for Galactic Archaeology (see Lindegren & Feltzing 2013 for a quantitative analysis). Also, before *Gaia* DR2, reliable stellar ages are still mostly confined to the solar vicinity (for exceptions using asteroseismology see Chiappini 2015; Martig et al. 2015; Casagrande et al. 2016; Anders et al. 2017b; Rodrigues et al. 2017; Miglio et al. 2017).

The wealth of new data, especially the high dimensionality of chemo-kinematics space, requires new statistical analysis methods to efficiently constrain detailed Milky-Way formation models (including e.g. stellar evolution, stellar chemical feedback, chemical evolution, and dynamical evolution). Traditionally, the metallicity distribution function and 2D chemical-abundance diagrams ([X/Fe] vs. [Fe/H]), and abundance gradients have been used to constrain the chemical evolution of stellar populations (e.g. Pagel 2009). On the other hand, it is also possible to *define* a stellar population by chemistry (e.g. carbon-enhanced metal-poor stars - Beers & Christlieb 2005; the chemical thick disc - Fuhrmann 1998; high-[α/Fe] metal-rich stars - Adibekyan et al. 2011), and to then study their structural and chemo-kinematic properties in detail. Abundance-space populations are usually defined in a simple fashion, by dissecting only one 2D abundance diagram.

More thorough multi-dimensional abundance-space studies using data-mining techniques have emerged over the past years. In a pioneer study, Ting et al. (2012) used principal-component analysis (PCA) to determine the effective dimensionality of abundance space accessible by spectroscopic surveys. da Silva et al. (2012, 2015), and Jofré et al. (2017) used tree clustering to find groups of stars with similar abundance patterns. Recently, Boesso & Rocha-Pinto (2018) studied a solar-vicinity literature compilation and combined hierarchical clustering and PCA to find peculiar chemical subgroups that do not follow the chemical-enrichment flow of the Galactic disc. Their results also suggest that 90% of the variance in the abundance data can be explained by two principal components that capture the main contributions to chemical enrichment. This is slightly at odds with the earlier work of Ting et al. (2012) who suggest that spectroscopic abundance space has at least an effective dimension of 4.

In this paper we explore the possibility of combining the information contained in various measured abundance ratios using the dimensionality reduction technique t-SNE (t-distributed stochastic neighbour embedding) to define more robust subpopulations and better identify outliers. In astronomical applications, t-SNE has mainly been used to identify objects with peculiar spectra (e.g. Matijevič et al. 2017; Valentini et al. 2017; Traven et al. 2017; Reis et al. 2018). During the writing of this paper, Kos et al. (2018) demonstrated in a complementary analysis that t-SNE can also be used as a chemical-tagging tool in chemical-abundance space: the authors were able to recover 7 out of 9 known open and globular clusters with high efficiency and low contamination using 13 chemical abundances from the GALAH survey (Martell et al. 2017), and they also found two new field member stars to known clusters with this technique.

Here we apply abundance-space t-SNE to the high-resolution solar-vicinity HARPS-GTO survey data of Delgado Mena et al. (2017), and demonstrate that this method provides a powerful visualisation and clustering tool also for field-star chemical-tagging studies. We identify, in a robust way, several distinct chemical-abundance substructures of the solar-vicinity disc population, as well as some peculiar stars. In a subsequent paper (Chiappini et al., in prep.) we discuss the main result: the detection of distinct chemical sub-populations in the high-[α/Fe] regime, a finding confirmed by other surveys that points to a different origin of the metal-poor and metal-rich part of the chemical thick disc.

The paper is structured as follows: Sec. 2 introduces t-SNE. Section 3 describes the t-SNE results for the high-resolution spectroscopic solar-vicinity survey of Delgado Mena et al. (2017), considering possible caveats in our analysis and characterising each of the found subpopulations. We finish with a discussion and conclusions in Sec. 4.

2. Dissecting chemistry space with t-SNE

Interpreting the multi-dimensional abundance distributions determined by spectroscopic surveys is not a trivial task, since different abundance diagrams contain different nucleosynthetic information and may be affected by different observational errors. A convenient way to simplify this problem is dimensionality reduction, i.e. the projection of the N-dimensional abundance space onto a lower-dimensional space in which the chemical similarity between two stars is reflected by their distance in that space. Possibly the best-known such method is PCA, widely used also in astronomical literature. For highly-correlated datasets such as spectral pixel spaces or chemical-abundance spaces, however, more sophisticated non-linear methods like IsoMap or locally linear embedding are known to perform much better (e.g. Matijevič et al. 2012; Ivezić et al. 2013).

In this paper, we reanalyse the high-resolution spectroscopic solar-vicinity survey of Delgado Mena et al. (2017) using a machine-learning algorithm called t-distributed stochastic neighbour embedding (t-SNE; Hinton & Roweis 2003; van der Maaten & Hinton 2008). This method is widely used in big-data analytics, and is able to efficiently project complex datasets onto a 2D plane in which the proximity between similar data points is preserved. We use the python implementation of t-SNE included in the `scikit-learn` package (Pedregosa et al. 2012) and refer to the original papers and the online documentation for details about the method and code. In short, the advantage of using t-SNE over other manifold-learning techniques is that it performs much better in revealing structure at many different scales (van der Maaten & Hinton 2008; Matijevič et al. 2017), which is a necessary feature when looking for chemical substructure in the Galactic disc.

How t-SNE works: For a given set of N high-dimensional datapoints $\mathbf{x}_1, \ldots, \mathbf{x}_N$ (images, spectra, or in our case chemical-abundance vectors), t-SNE first computes pairwise similarity probabilities p_{ij} for the points $\mathbf{x}_i$ and $\mathbf{x}_j$:

$$p_{j|i} = \frac{\exp(-||\mathbf{x}_i - \mathbf{x}_j||^2/2\sigma_i^2)}{\sum_{k\neq i}\exp(-||\mathbf{x}_i - \mathbf{x}_k||^2/2\sigma_i^2)}.$$

To circumvent problems with outliers, the symmetrised similarity of x_j and x_i is defined as

$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2N}.$$

In the next step, t-SNE attepts to learn a d-dimensional map $\mathbf{y}_1, \ldots, \mathbf{y}_N$ (in general $d = 2$) that reflects the similarities p_{ij} similarities between two points $\mathbf{y}_i$ and $\mathbf{y}_j$ in the low-dimensional map, defined as

$$q_{ij} = \frac{(1 + ||\mathbf{y}_i - \mathbf{y}_j||^2)^{-1}}{\sum_{k\neq m}(1 + ||\mathbf{y}_k - \mathbf{y}_m||^2)^{-1}}.$$

This metric uses Student's t distribution to avoid crowding problems in the low-dimensional map (van der Maaten & Hinton 2008). Starting from a random Gaussian distribution in the d-dimensional map, the locations of the points $\mathbf{y}_i$ are determined by minimizing the Kullback–Leibler divergence (Kullback & Leibler 1951) between the low- and high-dimensional similarity distributions Q and P:

$$KL(P||Q) = \sum_{i\neq j} p_{ij} \log \frac{p_{ij}}{q_{ij}},$$

using a gradient-descent method. The result of this optimization is a 2D (or 3D) map that reflects the similarities between the high-dimensional inputs (see e.g. Fig. 2).

The method has one main parameter, the so-called perplexity, p, which governs the bandwidth of the Gaussian kernels σ_i appearing in the similarities p_{ij}. As a result, the bandwidth is adapted to the density of the data: smaller values of σ_i are used in denser parts of the data space. The perplexity parameter can be thought of as a guess about the number of close neighbors each point has, and therefore the ideal value for p depends on the sample size. A change in perplexity has in many cases a complex effect on the resulting map, and different values for p should be explored (Wattenberg et al. 2016).

Recently, Linderman & Steinerberger (2017) demonstrated that two other hyper-parameters of t-SNE can be chosen optimally: the learning rate should be set to ~ 1, and the early-exaggeration parameter should be set to ~ 0.1 times the sample size. In the following, we use these recommendations.

In addition, t-SNE, as a genuine machine-learning technique, does have two drawbacks that are relevant for our science case. First, it does not account for individual uncertainties, and may therefore be affected by extremely heteroscedastic errors. We mitigate this shortcoming by performing a simple Monte-Carlo experiment (Sec. 3.2) to show that our results are robust to abundance uncertainties. Secondly, its current implementations do not allow to treat missing data, so that any star with a missing individual abundance measurement has to be excluded. We therefore decided to focus on the most inclusive set of chemical abundances (see Sec. 3).

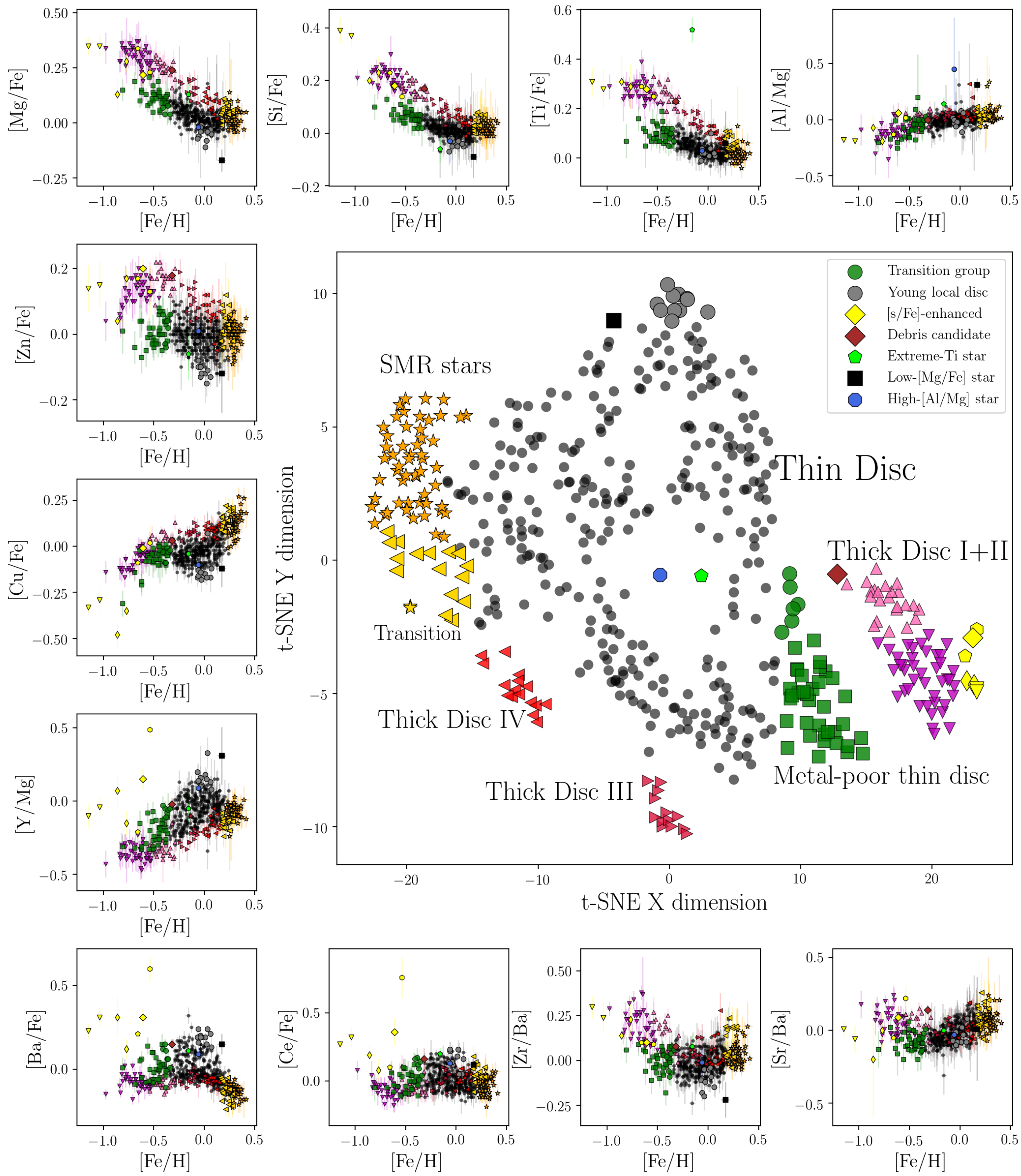

Fig. 1. Illustration of how t-SNE works in abundance space, using the Delgado Mena et al. (2017) sample. The small panels show eleven of the possible $\sim 20,000$ abundance diagrams that can be created from 13 elements. The resulting reference t-SNE projection of the full abundance space is shown in the big panel, and several identified subgroups are indicated.

3. Re-analysing the HARPS GTO sample

In an extensive series of papers, Adibekyan et al. (2011, 2012); Delgado Mena et al. (2014, 2015); Bertran de Lis et al. (2015); Suárez-Andrés et al. (2017); Delgado Mena et al. (2017, 2018) studied the chemical abundances of a sample of 1111 solar-vicinity FGK stars using the very high resolution of the HARPS spectrograph ($R \sim 115,000$). This sample mostly contains metal-rich warm dwarf and subgiant stars, but also includes a wide range of effective temperatures, gravities and metallicities. The HARPS sample initially served to detect and characterise exoplanets and may therefore contain some metallicity-related selection bias; however, e.g. Anders et al. (2014) have shown that the HARPS metallicity distribution (MDF) matches the MDF of high-quality local ($d < 1$ kpc) APOGEE DR10 red-giant stars that could be considered less chemically biased. The HARPS MDF also agrees with the MDF of the volume-complete sample of Fuhrmann (2011).

Delgado Mena et al. (2017) recently reanalysed this sample, employing a revised linelist (Tsantaki et al. 2013), improving the effective temperature calibration, and correcting spectroscopic gravities using the *Hipparcos* parallaxes of van Leeuwen (2007). They report chemical abundances for Mg, Al, Si, Ca, Ti,

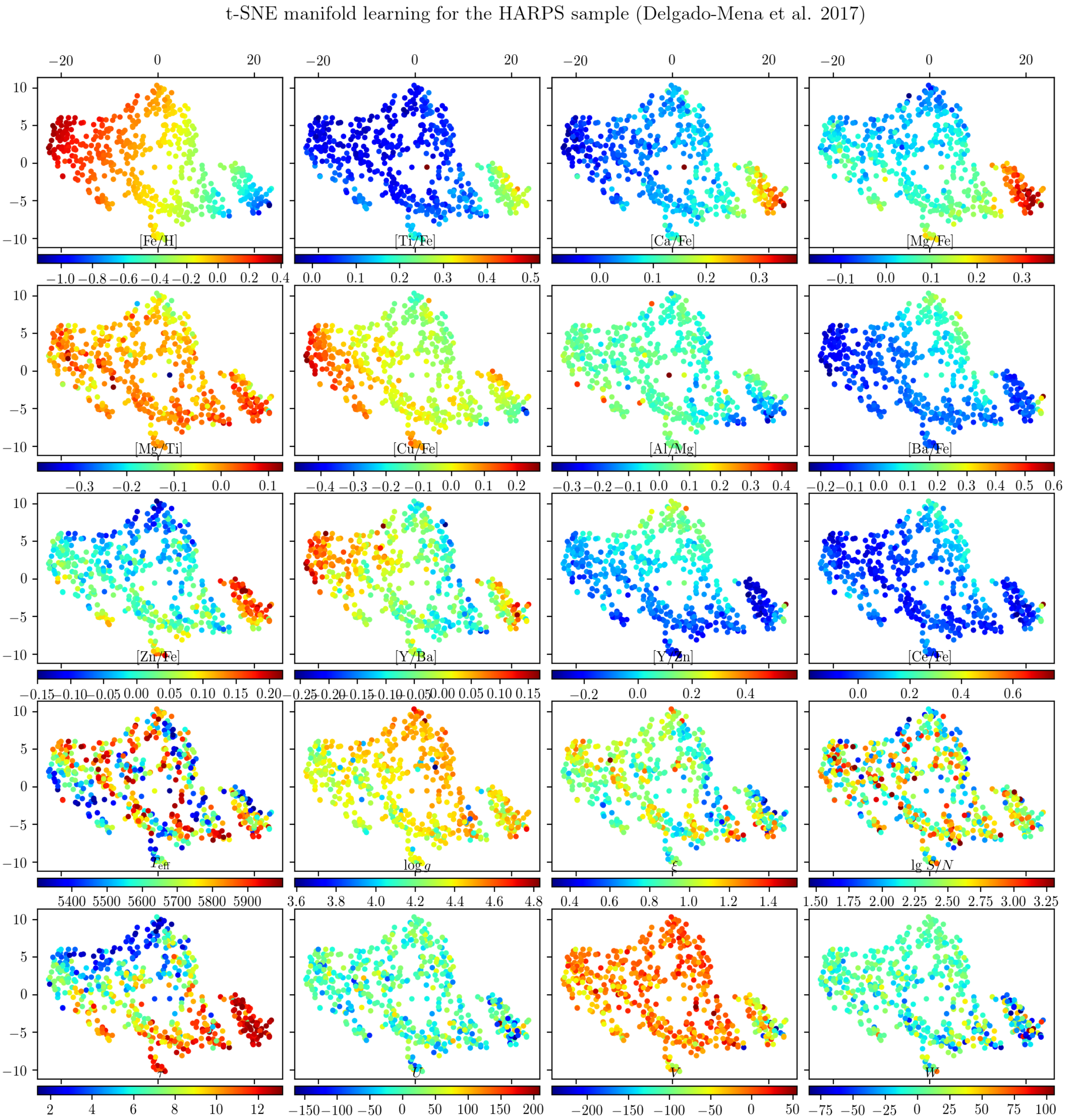

Fig. 2. Fiducial t-SNE projection of the Delgado Mena et al. (2017) sample (see big panel in Fig. 1), colour-coded by chemical abundances (top three rows), spectroscopic parameters (effective temperature $T_{\rm eff}$, surface gravity $\log g$, microturbulence ξ, and signal-to-noise ratio $\log SNR_{\rm HARPS}$; fourth row), age τ (fifth row, first panel) and space velocities (fifth row). We note that only [Fe/H] and the [X/Fe] ratios were used as input for the t-SNE run.

Fe, Cu, Zn, Sr, Y, Zr and Ba for 1059 stars (Ce, Nd and Eu are available for a substantial subset of these), derived using standard Local Thermodynamic Equilibrium (LTE) analysis using ARES (Sousa et al. 2007, 2015) to measure equivalent widths and MOOG (Sneden 1973) to measure abundances by comparing to Kurucz ATLAS9 atmospheres (Kurucz 1993). These chemical abundances were complemented by photometry from *Gaia* DR1 (Gaia Collaboration et al. 2016), APASS DR9 (Henden & Munari 2014), and 2MASS (Cutri et al. 2003), and by astrometry (parallaxes, proper motions) from the *Gaia* DR1/TGAS catalogue (Michalik et al. 2015; Gaia Collaboration et al. 2016), or when these were unavailable (135/1059 stars), from the re-reduced *Hipparcos* data (van Leeuwen 2007). Using the combined spectroscopic, photometric, and astrometric data, we computed precise stellar masses, ages, distances, and extinctions using the `StarHorse` code (Queiroz et al. 2018). For this run, we employed a fine grid ($\Delta \log \tau = 0.01$ dex, Δ[Z/H]=0.02 dex) of PARSEC 1.2S stellar models (Bressan et al. 2012; Tang et al. 2014; Chen et al. 2015), which significantly improved the precision of our ages with respect to the default grid ($\Delta \log \tau = 0.05$ dex, Δ[Z/H]=0.05 dex). The median age precision of the final t-SNE sample is 14%.

In this section we test the performance of abundance-space t-SNE on this most recent HARPS GTO sample compilation. The high number of measured abundances, in conjunction with the high precision of the measurements and the easily tractable sample size, makes the HARPS sample an ideal test case for machine-learning algorithms. Our first tests showed that, in order to obtain reliable t-SNE abundance maps, the sample needed to be analysed in a more restricted temperature range, because certain abundance trends seem to be dominated by underlying temperature trends. Therefore, similar to Delgado Mena et al. (2017), we chose an effective temperature range of 5300 K $< T_{\rm eff} <$ 6000 K (satisfied for 539 stars) for our analysis. We furthermore excluded one star with $\log g_{\rm HIP} < 3$, and required successful abundance determination for Mg, Al, Si, Ca, TiI, Fe, Cu, Zn, Sr, Y, ZrII, Ce and Ba that we use as input for t-SNE, leaving us with 533 stars.[1] To compensate the fact that t-SNE does not take into account individual (heteroscedastic) uncertainties in the data, we followed the approach of Hogg et al. (2016) and rescaled each abundance by the median uncertainty in that element, assuming an abundance uncertainty floor of 0.03 dex. In our final sample of 530 stars we also discarded 3 stars for which our age determination code, `StarHorse` (Santiago et al. 2016; Queiroz et al. 2018), did not converge. We verified that these choices do not significantly affect the resulting t-SNE maps.

The result of our reference t-SNE projection (perplexity p = 40) is illustrated in Fig. 1. The figure shows how the neighbourhood of points in several abundance diagrams (small panels) is reflected in the t-SNE map (big panel). We identified and named some of the substructures that emerge from the t-SNE projection.

Figure 2 again shows our reference t-SNE map for the HARPS sample, but now colour-coded by chemical-abundance ratios, stellar parameters, ages and kinematics. The panels in the first three rows show how t-SNE is grouping the stars with similar abundances in the two-dimensional plane. The panels coloured as a function of stellar parameters demonstrate that the sample is not subject to major systematic abundance shifts, but does show some residual trends with effective temperature, since it preferentially groups cooler stars in slightly different regions of the t-SNE map than hotter ones. Because part of this effect may be due to chemical evolution rather than systematic abundance errors, we refrained from applying ad-hoc corrections to the abundances.

Figure 3 shows the corresponding [X/Fe] abundance trends versus proton number for each of the substructures identified in Fig. 1. We now proceed to the discussion of these results.

3.1. The overall appearance of the t-SNE map

Our reference t-SNE projection shown in Figs. 1 and 2 reveals significant amounts of substructure in the local chemical-abundance space. The non-linearity of the method makes it difficult to attribute the overall appearance of the map to specific elemental abundances, which is why we limit this discussion to a qualitative level. In accordance with earlier studies of the dimensionality of abundance space (e.g. Ting et al. 2012; Boesso & Rocha-Pinto 2018), our results suggest that most of the variance of the data is in the metallicity and [α/Fe] abundance dimensions, corresponding to the different time-scales of supernovae type Ia and type II. In fact, the X dimension of the t-SNE map correlates very well with metallicity (Fig. 2, top right panel), which means that a) a lot of information about the chemical pattern of a star is already given by its metallicity, and b) the [Fe/H] abundance was measured with much higher precision, and therefore has more discriminative power than most of the other abundances.

Figure 2 also demonstrates that the t-SNE map's Y dimension, although it also correlates with [α/Fe] and [Zn/Fe] abundances, encodes information on s-process abundances, e.g. [Ba/Fe] and [Y/Zn], and consequently stellar age, a variable that was not included in the inference. In principle this opens up the possibility for calibrating multi-element chemical clocks.

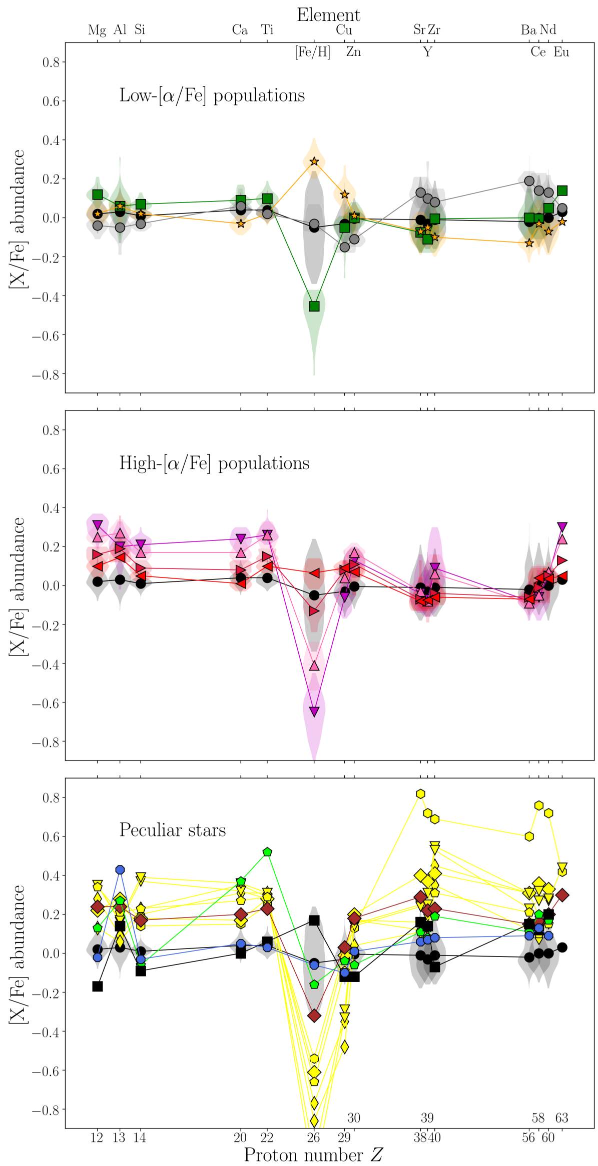

Fig. 3. Chemical-abundance patterns relative to iron for the t-SNE-selected subsamples of the HARPS survey, using the same symbols and colours as in Fig. 1. For each population we show the median abundance trend, as well as the full abundance distribution (for the top two panels). For visibility, we divided the sample into three groups that are shown separately in the three panels. The "thin disc" population (black circles) is shown in all panels for comparison.

[1] Carbon and oxygen abundances are available from previous studies (Suárez-Andrés et al. 2017; Bertran de Lis et al. 2015), but since they are based on previous stellar parameter estimates, we decided not to include them in the t-SNE runs and only use them in the interpretation. We also did not use Nd and Eu in the t-SNE run, because they were only available for about half of the sample (stars with the highest signal-to-noise ratios).

The fourth row of Fig. 2 also shows that the t-SNE projection responds to elemental-abundance trends with stellar parameters, although they have not been included as input parameters, and although we work in a narrow effective-temperature bin: t-SNE places stars with slightly different stellar parameters in slightly different places of the map. For example, we see some residual abundance trends with $T_{\rm eff}$ (fourth row, left panel), which may either be due to possible systematic abundance errors (see also Delgado Mena et al. 2017, or due to real stellar population trends (with stellar mass). In the case of $\log g$ (forth row, second panel), the trends are likely not due to systematic errors, but due to stellar and chemical evolution: at fixed $T_{\rm eff}$ on the main sequence, $\log g$ is a proxy for stellar age, and the abundance patterns are expected to vary with age.

By construction, t-SNE clusters similar-abundance stars in different places of the map. The several discernible islands on the map suggest that we are able to identify stars that were formed from gas with significantly different chemical enrichment than the bulk of the disc stars that live on the "main island" of the map. In the following subsection, we will show that most of the substructures identified in Fig. 1 are robust to abundance uncertainties and reasonable variations in our analysis.

3.2. The robustness of the t-SNE results

As discussed in Sec. 2, the overall appearance of the maps produced by t-SNE depend mainly on the perplexity parameter p, as well as on the chosen parameter space. In Fig. 4, we show the t-SNE maps for different perplexity values and different sets of input parameters, using the same colours and symbols as in Fig. 1. This experiment shows that:

1. The main features (i.e. neighbourhood relations between points) of the map are preserved (modulo map rotations/reflections) for a wide range of perplexities,.
2. The groups defined in Fig. 1 are also robustly recovered for different perplexities.
3. Using only [X/Fe] abundance ratios results in slightly different maps, which can be explained by the higher abundance precision of [Fe/H] with respect to the [X/Fe], and the thus higher weight of this dimension in the t-SNE projection. The [Fe/H] dimension alone, however, is not responsible for the emergence of the prominent subgroups.
4. Adding ages and/or kinematics to the input parameter space does not significantly improve the t-SNE projection, at least in this special case of very local, high-resolution, and high-signal-to-noise data. In the case of moving groups or globular clusters, however, adding kinematic dimensions to chemical tagging exercises does seem to help the recovery of known clusters (Chen et al. 2017).

We further tested the robustness of our reference map to abundance errors with a simple Monte-Carlo experiment (see Fig. 5: For each star, we created 50 mock stars with abundances drawn from a multi-dimensional Gaussian distribution centered on the measured abundance, and variance corresponding to the measured abundance uncertainties. Although t-SNE does not take into account uncertainties in the data, this procedure allows us to assure that the groups that we identified in the t-SNE map are not due to chance groupings.[2] The robustness test shown in

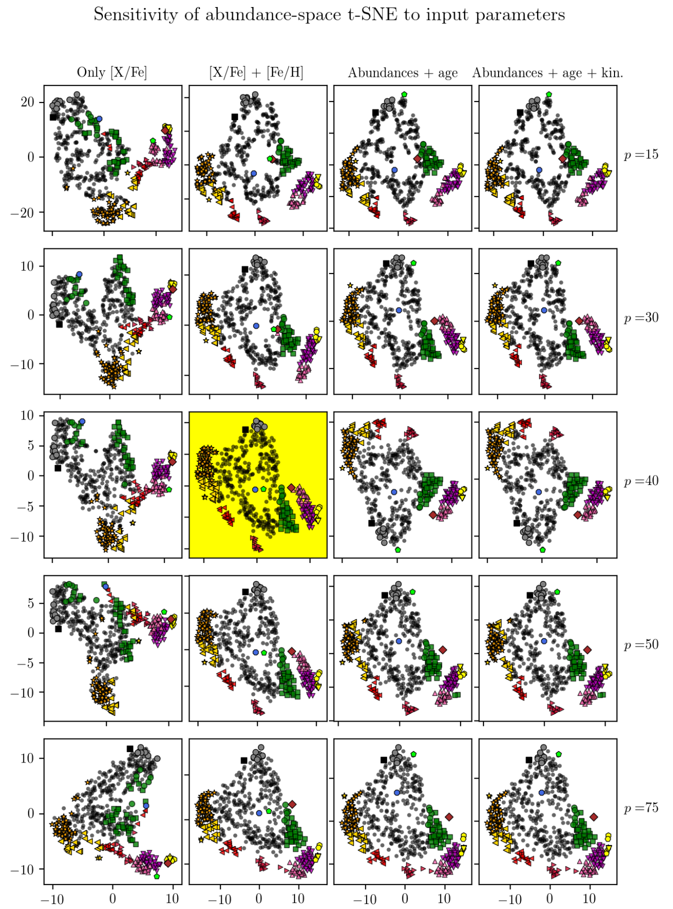

Fig. 4. t-SNE representations of the chrono-chemo-kinematics space spanned by the Delgado Mena et al. (2017) sample. Each column represents a combination of input information, while each row corresponds to a particular perplexity value, as indicated on the right side of the figure. The panel highlighted in yellow represents the results that we analyse in detail in this paper by defining chemical subpopulations based on this map.

Fig. 5 was also used for the definition of the chemical populations discussed in the next section.

3.3. Disc sub-populations

In this subsection, we discuss the main groups and features identified in Fig. 1 in more detail.

The thin-thick disc dichotomy: As discussed in the works of Adibekyan et al. (2011, 2012) and Delgado Mena et al. (2017), the HARPS-GTO data confirm the clear discontinuity between the high- and the low-[α/Fe] sequences in the [Mg/Fe] vs. [Fe/H] diagram (e.g. Edvardsson et al. 1993; Fuhrmann 1998, 2011; Fuhrmann et al. 2017). This discontinuity is reflected in a very clear manner in the t-SNE projection: We find a clear and obvious gap between the chemical thin- and thick-disc populations in the t-SNE diagram that remains very robust for different choices of the t-SNE hyper-parameters. Primarily, this means that the chemical patterns of thin and thick disc are indeed distinct, and can be disentangled by high-resolution spectroscopy. Secondly, our analysis of the full chemical information results in a much more accurate division of the chemically-thin and thick popula-

[2] In general, adding uncertainties to measured (i.e. already noisy) data will blur the true values even more. This means that if a signal disappears in the Monte-Carlo test, the test does not rule out its existence. On the other hand, if the signal persists, it is very unlikely to be due to a chance grouping.

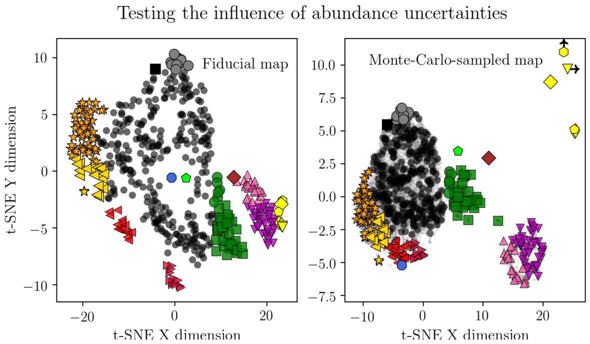

Fig. 5. Robustness test of our t-SNE-selected subsamples to abundance errors. The right panel shows the fiducial map, while the left panel shows the result of our Monte-Carlo test. For each star, 50 random stars were drawn from a Gaussian centered on the measured abundance, and with dispersions corresponding to the measured uncertainties. t-SNE was then run on this artificially increased sample. The resulting map demonstrates that our selected subgroups are robust to measurement errors.

tions. Indeed, if one only relies on one diagnostic, such as the [Mg/Fe] vs. [Fe/H] diagram (Adibekyan et al. 2011; Delgado Mena et al. 2017), several thin-disc stars would (most probably incorrectly) be identified as belonging to the chemical thick disc (see Fig. 1).

Thick-disc sub-populations: Adibekyan et al. (2011) first discovered a clear discontinuity between the metal-poor and metal-rich [α/Fe]-enhanced (or hαmr) disc populations (see also the earlier indications, e.g. Fuhrmann 2008, Fig. 30). In our t-SNE analysis of the Delgado Mena et al. (2017) sample, similar to the original paper, we also see a clear difference between at least two, maybe three or four populations (dubbed Thick Discs I/II and III/IV in Fig. 1). Even if ages and/or kinematics are included as additional dimensions in the analysis, this picture does not change much. The implications of this result, which we can also confirm with other high-resolution data covering larger volumes, will be discussed in depth in a companion paper (Chiappini et al., in prep.).

The middle panel of Fig. 3 shows the abundance profile with respect to iron for each of the four [α/Fe]-rich populations, compared to the chemical thin disc. The figure suggests that the most of the abundance variance among the four groups can be captured by one parameter (e.g. metallicity). There are, however, subtle deviations from this pattern: for example, group II is more enhanced in [Al/Fe] than group I. Figure 6 shows that the population age decreases monotonically from group I to IV.

Super-metal-rich stars: SMR stars ([Fe/H] $\gtrsim 0.3$; see Grenon 1972, 1989, 1999; Chiappini 2009 – the western-most stars in the t-SNE plane; orange stars in Fig. 1) have only slightly different abundance patterns from the bulk of the thin disc stars (black dots; see Fig. 3); however, Fig. 2 shows that they are enhanced in [Y/Ba] and [Cu/Fe] with respect to the local thin disc, indicative of an origin in the inner Milky-Way disc. Figures 6 and 7 show that most of these stars have ages between 4 and 8 Gyr (Trevisan et al. 2011; Casagrande et al. 2011; Anders et al. 2017b), and are on cold orbits ($e < 0.12$; e.g. Kordopatis et al. 2015), which again supports the idea that they have radially migrated from the inner disc (see e.g. Minchev et al. 2012, 2013, 2014; Vera-Ciro et al. 2014; Grand & Kawata 2016).

The transition from hαmr to SMR stars: Most literature measurements agree that the high- and low-[α/Fe] sequences in the [α/Fe] vs. [Fe/H] diagram merge at super-solar metallicities (e.g. Adibekyan et al. 2011; Anders et al. 2014; Hayden et al.

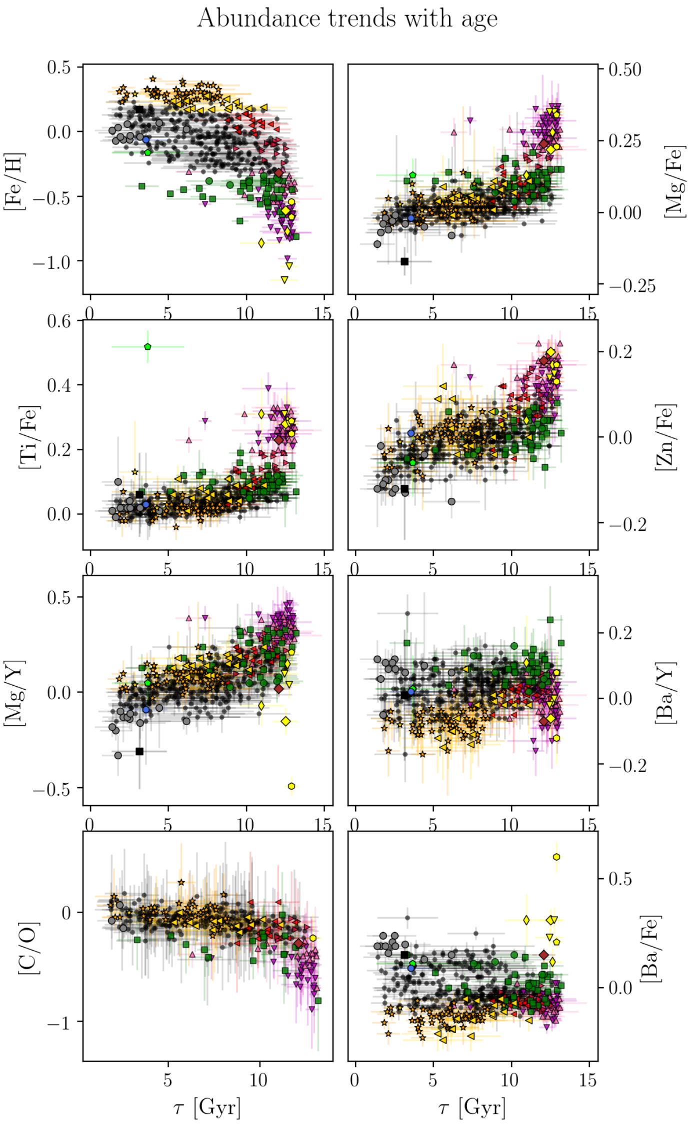

Fig. 6. Abundance trends of the HARPS-GTO abundances with stellar age, measured with the `StarHorse` code (Queiroz et al. 2018).

2015). In other words, the upper metallicity limit of the high-[α/Fe]/hαmr population is not yet firmly established. Our analysis shows that including the full chemistry information does not allow us to completely solve this question, since the border between thin-disc-like and thick-disc-like chemistry remains debatable in the t-SNE projection (e.g. the dark-yellow triangles in Fig. 1 have intermediate characteristics between Thick Disc IV and SMR stars). This could mean that the hαmr groups dubbed Thick Disc III/IV in Fig. 1 do not belong to the genuine thick disc, but actually are old inner-disc migrators (see Chiappini et al., in prep.).

The metal-poor thin disc: The green squares and circles in Fig. 1 correspond to the metal-poor thin disc ([Fe/H] ~ -0.5). Apart from metallicity, its main abundance differences with respect to the bulk of the chemical thin-disc population are: 1. a light elevation in all [α/Fe] ratios, as a consequence of the later onset of star formation in the outer disc, where this population is most likely to originate from (e.g. Nordström et al. 2004; Anders et al. 2014; Hayden et al. 2015; see also kinematic diagnostics in Fig. 7), 2. a slight underabundance of [Sr/Fe] and [Y/Fe] with respect to the thin-disc population, but solar-like second s-process peak abundances, and 3. a hint for a systematic r-process ([Eu/Fe]~ 0.2) enhancement (consistent with that in [α/Fe] with respect to the local disc.

The young locally-born disc: The grey circles in Fig. 1 denote a population that we call young local disc, because 1. they are among the youngest stars ($\sim$ 1 Gyr; see Fig. 6), 2. they follow the local rotation curve with a very low velocity dispersion (see Fig. 7, top right panel)The young local disc stars have near-solar metallicities and are slightly deficient in [Mg/Fe], [Al/Fe], and [Si/Fe] with respect to the Sun and the rest of the thin disc, as expected for young stars from the stronger contribution of SN Ia yields. They are also systematically deficient in [Cu/Fe] and [Zn/Fe], while being moderately in enhanced ([s/Fe]~ 0.15) in s-process elements (see Fig. 3. The second-peak s-process elements (Ba, Ce, Nd) are slightly more enhanced than first peak (Sr, Y, Zr), which is different from the rest of the thin disc, and slightly against the trends set by with metallicity-dependent yields from intermediate- and low-mass AGB stars (e.g. Cristallo et al. 2009, 2015; da Silva et al. 2016; Delgado Mena et al. 2017).

The remaining thin-disc component: The black dots in Fig. 1 stand for the remaining parts of the low-[α/Fe] solar-vicinity disc (7 kpc $\lesssim R_{\rm guide} \lesssim$ 9 kpc). The morphology of this population in the t-SNE map confirms that this reference "thin disc" is not be a homogeneous monolithic population either (possibly more substructure could be defined, although less robustly; see Fig. 5). But within the scope of this paper, we define the thin disc as a broad component that has a wide range of ages and birth places, and therefore could in principle also cover a wider range of chemical abundances. Fig. 3 shows, however, that our "thin disc" population, while covering a considerable metallicity range from -0.3 to 0.24, has quite small spreads in each elemental abundance relative to iron, and closely follows the solar abundance pattern. This suggests that the chemical evolution of the interstellar medium in the disc near must have been slow and very homogeneous for the past $\sim$ 10 Gyr. The significant spread in the age-metallicity relation of the solar-neighbourhood thin disc (Fig. 6) can be explained by the presence of a strong radial metallicity gradient, together with radial mixing (e.g. Haywood 2006; Minchev et al. 2013; Anders et al. 2017a, Minchev et al., in prep.).

3.4. Chemically peculiar stars

In addition to the main disc populations discussed in the previous subsection, Fig. 1 also highlights a number of outliers and chemically peculiar stars revealed by the t-SNE projection. Some of them are known peculiar objects, some are solid, and some are dubious candidates. Their abundance patterns relative to iron are shown in the bottom panel of Fig. 3. Here we discuss each of them briefly.

s-process-enhanced stars: Our method clearly singles out a small group of stars with dwarf-galaxy- or globular-cluster-like, and s-process-enhanced abundance patterns (a few more were lost due to the temperature and abundance quality cuts). These seven stars (yellow points in Fig. 1) are all enhanced in the measured s-process elements with respect to both the thin and thick disc populations (see Fig. 3, bottom panel). They are also enhanced in [α/Fe], although there is considerable star-to-star variance, and all of them are [Al/Mg]-poorer than the thick-disc populations, placing them in an abundance regime somewhere in-between Galactic thick disc, the halo, and massive dwarf-galaxies.

The most extreme abundance outlier in this group, as already noted by Delgado Mena et al. (2017), is HD11397 (yellow hexagon), which shows the highest s-process abundances of the entire sample ([s/Fe]~ 0.7). It was classified as a so-called mild barium star by Pompéia & Allen (2008) who also showed that its s-process abundance pattern is compatible with typical AGB stellar yields, possibly accreted from an unseen companion. Another star that was noted as a mildly s-enhanced thick-disc star by Delgado Mena et al. (2017), is HD126803 (yellow square). The last mild s-enhancement candidate of Delgado Mena et al. (2017), CD-436810, did not satisfy our $T_{\rm eff}$ criterion, and was therefore not included in our analysis.

s-process-enhanced stars with halo kinematics: HD175179 (yellow pentagon) as well as BD+083095 and CD-4512460 (yellow diamonds) are mildly [s/Fe]-enhanced old halo-kinematic stars with very similar abundance patterns and metallicities between -0.66 and -0.86. BD+083095 and CD-4512460 are shown with the same symbol in all figures because they also have similar kinematics.

A high-confidence s-process-enhanced abundance pair: The yellow triangles in all figures correspond to the nearby high proper-motion stars HD91345 and HD126681.[3] We find that the HARPS-derived abundances of Delgado Mena et al. (2017) are so similar for these two stars that they can be considered abundance-ratio twins (see Table 1). With the exception of metallicity (2σ-deviation), all [X/Fe] abundances are consistent with each other within the respective 1σ uncertainties. They have the two highest [Si/Fe] enrichments of the sample. In connection with their similar ages and space velocities (except for the discrepant v_Z component), and considering the rareness of [s/Fe]-enhanced metal-poor disc stars, we propose that the two stars could have been born in the same stellar system (possibly a globular cluster) that has long since been disrupted by the Milky Way.

Another debris candidate at higher metallicity: HD28701 (brown diamond in Fig. 1) is another interesting object with similar s-process enhancements as the [s/Fe]-enhanced stars discussed above, but at higher metallicity ([Fe/H]$= -0.32$). Like the group of yellow stars, it shows mildly enhanced ([s/Fe]~ 0.2) abundances of Sr, Y and Zr when compared to thick-disc stars of similar metallicity, and not as much enhancement in the second s-process peak elements Ba, Ce and Nd. It is also enhanced in the r-process element europium ([Eu/Fe]$= 0.30 \pm 0.06$).

Extreme-[Ti/Fe] star: HD14452 (limegreen pentagon, SNR$_{\rm HARPS}$ = 89) has possibly the most extraordinary abundance pattern of the Delgado Mena et al. (2017) sample: It has a metallicity of [Fe/H]$= -0.16 \pm 0.02$ and is highly enriched in the heavier α-elements titanium and calcium ([Ti/Fe]$= 0.52 \pm 0.05$, [Ca/Fe]$= 0.37 \pm 0.10$), while being only slightly enhanced in [Mg/Fe], and not at all in [Si/Fe]. Also the elevated [Al/Mg] ratio is puzzling. This abundance pattern suggests peculiar enrichment, e.g. by a $\sim 25\ M_\odot$ type-II supernova that did not produce light α elements but large amounts of Ca and Ti (e.g. Ritter et al. 2017, Fig. 26).

Low-[Mg/Fe] candidate: HD113513 (black square, SNR$_{\rm HARPS}$ = 26) is not really an outlier in the t-SNE map, but the star with the lowest [α/Fe] ratios of the sample, and

[3] Both stars were observed by various solar-vicinity spectroscopic surveys such as RAVE (Steinmetz et al. 2006; Kunder et al. 2017) and GCS (Nordström et al. 2004; Casagrande et al. 2011), resulting in compatible spectroscopic parameter determinations (although of lower quality). Using high-resolution spectroscopy, Bensby et al. (2014) measured a very similar abundance profile to the HARPS one analysed here for HD126681.

Table 1. Details of the abundance pair HD91345 and HD126681.

Property	HD91345	HD126681
$\mathrm{SNR}_{\mathrm{HARPS}}$	160	244
T_{eff}	5658 ± 39 K	5570 ± 34 K
$\log g_{\mathrm{HIP}}$	4.38 ± 0.08	4.64 ± 0.08
[Fe/H]	-1.04 ± 0.03	-1.15 ± 0.03
[Mg/Fe]	0.35 ± 0.05	0.35 ± 0.04
[Al/Fe]	0.16 ± 0.01	0.17 ± 0.01
[Si/Fe]	0.37 ± 0.04	0.39 ± 0.05
[Ca/Fe]	0.32 ± 0.09	0.36 ± 0.06
[TiI/Fe]	0.28 ± 0.08	0.31 ± 0.06
[Cu/Fe]	0.29 ± 0.05	0.33 ± 0.03
[Zn/Fe]	0.15 ± 0.07	0.14 ± 0.07
[Sr/Fe]	0.25 ± 0.07	0.24 ± 0.05
[Y/Fe]	0.31 ± 0.05	0.25 ± 0.07
[ZrII/Fe]	0.55 ± 0.06	0.53 ± 0.05
[Ba/Fe]	0.31 ± 0.06	0.23 ± 0.04
[Ce/Fe]	0.32 ± 0.08	0.27 ± 0.03
[Nd/Fe]	0.27 ± 0.06	0.28 ± 0.09
[Eu/Fe]	0.44 ± 0.32	
Mass	$0.762^{+0.004}_{-0.011}\ M_{\odot}$	$0.711^{+0.008}_{-0.006}\ M_{\odot}$
Age	$12.9^{+0.2}_{-0.8}$ Gyr	$12.7^{+0.4}_{-1.0}$ Gyr
Distance	59 ± 1 pc	54.3 ± 0.6 pc
v_X	11.5 ± 0.7 km/s	10.3 ± 0.1 km/s
v_Y	185.2 ± 0.3 km/s	182.5 ± 0.7 km/s
v_Z	-7.5 ± 0.1 km/s	-70.2 ± 0.5 km/s

it also sticks out in several of the abundance diagrams shown in Fig. 1. Its abundance profile is similar to the "young local disc" population defined above, except for its higher metallicity and the elevated [Al/Mg] ratio. We note the low signal-to-noise ratio of the spectrum, and the consequently higher abundance uncertainties.

High-[Al/Mg] candidate: HD29428 (blue octogon) is most likely not a true oddball, but a typical youngish thin-disc star with a very uncertain Al measurement ([Al/Fe]= 0.43 ± 0.46) that ended up as an outlier in the t-SNE map because the method cannot account for heteroscedastic errors.

3.5. Trends with age and kinematics

Figure 6 presents the dependence of several stellar abundance ratios on stellar age, as derived by the `StarHorse` code (Queiroz et al. 2018), for each of the abundance populations and groups defined above. Haywood et al. (2013) computed ages for the HARPS sample of Adibekyan et al. (2012), and studied the [α/Fe] and [Fe/H] abundance trends of the chemically-defineds thin, metal-poor thin, and thick discs. More recently, Delgado Mena et al. (2018) also computed ages for the re-analysed HARPS-GTO sample of Delgado Mena et al. (2017), using the PARAM tool (da Silva et al. 2006). Fig. 6 presents similar age-abundance trends with respect to those authors, but focussing on our t-SNE-defined stellar populations. An in-depth discussion of these results is beyond the scope of this paper; the age-metallicity and age-[α/Fe] relations will be discussed in more detail in Chiappini et al. (in prep.) and Minchev et al. (in prep.).

For similar reasons of completeness, we show some kinematic relations and age-kinematic trends in Fig. 7. The kinematic results are based on *Gaia* DR1/TGAS positions and proper motions, radial velocities from Adibekyan et al. (2012), our `StarHorse` distances, and the orbit-integration tool `galpy` (Bovy 2015), using the new Staeckel approximation implementation of Mackereth & Bovy (2018) to determine orbital eccentricities e and maximum heights above the plane Z_{max}. Again, we discuss the implications of these results in Chiappini et al. (in prep.).

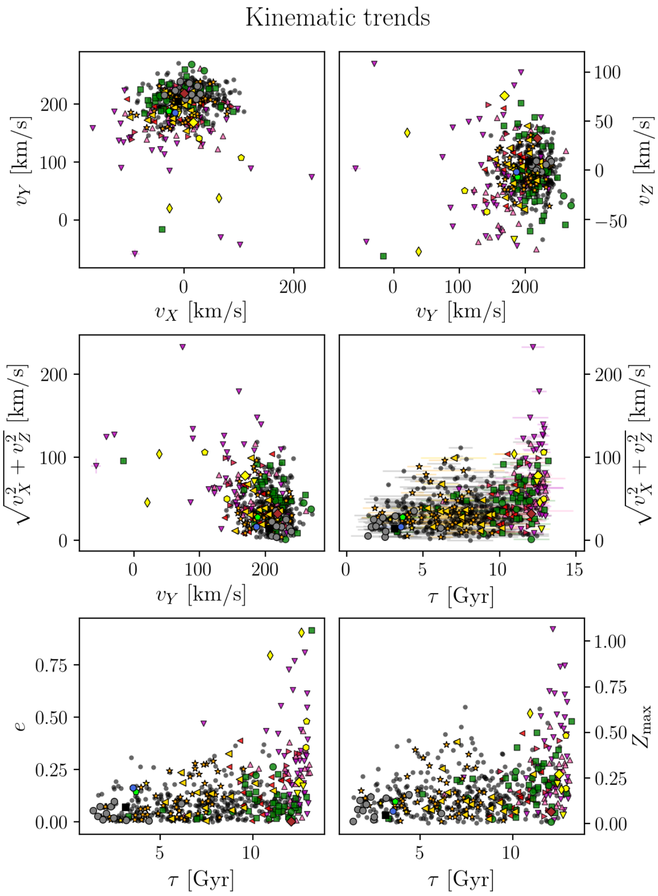

Fig. 7. Kinematic trends of the HARPS-GTO sample. The upper panels show the distribution of the sample in the $v_X - v_Y$ (or UV) and $v_Y - v_Z$ (VW) planes. We note that the 1σ errorbars are in most cases smaller than the symbols. The middle left panel shows the classic Toomre diagram (e.g. Feltzing et al. 2003), and the middle right panel a diagram of orbital heating (as measured by v_X and $v - Z$) as a function of age. The bottom panels display the orbital eccentricities and maximum heights above the Galactic plane, respectively, as a function of age.

4. Discussion and conclusions

The solar vicinity comprises a well-established mixture of stellar populations, among them halo stars, thick- and thin-disc stars, stars in streams, stars passing by on eccentric orbits, stars on circular orbits that have radially migrated, chemically peculiar stars, and even stars with possibly extragalactic origins (e.g. members of disrupted dwarf galaxies or globular clusters). In this paper we have demonstrated the use of the dimensionality reduction algorithm t-SNE to better define subpopulations in abundance space. While the non-parametric non-linear behaviour of the technique makes it difficult to estimate the significance of found subgroups or clusters, we have verified that our results depend little on the t-SNE parameter choices and are robust to abundance errors. As in other differential abundance studies, it is important to confine the analysis to narrow regions in atmospheric-parameter space to avoid spurious abundance trends induced by differences in atmospheric parameters. The t-SNE method could in principle even be coupled to a genuine cluster finding algorithm.

Our approach allowed us to define chemical subpopulations in the solar vicinity in a more reliable way than by just looking at 2D abundance diagrams. The gap between the chemical thin and thick discs is much more prominent, as is the separation between the genuine thick-disc and the high-α metal-rich population. The high-[α/Fe] population may even be composed of more than two distinct populations, but this affirmation is not as robust to abundance uncertainties. The metal-rich end of the high-α metal-rich population and the super-metal-rich thin-disc stars are not clearly separated in our t-SNE map, which suggests that their chemical evolution is connected (both have origins in the inner disc, Adibekyan et al. 2012; Haywood et al. 2018, see also Chiappini et al., in prep.).

We also re-characterise the chemical thin-disc component, even excluding the metal-poor and and super-metal-rich parts corresponding to stars originating from the inner and outer disc, respectively. This broad component still covers a considerable metallicity range (~ -0.3 to $+0.25$) and a wide range of ages ($\sim$ 10 Gyr), but has surprisingly similar abundances as the Sun, and very small spreads in the α abundances ($\sigma[\alpha_i/\mathrm{Fe}]\simeq 0.03$ – smaller than the individual abundance uncertainties except for [Mg/Fe]), while [s/Fe] abundances dispersions are higher ($\sim$ 0.08 dex) because s-process elements are sensitive to age and birth radius). In accordance with previous literature, we attribute these facts to a slow and homogeneous chemical evolution in the disc that is mainly characterised by a negative radial metallicity gradient, as well as strong radial migration that brought stars from various Galactic radii into the solar vicinity.

Our identification of the s-process-rich abundance pair HD91345/HD126681 in Sec. 3.4 further demonstrates the potential of abundance-space t-SNE for chemical tagging. The viability of t-SNE for strong chemical tagging (finding dispersed members of open clusters) is still not completely clear, though. The GALAH results of Kos et al. (2018) suggest that it is possible to recover a large fraction of open clusters with abundance-space t-SNE, and to even find extratidal cluster members with this technique. On the other hand, the recent APOGEE paper by Ness et al. (2018) adopts a more pessimistic view on strong chemical tagging: the authors find that most 0.03-dex level abundance pairs (at solar metallicity) were probably not born in the same cluster, but are rather "doppelgänger-abundance" stars than actual twins. Our abundance pair is arguably a rarer case than a solar-abundance pair, and may well be a real stellar twin, as also suggested by the very similar (and precise) ages, as well as v_X and v_Y space velocities. The question why the v_Z velocities are so different remains to be resolved, as well as the question if we can find more stars with similar abundance patterns.

References

Adibekyan, V. Z., Santos, N. C., Sousa, S. G., & Israelian, G. 2011, A&A, 535, L11
Adibekyan, V. Z., Sousa, S. G., Santos, N. C., et al. 2012, A&A, 545, A32
Anders, F., Chiappini, C., Minchev, I., et al. 2017a, A&A, 600, A70
Anders, F., Chiappini, C., Rodrigues, T. S., et al. 2017b, A&A, 597, A30
Anders, F., Chiappini, C., Santiago, B. X., et al. 2014, A&A, 564, A115
Beers, T. C. & Christlieb, N. 2005, ARA&A, 43, 531
Bensby, T., Feltzing, S., & Oey, M. S. 2014, A&A, 562, A71
Bertran de Lis, S., Delgado Mena, E., Adibekyan, V. Z., Santos, N. C., & Sousa, S. G. 2015, A&A, 576, A89
Boesso, R. & Rocha-Pinto, H. J. 2018, MNRAS, 474, 4010
Bovy, J. 2015, ApJS, 216, 29
Bressan, A., Marigo, P., Girardi, L., et al. 2012, MNRAS, 427, 127
Casagrande, L., Schönrich, R., Asplund, M., et al. 2011, A&A, 530, A138
Casagrande, L., Silva Aguirre, V., Schlesinger, K. J., et al. 2016, MNRAS, 455, 987
Chen, B., D'Onghia, E., Pardy, S., et al. 2017, ApJ, submitted, arXiv:1709.03987
Chen, Y., Bressan, A., Girardi, L., et al. 2015, MNRAS, 452, 1068
Chiappini, C. 2009, in IAU Symposium, Vol. 254, IAU Symposium, ed. J. Andersen, B. Nordström, & J. Bland-Hawthorn, 191–196
Chiappini, C. 2015, in EAS Publications Series, Vol. 73, EAS Publications Series, 309–341
Cristallo, S., Straniero, O., Gallino, R., et al. 2009, ApJ, 696, 797
Cristallo, S., Straniero, O., Piersanti, L., & Gobrecht, D. 2015, ApJS, 219, 40
Cutri, R. M., Skrutskie, M. F., van Dyk, S., et al. 2003, 2MASS All Sky Catalog of point sources.
da Silva, L., Girardi, L., Pasquini, L., et al. 2006, A&A, 458, 609
da Silva, R., Lemasle, B., Bono, G., et al. 2016, A&A, 586, A125
da Silva, R., Milone, A. d. C., & Rocha-Pinto, H. J. 2015, A&A, 580, A24
da Silva, R., Porto de Mello, G. F., Milone, A. C., et al. 2012, A&A, 542, A84
Delgado Mena, E., Bertrán de Lis, S., Adibekyan, V. Z., et al. 2015, A&A, 576, A69
Delgado Mena, E., Israelian, G., González Hernández, J. I., et al. 2014, A&A, 562, A92
Delgado Mena, E., Tsantaki, M., Adibekyan, V. Z., et al. 2017, A&A, 606, A94
Delgado Mena, E., Tsantaki, M., Adibekyan, V. Z., et al. 2018, IAU Symposium 330 Proceedings, arXiv:1707.05156
Deng, L.-C., Newberg, H. J., Liu, C., et al. 2012, Research in Astronomy and Astrophysics, 12, 735
Edvardsson, B., Andersen, J., Gustafsson, B., et al. 1993, A&A, 275, 101
Feltzing, S., Bensby, T., & Lundström, I. 2003, A&A, 397, L1
Fuhrmann, K. 1998, A&A, 338, 161
Fuhrmann, K. 2008, MNRAS, 384, 173
Fuhrmann, K. 2011, MNRAS, 414, 2893
Fuhrmann, K., Chini, R., Kaderhandt, L., & Chen, Z. 2017, MNRAS, 464, 2610
Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2016, A&A, 595, A2
Gilmore, G. 2012, in Astronomical Society of the Pacific Conference Series, Vol. 458, Galactic Archaeology: Near-Field Cosmology and the Formation of the Milky Way, ed. W. Aoki, M. Ishigaki, T. Suda, T. Tsujimoto, & N. Arimoto, 147
Grand, R. J. J. & Kawata, D. 2016, Astronomische Nachrichten, 337, 957
Grenon, M. 1972, in IAU Colloq. 17: Age des Etoiles, ed. G. Cayrel de Strobel & A. M. Delplace, 55
Grenon, M. 1989, Ap&SS, 156, 29
Grenon, M. 1999, Ap&SS, 265, 331
Hayden, M. R., Bovy, J., Holtzman, J. A., et al. 2015, ApJ, 808, 132
Haywood, M. 2006, MNRAS, 371, 1760
Haywood, M., Di Matteo, P., Lehnert, M., et al. 2018, A&A, submitted, arXiv:1802.09887
Haywood, M., Di Matteo, P., Lehnert, M. D., Katz, D., & Gómez, A. 2013, A&A, 560, A109
Henden, A. & Munari, U. 2014, Contributions of the Astronomical Observatory Skalnate Pleso, 43, 518
Hinton, G. E. & Roweis, S. T. 2003, in Advances in neural information processing systems, 857–864
Hogg, D. W., Casey, A. R., Ness, M., et al. 2016, ApJ, 833, 262
Ivezić, Ž., Connolly, A., VanderPlas, J., & Gray, A. 2013, Statistics, Data Mining, and Machine Learning in Astronomy
Jofré, P., Das, P., Bertranpetit, J., & Foley, R. 2017, MNRAS, 467, 1140
Kordopatis, G., Binney, J., Gilmore, G., et al. 2015, MNRAS, 447, 3526
Kos, J., Bland-Hawthorn, J., Freeman, K., et al. 2018, MNRAS, 473, 4612
Kullback, S. & Leibler, R. A. 1951, Ann. Math. Statist., 22, 79
Kunder, A., Kordopatis, G., Steinmetz, M., et al. 2017, AJ, 153, 75
Kurucz, R. 1993, ATLAS9 Stellar Atmosphere Programs and 2 km/s grid. Kurucz CD-ROM No. 13. Cambridge, Mass.: Smithsonian Astrophysical Observatory, 1993., 13
Lindegren, L. & Feltzing, S. 2013, A&A, 553, A94
Linderman, G. C. & Steinerberger, S. 2017
Mackereth, J. T. & Bovy, J. 2018, MNRAS, submitted, arXiv:1802.02592
Majewski, S. R., Schiavon, R. P., Frinchaboy, P. M., et al. 2017, AJ, 154, 94
Martell, S. L., Sharma, S., Buder, S., et al. 2017, MNRAS, 465, 3203
Martig, M., Rix, H.-W., Aguirre, V. S., et al. 2015, MNRAS, 451, 2230
Matijevič, G., Chiappini, C., Grebel, E. K., et al. 2017, A&A, 603, A19
Matijevič, G., Zwitter, T., Bienaymé, O., et al. 2012, ApJS, 200, 14
Michalik, D., Lindegren, L., & Hobbs, D. 2015, A&A, 574, A115
Miglio, A., Chiappini, C., Mosser, B., et al. 2017, Astronomische Nachrichten, 338, 644
Minchev, I., Chiappini, C., & Martig, M. 2013, A&A, 558, A9
Minchev, I., Chiappini, C., & Martig, M. 2014, A&A, 572, A92
Minchev, I., Famaey, B., Quillen, A. C., et al. 2012, A&A, 548, A127
Ness, M., Rix, H.-W., Hogg, D. W., et al. 2018, ApJ, 853, 198
Nissen, P. E. 2015, A&A, 579, A52
Nissen, P. E. 2016, A&A, 593, A65

Nordström, B., Mayor, M., Andersen, J., et al. 2004, A&A, 418, 989
Pagel, B. E. J. 2009, Nucleosynthesis and Chemical Evolution of Galaxies
Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2012, ArXiv e-prints, arXiv:1201.0490
Pompéia, L. & Allen, D. M. 2008, A&A, 488, 723
Queiroz, A. B. A., Anders, F., Santiago, B. X., et al. 2018, MNRAS, in press, arXiv:1710.09970
Reis, I., Poznanski, D., Baron, D., Zasowski, G., & Shahaf, S. 2018, MNRAS, in press, arXiv:1711.00022
Ritter, C., Herwig, F., Jones, S., et al. 2017, MNRAS, submitted, arXiv:1709.08677
Rodrigues, T. S., Bossini, D., Miglio, A., et al. 2017, MNRAS, 467, 1433
Santiago, B. X., Brauer, D. E., Anders, F., et al. 2016, A&A, 585, A42
Sneden, C. A. 1973, PhD thesis, University of Texas at Austin
Sousa, S. G., Santos, N. C., Adibekyan, V., Delgado-Mena, E., & Israelian, G. 2015, A&A, 577, A67
Sousa, S. G., Santos, N. C., Israelian, G., Mayor, M., & Monteiro, M. J. P. F. G. 2007, A&A, 469, 783
Steinmetz, M., Zwitter, T., Siebert, A., et al. 2006, AJ, 132, 1645
Suárez-Andrés, L., Israelian, G., González Hernández, J. I., et al. 2017, A&A, 599, A96
Tang, J., Bressan, A., Rosenfield, P., et al. 2014, MNRAS, 445, 4287
Ting, Y.-S., Freeman, K. C., Kobayashi, C., De Silva, G. M., & Bland-Hawthorn, J. 2012, MNRAS, 421, 1231
Traven, G., Matijevič, G., Zwitter, T., et al. 2017, ApJS, 228, 24
Trevisan, M., Barbuy, B., Eriksson, K., et al. 2011, A&A, 535, A42
Tsantaki, M., Sousa, S. G., Adibekyan, V. Z., et al. 2013, A&A, 555, A150
Valentini, M., Chiappini, C., Davies, G. R., et al. 2017, A&A, 600, A66
van der Maaten, L. & Hinton, G. 2008, The Journal of Machine Learning Research, 9, 85
van Leeuwen, F., ed. 2007, Astrophysics and Space Science Library, Vol. 350, Hipparcos, the New Reduction of the Raw Data
Vera-Ciro, C., D'Onghia, E., Navarro, J., & Abadi, M. 2014, ApJ, 794, 173
Wattenberg, M., Viégas, F., & Johnson, I. 2016, Distill, doi:10.23915/distill.00002
Yanny, B., Rockosi, C., Newberg, H. J., et al. 2009, AJ, 137, 4377

Acknowledgements. FA would like to thank Elisa Delgado-Mena for sharing the re-reduced HARPS-GTO data prior to publication, and Guillaume Guiglion, Katia Cunha, Paula Jofré, Bertrand Lemasle and the other participants of the IAU symposium 334 in Potsdam, as well as David W. Hogg, for their encouragement and critical thoughts.