Principal Component Analysis on Chemical Abundances Spaces

Y. S. Ting,[1,2*] K. C. Freeman,[1] C. Kobayashi,[1,3] G. M. De Silva[4] and
J. Bland-Hawthorn[5]

[1] *Research School of Astronomy & Astrophysics, The Australian National University, Cotter Road, Weston Creek, ACT 2611, Australia*
[2] *Department of Physics, National University of Singapore, Singapore 119260*
[3] *Centre for Astrophysics Research, University of Hertfordshire, Hatfield AL10 9AB, United Kingdom*
[4] *Australian Astronomical Observatory, PO Box 296, NSW 1710, Australia*
[5] *Institute of Astronomy, School of Physics, University of Sydney, NSW2006, Australia*

20 March 2012

ABSTRACT

In preparation for the *HERMES* chemical tagging survey of about a million Galactic FGK stars, we estimate the number of independent dimensions of the space defined by the stellar chemical element abundances [X/Fe]. This leads to a way to study the origin of elements from observed chemical abundances using Principal Component Analysis. We explore abundances in several environments, including solar neighbourhood thin/thick disk stars, halo metal-poor stars, globular clusters, open clusters, the Large Magellanic Cloud and the Fornax dwarf spheroidal galaxy. By studying solar-neighbourhood stars, we confirm the universality of the *r*-process that tends to produce [*n*-capture elements/Fe] in a constant ratio. We find that, especially at low metallicity, the production of *r*-process elements is likely to be associated with the production of α-elements. This may support the core-collapse supernovae as the *r*-process site. We also verify the over-abundances of light *s*-process elements at low metallicity, and find that the relative contribution decreases at higher metallicity, which suggests that this lighter elements primary process may be associated with massive stars. We also verify the contribution from the *s*-process in low-mass AGB stars at high metallicity. Our analysis reveals two types of core-collapse supernovae: one produces mainly α-elements, the other produces both α-elements and Fe-peak elements with a large enhancement of heavy Fe-peak elements which may be the contribution from hypernovae. Excluding light elements that may be subject to internal mixing, K and Cu, we find that the [X/Fe] chemical abundance space in the solar neighbourhood has about 6 independent dimensions both at low metallicity ($-3.5 \lesssim$ [Fe/H] $\lesssim -2$) and high metallicity ([Fe/H] $\gtrsim -1$). However the dimensions come from very different origins in these two cases. The extra contribution from low mass AGB stars at high metallicity compensates the dimension loss due to the homogenization of the core-collapse supernovae ejecta. Including the extra dimensions from [Fe/H], K, Cu and the light elements, the number of independent dimensions of the [X/Fe]+[Fe/H] chemical space in the solar neighbourhood for *HERMES* is about 8 to 9. Comparing fainter galaxies and the solar neighbourhood, we find that the chemical space for fainter galaxies such as Fornax and the Large Magellanic Cloud has a higher dimensionality. This is consistent with the slower star formation history of fainter galaxies. We find that open clusters have more chemical space dimensions than the nearby metal-rich field stars. This suggests that a survey of stars in a larger Galactic volume than the solar neighbourhood may show about 1 more dimension in its chemical abundance space.

Key words: methods: data analysis – ISM: abundances – ISM: evolution – stars: abundances – stars: AGB and post-AGB – stars: supernovae: general.

1 INTRODUCTION

Stars are believed to form in aggregates which are mostly short-lived. They disrupt through the action of mass

---

* E-mail: ting.yuansen@gmail.com

loss from stellar evolution, two-body effects and the tidal field of the Galaxy (e.g. Chernoff & Weinberg 1990; Odenkirchen et al. 2003). After the aggregates dissolve, their debris disperses and after several Galactic rotation periods becomes mixed throughout an annular region around the Galaxy. In this way, the stellar disk was gradually built up. The goal of chemical tagging is to use element abundances to reconstruct these ancient clusters in which the stars were born (Freeman & Bland-Hawthorn 2002). Individual clusters are observed to be chemically homogeneous, at least in the elements heavier than Na (e.g. De Silva et al. 2006; Mikolaitis et al. 2010, 2011). Stars that were born in the same cluster and have now dispersed will have similar element abundance patterns reflecting the chemical evolution of the gas from which they formed. This gas had its own history of pollution by ejecta from core-collapse supernovae, type Ia supernovae (SNe Ia) and Asymptotic Giant Branch (AGB) stars.

Chemical space ($\mathcal{C}$-space) is a space defined by the abundances of the chemical elements that we are able to measure. Finding the debris from long-dispersed clusters using chemical tagging is an exercise in group finding in this multidimensional $\mathcal{C}$-space. For stars which formed in clusters within the Milky Way (MW), chemical tagging appears relatively straightforward. The stars were born in chemically homogeneous clusters and the detailed abundance pattern changes from cluster to cluster. The debris from a single star cluster will be tightly clustered in chemical space. Stars can also come into the Galaxy in small accreted galaxies; their stars are again disrupted by the Galactic tidal field and dispersed. If these small galaxies are like the dwarf spheroidal galaxies (dSph) and ultra-faint galaxies now around the Milky Way, they would not have been be chemically homogeneous (Tolstoy et al. 2009). Their broad and different element abundance patterns are defined by their individual star formation histories (SFH) and are unlike those of the Galactic disk. The debris of such accreted galaxies would not lie in a tight clump but in a streak through $\mathcal{C}$-space which reflects the element abundance pattern in the parent dwarf.

To make a significant recovery of dispersed aggregates, simulations show that a large sample of stars, of order one million, is needed (Bland-Hawthorn & Freeman 2004). The GALAH[1] survey using the *HERMES*[2] instrument on the Anglo Australian Telescope is designed to do such a chemical tagging study. It will measure abundances of about 25 elements using multi-object high resolution (R = 28,000) spectroscopy of about a million stars. Its $\mathcal{C}$-space will have about 25 dimensions but the abundances of these 25 elements do not all vary independently. The abundances of some elements are correlated because of the underlying nucleosynthetic processes, and the effective dimensionality of the $\mathcal{C}$-space will be less than 25. Based on the existence of the various element groups (light, light odd-Z, α, Fe-peak, light and heavy s-process and mostly r-process) represented in the *HERMES* spectra, and from various abundance patterns observed in field and cluster stars, we think that the dimensionality of the $\mathcal{C}$-space will be about 7 to 9. The higher the dimensionality of $\mathcal{C}$-space, the more power chemical tag-

ging will have to identify the debris of disrupted systems. We would like to determine the dimensionality of $\mathcal{C}$-space more rigorously and that is one of the main purposes of this paper.

It seems likely that the dimensionality of $\mathcal{C}$-space will depend on the metallicity[3] of the stars defining the $\mathcal{C}$-space. For example, stars of low metallicity may have formed within a brief period of time and the gas from which they formed may have been enriched by only a few supernovae events (Audouze & Silk 1995). Among the neutron-capture (n-capture) elements, there would not have been time for AGB stars to provide enrichment of s-process elements, so the r-process may dominate the chemical evolution. On the other hand, the more metal-rich stars may have had a long history of chemical evolution with more than one process contributing to the n-capture element abundances.

We will use Principal Component Analysis (PCA) to determine the dimensionality of $\mathcal{C}$-space for various samples of stars, including field stars in different metallicity intervals, star clusters and stars in the Fornax dSph galaxy and the Large Magellanic Cloud (LMC). This is because the components will be different in the different situations. We will attempt to interpret the principal components in terms related to nucleosynthesis processes and hopefully gain some insight into the dominant processes in each situation.

This paper is organized as follows: Section 2 gives a brief discussion of some related aspects of chemical evolution processes and sites. In Section 3, we give details on the samples that we chose to study. In Section 4, we will describe the PCA method and a way to estimate intrinsic correlation using Monte-Carlo simulations. In Section 5, we will present the results of PCA analysis and then in Section 6, we will interpret the results and discuss their origins. We summarize our main conclusions in Section 7.

2 CHEMICAL EVOLUTION PROCESSES

Stellar element abundance studies are usually presented in terms of [X/Fe]–[Fe/H][4] diagrams. These illustrate many basic element behaviour patterns, such as decreasing trend of [α/Fe] toward higher [Fe/H] which is usually attributed to the increasing contribution of SNe Ia to the chemical evolution at later times. These two-dimensional plots do not readily reveal the interplay between different families of elements during the different stages of chemical evolution. This interplay is of particular interest to us in trying to determine the dimensionality of $\mathcal{C}$-space. As chemical evolution proceeds, different groups of elements evolve together and define individual dimensions of the $\mathcal{C}$-space. We seek an approach that is able to display these groups of elements, using PCA that we will present here. This will hopefully benefit the analysis of elemental abundances that will be obtained soon for large samples of stars with high resolution multi-object surveys such as *APOGEE*[5] (Allende Prieto et al. 2008), *HER-*

---

[3] We assume that the iron abundance [Fe/H] is a pertinent tracer of the metallicity in this paper.
[4] By definition, $[X/Y] \equiv \log_{10}(N_X/N_Y)_{\text{star}} - \log_{10}(N_X/N_Y)_{\odot}$, where N_X and N_Y are the abundances of element X and element Y respectively.
[5] The Apache Point Observatory Galactic Evolution Experiment

MES and the *ESO-Gaia VLT* survey. First we briefly discuss major production sites for different families of elements.

α-elements such as O, Ne, Mg, Si, S, and Ca are mainly produced by core-collapse supernovae (i.e. Type II, Ib and Ic Supernovae), while Fe-peak elements such as Cr, Mn, Ni, and Fe are mainly produced by SNe Ia. Although the progenitor models are still debated, the onset delay for these SNe is in a range of 0.1 – 2 Gyr after star formation begins (see model from Kobayashi & Nomoto 2009) . Observations show complications such as over-abundances of heavy Fe-peak elements such as Zn and Co at low metallicity (McWilliam et al. 1995; Ryan, Norris & Beers 1996; Norris, Ryan & Beers 2001; Cayrel et al. 2004; Chen, Nissen & Zhao 2004). These cannot be understood from core-collapse supernovae models with explosion energy of about 10^{51} erg. Umeda & Nomoto (2002, 2005) showed that more energetic supernovae (hypernovae) are responsible for the over-abundances of Zn and Co. This is supported by Chieffi & Limongi (2002). Heger & Woosley (2010) argued that over-abundances of Zn can be explained without involving hypernovae. However, Izutani & Umeda (2010) showed that it is necessary to include hypernovae in order to explain both Co and Zn observations.

Neutron-capture elements ($A \gtrsim 65$) can provide independent probes for Galactic chemical evolution. There is not yet a consistent scenario to explain all observations of n-capture elements. Physically, the n-capture processes are divided into slow ($s-$) and rapid ($r-$) processes. In the r-process the neutron flux is intense and the time-scale of n-capture is much shorter than that of the β-decay. This causes the seed nucleus to grab many neutrons before it β-decays to the valley of stability. The observed abundance patterns of r-process elements show very small star-to-star variation, and are in excellent agreement with the scaled solar r-process curve, at least for elements $56 < Z < 72$ (e.g. Sneden et al. 1996; Westin et al. 2000; Hill et al. 2002; Cowan et al. 2002, 2005). This suggests the universality of the r-process nucleosynthesis. However, the site(s) of the r-process remain uncertain. The requirements on the physical conditions are neutron rich (i.e. low electron fraction Y_e), high entropy and short dynamic time-scales (Wanajo & Ishimaru 2006); but see caveat from Freiburghaus et al. (1999a). A possible site is the neutrino-driven neutron-wind during the formation of neutron stars in core-collapse supernovae (e.g. Woosley et al. 1994). Various scenarios of this possibility have been studied, including association with low-mass supernovae (e.g. Wanajo et al. 2009) or massive supernovae (Truran et al. 2002; Wanajo & Ishimaru 2006), such as the ν-driven He-shell mechanism (Banerjee, Haxton & Qian 2011). Another alternative site is neutron star mergers (Freiburghaus, Rosswog & Thielemann 1999b; Rosswog et al. 1999). It has been argued that the stringent mass range might be responsible for the large scatter that has been observed for n-capture elements (Wanajo & Ishimaru 2006). The homogenizing effect of n-capture element evolution is less apparent than for the α-elements that are produced by the whole mass range of core-collapse supernovae.

In contrast, for the s-process, the neutron flux is not so intense and the n-capture rate is compara-

ble to the β-decay rate. The major sites for the s-process are believed to be low-mass (1.5 – 3 $M_\odot$) AGB stars (for a review, see Busso, Gallino & Wasserburg 1999; Herwig 2005; Käppeler et al. 2011) with the input of the ^{13}C pocket (Iben & Renzini 1982; Hollowell & Iben 1988). However, the progenitor mass range, metallicity dependency and the impact of rotation (Langer et al. 1999; Meynet & Maeder 2002) are still uncertain. In the s-process, the major proposed neutron sources are from ^{13}C$(\alpha, n)^{16}$O (e.g. Käppeler et al. 1990; Stancliffe & Jeffery 2007; Cristallo et al. 2009). More massive stars (> 4 $M_\odot$) achieve higher temperatures and another neutron source ^{22}Ne$(\alpha, n)^{25}$Mg could be dominant (Iben 1975; Truran & Iben 1977). Stars in the mass range of 8 – 12 $M_\odot$ might evolve into super AGB phase (Siess 2006, 2007, 2010). However, the s-process in super AGB phase is still yet to be understood. S-process also operates in massive stars that do not go through AGB/super AGB phase. This *weak s-process* can produce elements significantly up to Sr (e.g. Prantzos, Hashimoto & Nomoto 1990; Pignatari et al. 2010). Due to the long time-scales of low-mass AGB stars, it is believed that AGB stars will not contribute significantly below metallicity [Fe/H] $= -2$ (e.g. Roederer et al. 2010; Kobayashi et al. 2011c). However, the contribution of AGB stars can appear even at low metallicity in the form of some classes of carbon enhanced metal-poor (CEMP) stars through binary mass transfer (McClure 1984; Johnson & Bolte 2002b, 2004; Lucatello et al. 2003; Sivarani et al. 2004; Goswami et al. 2006; Aoki et al. 2007, 2008; Masseron et al. 2010). The AGB star ejecta is transfered to the companion star and causes them to be enriched in carbon and s-process elements (see models from Stancliffe & Glebbeek 2008; Stancliffe 2009; Bisterzo et al. 2009). In addition to n-capture element production from the s-process, low-mass stars also produce a small amount of Mg and O in the dredge-up (Marigo 2001; Karakas & Lattanzio 2003), and might pollute the interstellar medium (ISM) with lighter elements due to the mass loss from the outer envelope (e.g. Reimers 1977; Vassiliadis & Wood 1993). However, these contributions to the ISM are marginal, especially at high metallicity (e.g. Kobayashi et al. 2006; Karakas 2010).

For the elements at $A \simeq 90$ including Sr, Y and Zr, the origin is even more problematic (e.g. Honda et al. 2004b; Aoki et al. 2005; Cowan et al. 2005). The observed abundance ratios of n-capture elements suggest the overproduction of these elements. Travaglio et al. (2004) showed that this overproduction cannot be explained by the weak s-process alone and named it as the Lighter Element Primary Process (LEPP). It has been shown that these elements could be produced by collapse of rotating massive stars, i.e. collapsars (Pruet, Thompson & Hoffman 2004) or the weak r-process with slightly low Y_e matter that is naturally expected in core-collapse supernovae (Izutani et al. 2009). Qian & Wasserburg (2007, 2008) pointed out that these elements might be formed through charged particle reaction (CPR). More recently, Boyd et al. (2011) suggested that this anomaly could be explained by truncated r-process for massive stars, in which the produced heavier r-process elements are consumed by the collapse of neutron star to a black hole.

Light elements (Li, C, N, O, Na) will be depleted or internally-mixed during their Red Giant Branch

(RGB)/AGB phase either from the CNO cycle, the NeNa cycle (e.g. Andrievsky et al. 2007) or the hot bottom burning process (e.g. Karakas & Lattanzio 2003). For instance, Li in giants will be depleted (Norris et al. 1997; Spite et al. 2005; Bonifacio et al. 2007; Sbordone et al. 2010), whereas N will be enhanced at the expense of C and O. To avoid complication due to internal mixing and to better study the evolutionary state of the ISM when the stars were formed, we will leave the discussion of Li, C, N, O, and Na to a later paper.

We will also study the $\mathcal{C}$-space properties for some dwarf galaxies, because dwarf galaxies are often proposed as building blocks for larger galaxies like the Milky Way (for a general review on local group dwarf galaxies, see Mateo 1998). The chemical properties of the dwarf galaxies that are orbiting the Milky Way, such as Sculptor, Fornax, Sagittarius, Sextans and the LMC[6] (Shetrone, Côté & Sargent 2001; Shetrone et al. 2003; Geisler et al. 2005; Sbordone et al. 2007) reveal that the present-day dwarfs are chemically too dissimilar to the Milky Way to be realistic building blocks (but see Frebel, Kirby & Simon 2010b). However the ultra-faint dSph galaxies such as Coma Berenices, Hercules, Leo IV, Leo T, Ursa Major I and II have abundance patterns more like the pattern of the Galactic metal-poor halo and appear to be plausible building blocks for the halo (Kirby et al. 2008; Koch et al. 2008; Frebel et al. 2010a), although their baryonic masses are very small.

We briefly summarize the difference in abundance patterns between the brighter dwarf galaxies and the Milky Way; for a more complete discussion, see Venn et al. (2004); Venn & Hill (2008); Tolstoy, Hill & Tosi (2009). Some dwarf galaxies show lower $[\alpha/\text{Fe}]$ at $-2 \lesssim [\text{Fe/H}] \lesssim -1$ than the metal-poor stars in the solar neighbourhood, but reach the same metal-poor plateau ($[\alpha/\text{Fe}] = 0.3 – 0.5$) at $[\text{Fe/H}] \lesssim -2$. This is often explained by the slower star formation histories (SFH) of the dwarf galaxies, so SNe Ia and AGB stars that have longer time-scales can contribute more to the chemical evolution at lower metallicity than they do for the nearby metal-poor stars. This produces a 'knee' in the $[\alpha/\text{Fe}]$–$[\text{Fe/H}]$ diagram but at lower $[\text{Fe/H}]$. However, Kobayashi & Nomoto (2009) argued that if this is due to the SNe Ia contribution, $[\text{Mn/Fe}]$ should also show an increasing trend at the same $[\text{Fe/H}]$, while $[\text{Mn/Fe}]$ ratios in dwarf galaxies are as low as in the Galactic halo stars (McWilliam, Rich & Smecker-Hane 2003). They claimed that the low $[\alpha/\text{Fe}]$ and low $[\text{Mn/Fe}]$ abundances patterns in the dwarfs are more consistent with the lack of massive star contribution. Or perhaps both the slower SFH and the lack of massive stars play a role in explaining the chemical profile of the dwarf galaxies.

The slow SFH also coincides with the observed over-abundances of heavy s-process elements (hs) such as Ba and La relative to their Galactic counterparts at $[\text{Fe/H}] \gtrsim -1.5$, since there is more time for the hs elements to be produced by the s-process. On the other hand, at low metallicity $[\text{Fe/H}] \lesssim -2.5$, Aoki et al. (2009) and Tafelmeyer et al. (2010) found that dwarf galaxies do not

show over-abundances of $[hs/\text{Fe}]$ and tend to follow the Galactic halo trend. They argued that, at this metallicity, all of the n-capture elements are produced by the r-process. The light s-process elements (ls) such as Sr, Y and Zr do not however show over-abundances for dwarf galaxies at $[\text{Fe/H}] \gtrsim -1.5$, and are sometimes under-abundant, similar to $[\alpha/\text{Fe}]$. This is not unexpected. Although the s-process in Galactic low-mass AGB stars produces both ls and hs elements (Herwig 2005; Käppeler et al. 2011), it has been proposed that the relative under-abundances of the ls elements could be explained by the metallicity-dependence of the s-process (Tolstoy et al. 2009; Käppeler et al. 2011). Due to the *primary* nature (i.e. independent of metallicity) of the ^{13}C pocket as the major neutron source, theoretical studies suggest that metal-poor AGB stars will preferentially produce heavier s-process elements due to the high neutron-to-seed ratio at low metallicity (Gallino et al. 1998; Busso et al. 2001; Cristallo et al. 2009; Bisterzo et al. 2010).

3   DATA SELECTION

Different mechanisms contribute to different metallicity intervals. For example, mechanisms that are associated with more massive progenitors will preferentially contribute at lower $[\text{Fe/H}]$ due to the short evolution time, whereas the s-process in low-mass AGB stars will only contribute at higher $[\text{Fe/H}]$ due to their longer evolution time. The $\mathcal{C}$-space dimensionality and its interpretation may then be different in different metallicity ranges. Therefore we separate our discussion according to two major metallicity intervals: $-3.5 \lesssim [\text{Fe/H}] \lesssim -2$ and $-1 \lesssim [\text{Fe/H}] \lesssim 0$. These correspond roughly to the low metallicity halo and the high metallicity (thick + thin) disks of the Galaxy. In some cases we will also study the intermediate interval of $-2.5 \lesssim [\text{Fe/H}] \lesssim -1$. These metallicity ranges are just a rough guide and depend on the samples that we adopt from the literature.

There are many observational studies of elemental abundances in the literature (Edvardsson et al. 1993; McWilliam & Rich 1994; McWilliam 1998; Nissen & Schuster 1997, 2010; Hanson et al. 1998; Prochaska et al. 2000; Israelian et al. 2001; Carretta et al. 2002; Johnson & Bolte 2002a; Stephens & Boesgaard 2002; Chen et al. 2002, 2003; Nissen et al. 2002, 2004, 2007a,b; Gratton et al. 2003a,b; Bensby, Feltzing & Lundström 2003; Bensby et al. 2005; Akerman et al. 2004; Arnone et al. 2005; Jonsell et al. 2005; Caffau et al. 2005b; Asplund et al. 2005a, 2006; García Pérez et al. 2006; Allende Prieto et al. 2006; Preston et al. 2006; Fulbright, McWilliam, & Rich 2007; Lai et al. 2007, 2008; Ruchti et al. 2010; Fuhrmann 2011), but the analyzing models, solar abundances and the sources of stellar parameters vary from author to author. This may cause systematic differences in elemental abundances, and produce spurious dimensions in the PCA analysis. Therefore, where possible we do not use compilations of abundances from multiple sources. Many surveys are relatively restricted in the number of elements measured or number of stars (< 50) observed. Some surveys are strongly biased to specific classes of metal-poor stars, such as CEMP stars (e.g. Cohen et al. 2006), r-enhanced stars or related to planet hunting studies (e.g. Sousa et al. 2008; Neves et al. 2009) and therefore not suitable for our study,

---

[6] Some exclude LMC from the dwarf galaxies category (e.g. Mateo 1998). However, for simplicity, we do not distinguish fainter satellite galaxies and dwarf galaxies in this paper
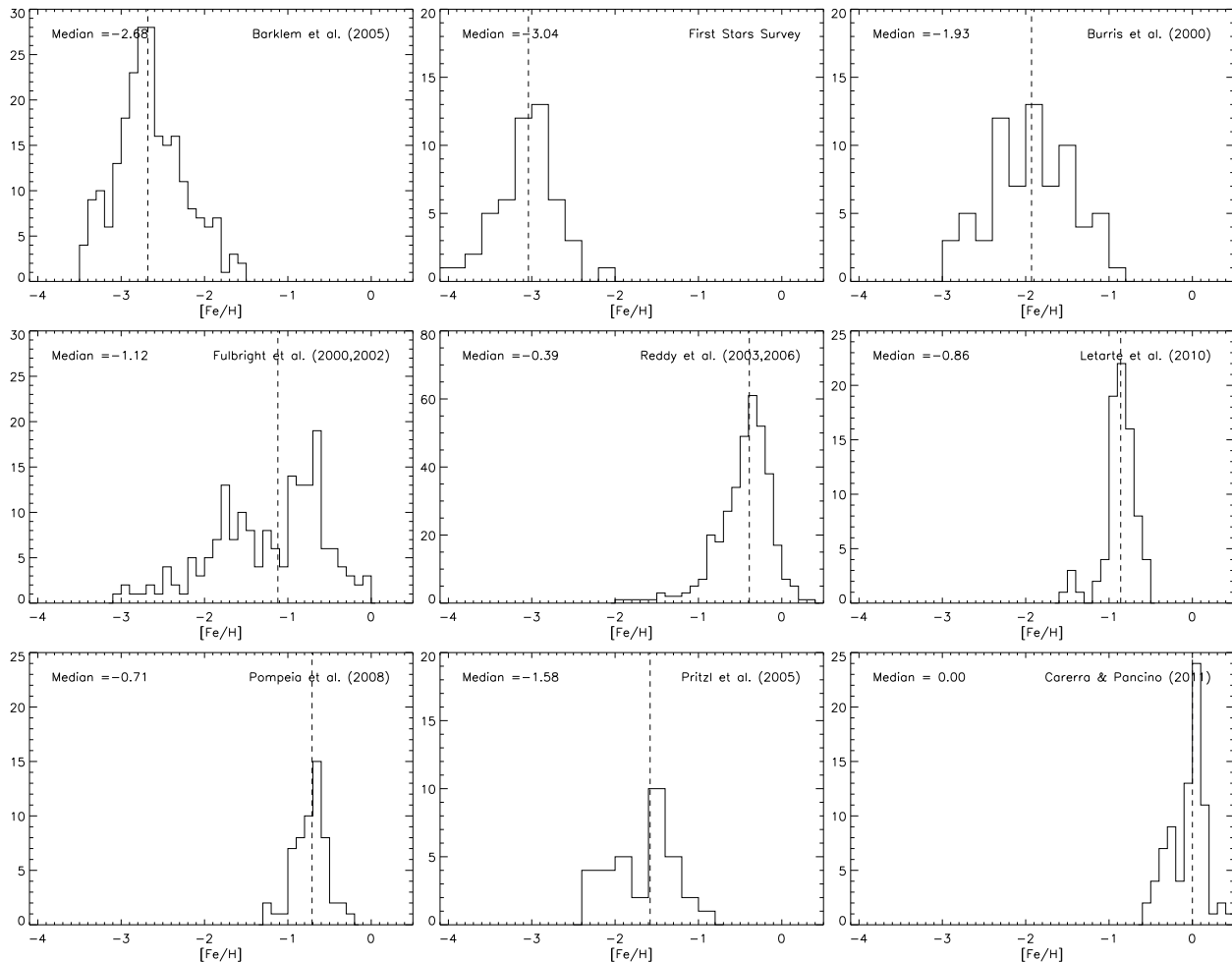
Figure 1. This figure shows the metallicity distribution of each sample that we have adopted. For solar neighbourhood stars, we have *low metallicity* stars from Barklem et al. (2005) and First Stars Survey; *intermediate metallicity* stars from Burris et al. (2000), Fulbright (2000, 2002); and *high metallicity* stars from Reddy et al. (2003, 2006). We also have the *Fornax dSph galaxy* sample from Letarte et al. (2010), the *LMC* sample from Pompeia et al. (2008), *MW globular cluster* compilation from Pritzl et al. (2005) and *MW open cluster* compilation from Carrera & Pancino (2011). The dashed lines show the median abundance of each sample.

because our goal is to measure the overall dimensionality of the $\mathcal{C}$-space. For each metallicity range, we focus on a few relatively unbiased, large samples of homogeneous data, and adopted their abundances directly from the original papers. For the open and globular clusters, we had no alternative but to use compilations.

3.1 Low metallicity

For solar neighbourhood low metallicity halo stars in this study, we use the observational data from Barklem et al. (2005) and the First Stars Survey (Cayrel et al. 2004; François et al. 2007; Bonifacio et al. 2009). From Barklem et al. (2005), we consider only stars in the range $-3.5 <$ [Fe/H] < -1.5. We exclude CEMP stars and blue stragglers as discussed below. In total, after culling, we have 231 stars from Barklem et al. (2005) and 50 stars from First Stars Survey.

Blue stragglers (Preston & Sneden 2000) and *some*

classes of CEMP stars are believed to have suffered binary mass transfer (cf. Section 2) and therefore might not reflect the abundances profile of the ISM from which they formed. We exclude known blue stragglers and CEMP stars with [C/Fe] > 1: this criterion is from Beers & Christlieb (2005). Having made this exclusion, the dimensionality of $\mathcal{C}$-space derived by our PCA method should be regarded as a lower limit, in case the carbon enhancement of some CEMP stars is not due to binary mass transfer (e.g. Aoki et al. 2002; Depagne et al. 2002), but to some distinctive astrophysical origins such as *faint* supernovae (Tsujimoto & Shigeyama 2003; Umeda & Nomoto 2003; Karlsson 2006; Kobayashi, Tominaga & Nomoto 2011b) or massive rotating stars (Meynet et al. 2006).

As the metallicity decreases, departures from Local Thermodynamic Equilibrium (LTE) are expected to become more pronounced (for a review, see Asplund 2005b). Therefore Non-LTE (NLTE) calculation are needed (e.g. Baumüeller, Butler & Gehren 1998; Gratton et al. 1999;

Mashonkina & Gehren 2001; Korn, Shi & Gehren 2003; Takeda et al. 2003; Andrievsky et al. 2009, 2010, 2011; Bergemann, Pickering & Gehren 2010a). However, note that Barklem et al. (2011) cautioned that estimating the inelastic hydrogen atom collisions with Drawin's formula in these studies might not be appropriate. As a large homogeneous NLTE-corrected sample is not yet available, and combining heterogeneous samples will give spurious dimensions in the PCA analysis due to the systematic differences, we decided to use available homogeneous 1D-LTE abundances. All the chemical abundances from 1D-LTE models are taken directly from the original papers, with the exception of the element Al.

We include Al in our analysis because, unlike Na and O, Al abundances do not suffer from significant internal mixing in Galactic stars (Andrievsky et al. 2008; Bonifacio et al. 2009). However, Al abundances are affected by NLTE effects at low metallicity. We adopted a +0.6 NLTE correction for [Al/Fe] as suggested by Baumüeller & Gehren (1997) and Cohen et al. (2004). Our results would not alter if we were to exclude Al. Al is the only NLTE-corrected abundance in this study. Our results should be reviewed once a large homogenized NLTE-corrected sample is available.

3.2 Intermediate, high metallicity

For solar neighbourhood intermediate metallicity stars, we chose to use the Burris et al. (2000) sample (70 stars) and Fulbright (2000, 2002) sample (178 stars). For high metallicity disk stars, we use the Reddy et al. (2003) and Reddy, Lambert & Allende Prieto (2006) samples (357 stars). These samples are adopted without further modification beyond culling obvious outliers (e.g. [Al/Fe] > 1, [V/Fe] > 0.8, [Co/Fe] > 0.7, [Nd/Fe] > 5, [Eu/Fe] > 5).

3.3 Dwarf galaxies

To compare the state of chemical evolution in the Milky Way and fainter satellite galaxies, we study the Fornax dSph galaxy using the Letarte et al. (2010) sample (80 stars). We exclude the star B058 because it is a metallicity outlier ([Fe/H] = −2.58) for this sample (−1.2 < [Fe/H] < −0.6). We also studied the LMC, using the homogeneous sample (57 stars) from Pompeia et al. (2008). For dwarf galaxies, to our knowledge, these are the only public available datasets with large (> 50 stars) homogeneous samples and elemental abundances measured at high resolution.

3.4 Globular and open clusters

The halo/disk star samples are mostly in the solar neighbourhood with $7.5 \lesssim r_G \lesssim 8.5$ kpc, where r_G is the distance from the Galactic centre. We would like to know whether stars over a larger Galactic volume would give us more inhomogeneity and therefore more independent dimensions in the $\mathcal{C}$-space. To probe wider regions, globular clusters (Searle & Zinn 1978), moving groups (e.g. De Silva et al. 2007; Bubar & King 2010) and open clusters (e.g. Yong, Carney & Teixera de Almeida 2005; Friel, Jacobson & Pilachowski 2010; Pancino et al.

2010; Andreuzzi et al. 2011; Jacobson, Friel & Pilachowski 2011) are suitable objects.

No large homogeneous survey of globular clusters including a wide variety of elements is available to date (but see Carretta et al. 2009, for a homogeneous survey with a restricted number of elements). We have to use the compilation of Pritzl, Venn & Irwin (2005) from different authors. In this compilation, mean abundances for each globular cluster were derived. Since we do not study light-elements like C, N, O, Na that have been shown to have star-to-star dispersion within a cluster, other elements should have small star-to-star dispersion within a cluster and therefore it is justified to take mean abundances for most of the elements in this study. However, we are aware that recently Roederer & Sneden (2011) have shown star-to-star dispersion in heavy *n*-capture elements (like La, Eu) and therefore taking mean abundances of each globular cluster will only give a lower limit of the $\mathcal{C}$-space dimensionality. We exclude objects identified with a dwarf spheroidal galaxy debris stream, such as Rup106, Pal12, Ter7 (e.g. Caffau et al. 2005a), and M68. Metallicity outliers are excluded by restricting [Fe/H] to the range −2.5 to −1. This leaves us with a total of 33 clusters.

We also study a recent open clusters compilation from Carrera & Pancino (2011) (private communication). We exclude those open clusters that have only [Fe/H] measurements. For clusters with multiple measurements, we take the mean abundance for each element. The sample then contains a total of 78 clusters, with Galactocentric radii $6.4 \leqslant r_G \leqslant 20.8$ kpc and $-0.57 \leqslant$ [Fe/H] $\leqslant 0.41$. For this compilation, we found that different model parameters lead to systematic differences of up to about 0.1 dex. Other sources of systematics such as differences in model atmospheres and methodology (e.g. equivalent width vs. spectrum synthesis), are difficult to quantify but are likely to be smaller. Thus we estimate that the systematic differences can be up to 0.2 dex depending on the elements.

Since we will not perform PCA combining multiple samples other than globular clusters and open clusters, and the solar abundance differences for elements heavier than Na are usually small (e.g. Anders & Grevesse 1989; Grevesse & Sauval 1998; Asplund, Grevesse & Sauval 2005c; Asplund et al. 2009; Lodders, Palme & Gail 2009), we do not homogenize the solar abundances adopted by different authors. The metallicity distribution of each sample is shown in Fig. 1 and we summarize our adopted samples in Table 1.

4 ANALYSIS METHOD

4.1 PCA

We start with the $\mathcal{C}$-space defined by the set of element ratios [X$_i$/Fe]. We form the matrix of the correlation coefficients between all pairs of element ratios [X$_i$/Fe], [X$_j$/Fe] and diagonalize this matrix. One can show (Appendix A) that the eigenvector corresponding to the largest eigenvalue is the direction where we have the largest variance in the *mean-shifted, normalized* data set and so on for the successively smaller eigenvalues. Furthermore, the variances along those directions are given by the corresponding eigenvalues. Note

Table 1. Summary of adopted samples in this study.

References	Categories	Count*
MW Solar Neighbourhood		
Barklem et al. (2005)	Metal-poor halo stars	231
First Stars Survey[†]	Metal-poor halo stars	50
Reddy et al. (2003, 2006)	Metal-rich disk stars	357
Fulbright (2000, 2002)	In between	178
Burris et al. (2000)	In between	70
Others		
Pritzl et al. (2005)	MW globular clusters	33**
Carrera & Pancino (2011)	MW open clusters	78**
Letarte et al. (2010)	Fornax dSph galaxy	80
Pompeia et al. (2008)	The LMC	57

* After restricting metallicity range; culling outliers, CEMP and blue stranglers
[†] Cayrel et al. (2004); Bonifacio et al. (2007); François et al. (2007)
** Number of clusters

that the matrix is symmetric, therefore the eigenvectors are orthogonal. These eigenvectors are the *principal components* and we can, *to some extent*, interpret these eigenvectors in terms of nucleosynthetic processes. We illustrate the procedure with toy models.

4.1.1 Toy models

Take a 3-dimensional space defined by three elements which we denote as El1, El2 and El3, and assume that there are two element-producing mechanisms. The first produces exactly the same amount of El1, El2 and El3 and the second mechanism only produces El1 and El2 with production ratio El1/El2 $= \gamma$. Figs. 2 and 3 illustrate the cases $\gamma = 1$ and $\gamma = 2$. For each figure, panel (a) shows the scenario where only mechanism 1 is working: all elements are produced in exactly the same ratio. Panel (b) shows the scatter plot with both mechanism 1 and mechanism 2 contributing to the element abundances: e.g. in panel (b), the points have now been randomly translated in the direction (1,1,0) by the action of the second mechanism.

The first principal component of the PCA analysis, shown in panel (b) as the *blue solid line*, represents the direction that has largest variance in the *normalized, mean shifted* $\mathcal{C}$-space. After determining the first component (with the largest eigenvalue), the PCA machinery then projects orthogonally all data points on to the hyperplane normal to the first principal component, as shown in panel (c) and panel (d).

The second principal component is then the direction in the hyperplane that shows the largest variance among the projected points. Panel (d) shows the data points projected on the hyperplane of the first principal component. The second principal component is shown as the *red solid line*. The composition of the normalized eigenvectors in the 3-dimensional $\mathcal{C}$-space are shown in the bar chart (e), with the first eigenvector as the three bars on the left and the second eigenvector as the three bars to the right. The first eigenvector has approximately equal components in each element and represents the contribution of mechanism 1. Mech-

anism 1 has equal components, but the eigenvector shows small departures from equality which come from projections of random amounts of mechanism 2 on to this eigenvector.

The first eigenvector has a clear interpretation: in the bar chart (panel e), the contributions from mechanism 1 have similar components for all elements, as expected. The second principal component represents the contribution of mechanism 2. In the bar chart, El1 and El2 have the same sign but El3 has the opposite sign. It is important to note that, although the second mechanism does not produce El3, the contribution of El3 to the second principal component is not zero. This is intuitively clear, because the second component must be orthogonal to the first and must lie in the hyperplane shown in panel (c). The direction where we have both positive El1 and El2 in the tilted hyperplane is pointing in the negative direction of El3. This illustrates an issue with interpreting the eigenvectors in terms of nucleosynthetic processes in more realistic systems. Although both principal components are well defined, the first eigenvector has a clear interpretation, but the interpretation of the second and later eigenvectors is less straightforward.

Comparing the second eigenvector for $\gamma = 1$ and $\gamma = 2$, El1 and El2 show similar contributions for $\gamma = 1$, but for $\gamma = 2$ the contribution from El1 is larger. This is not surprising; as we have seen in panel (d), the second eigenvector on the hyperplane shows a ratio of El1 to El2 > 1 when $\gamma = 2$. As before, although both principal components are well defined by the PCA, the first eigenvector has a clear interpretation but it is less obvious for the second eigenvector. When one component is in the positive diagonal direction (i.e. all elements contribute with the same sign) and the second eigenvector components show opposite signs of different families (here El1 & El2 vs El3), this suggests that there is contribution from a first mechanism that produces both families, and contribution from a second mechanism that preferentially produces El1 & El2 or preferentially produces El3 but we cannot tell which without further diagnostics. Section 5.1 gives an example of a situation in which further diagnostics can illuminate the interpretation of second and later eigenvectors.

4.1.2 Dealing with incomplete data sets

Given an incomplete data set (i.e. some of the element abundances are not available for all stars), in principle we can still calculate Pearson's correlation for any two chemical abundances by using only the data points that have both chemical abundances, and therefore we can construct the correlation matrix entry by entry. However in this case the correlation matrix is clearly not a Gram matrix[7] and therefore might not necessarily be semi-positive-definite and we might have unphysical negative eigenvalues.

This is a well-known problem in finance analysis. We use the algorithm suggested by Rebonato & Jäckel (1999). The idea is to search for a closest semi-positive-definite matrix **C'** that resembles the correlation matrix (cf. Appendix B)[8].

[7] A $n \times n$ matrix is Gramian if and only if for every $i, j \in \{1, 2, \ldots, n\}$, the (i,j)-th entry is given by the inner product $\langle \boldsymbol{v_i}, \boldsymbol{v_j} \rangle$ of the same set of vectors $\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_n}$
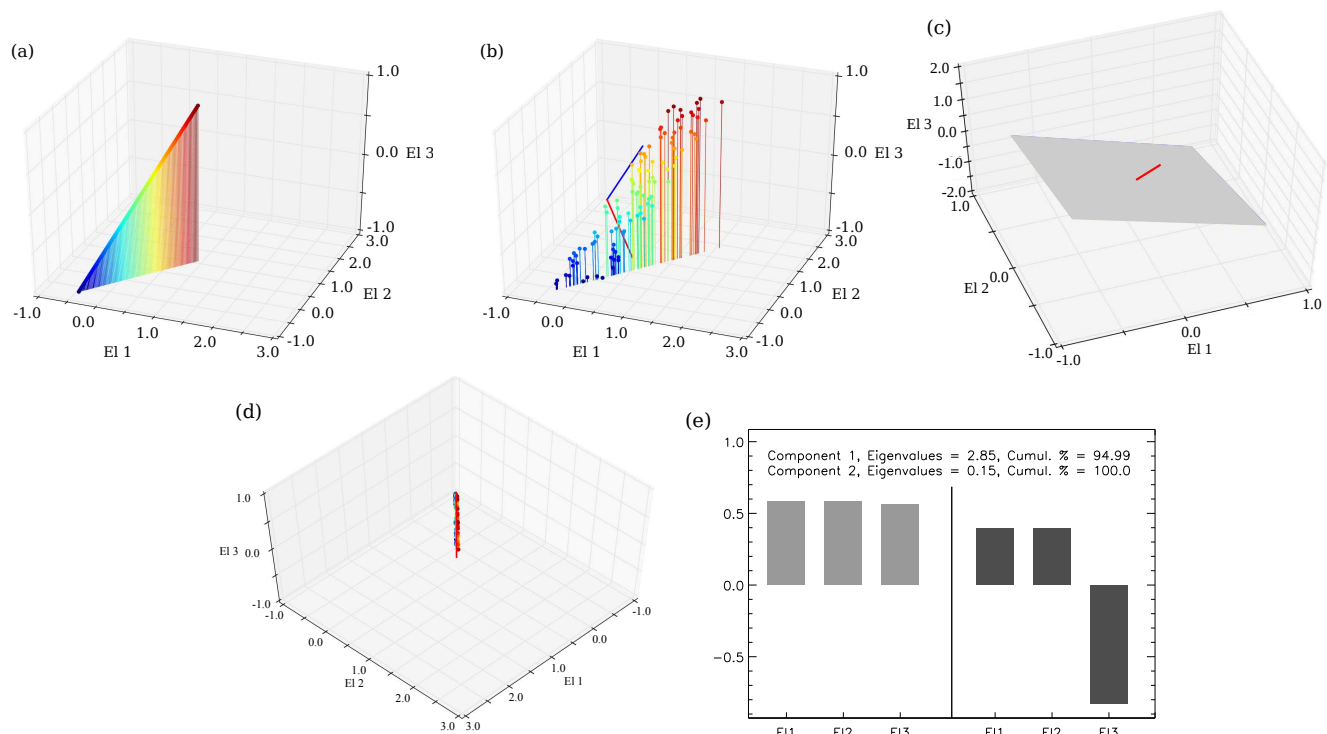[8] Another method to deal with missing data is by guessing the

Figure 2. This is a model ($\gamma = 1$) to illustrate how PCA works. The vertical lines in panel (a) and (b) are visual guides. Panel (a) shows the scatter plot of the chemical abundances where only mechanism 1 is working (see text). Panel (b) is the same as (a) but both mechanisms 1 & 2 are contributing to the chemical abundances. Panel (b) also shows the first principal component in *blue solid line* and the second principal component in *red solid line*. Panel (c) shows the the hyperplane of the first principal component in *grey* and the second principal component in *red solid line*. Panel (d) shows the data points projected onto the hyperplane of the first principal component and the *red solid line* is the second principal component. Panel (e) shows the composition of the first two normalized eigenvectors. Colour version of all figures in this paper are available in the online version of the journal.
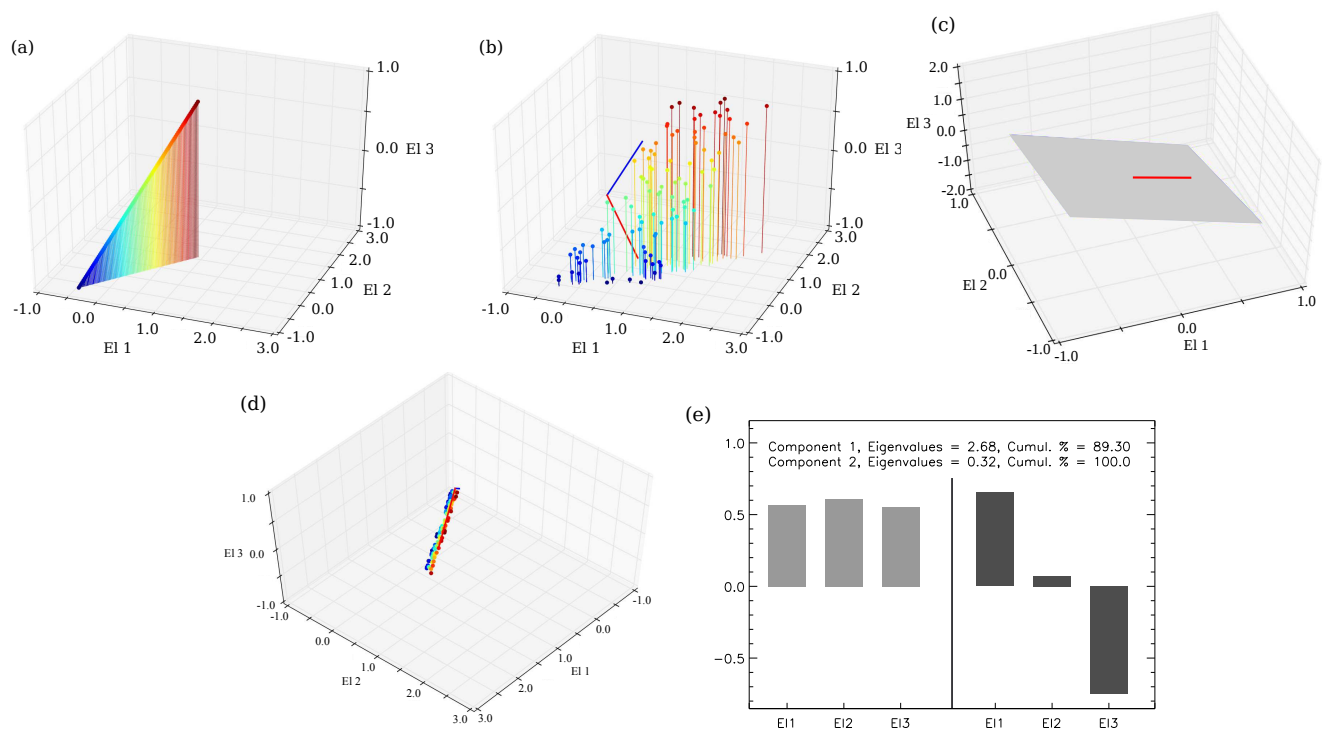

Figure 3. Same as Fig. 2, but with $\gamma = 2$.

We can measure the differences between these two matrices using the canonical matrix norm or the quadratic sum of the differences of the eigenvalues (cf. Appendix B). Throughout our study, the differences of either norm over the number of entries gave value $\ll 1$ and therefore the results are robust.

After *orthogonally* diagonalizing the matrix **C'**, we rank the eigenvalues in decreasing order and calculate their cumulative percentages. The cumulative percentages of ranked-eigenvalues represents the accounted total variances of the data cloud.

4.1.3 Best cut-off for ranked-eigenvalues cumulative percentages

Given that the measurements are not perfect, what percentage of the variance of the data cloud is due to measurement uncertainty? For example, Section 5.1 shows the cumulative percentages for the ranked-eigenvalues for the n-capture elements at low metallicity abundance. The first component provides about 80% of the variance. The next two provide almost all of the rest of the variance. Our goal is to find the dimensionality of the $\mathcal{C}$-space. How many of the principal components should we accept as real? We attempted to answer this question by performing Monte Carlo simulations, in which we created mock data sets lying in an n-dimensional spaces and tracing a non-tilted m-dimensional flat manifold[9]. The mock data set has a spread of $\Delta = 0.3$ to 0.7 dex mimicking the real situation with element abundance distributions. We performed the simulation by rotating the manifold with a random orthogonal matrix. For each simulation, we added 0.1 dex of measurement uncertainty and then performed PCA on the noisy data. We recorded the cumulative percentages corresponding to the m-th principal component. In this way, we tried to estimate the cumulative percentage at which we should stop accepting principal components as real, in order to deduce the correct m-dimensions of the manifold.

For each set of values of n and m, we performed 10000 simulations. We varied n, m and Δ in the range $n \in [7, 15]$,

---

missing data from the best fit model, where the best fit model is deduced from the rest of the data. However in this case, unavoidably, we will strengthen our correlation unnecessarily. This illustrates a problem with most of the manifold learning methods such as Locally Linear Embedding (LLE), ISOMAP, Diffusion Map etc. These methods require a complete data set. Therefore there is no alternative to guessing the missing data. We argue that the method that we introduce here is a more democratic way to deal with missing data without imposing any prior.

[9] A m-dimensional flat manifold in n-dimensional space is the generalization of a 2-dimensional plane in 3-dimensional space. In this study, we made two assumptions: (1) If correlated, $[X_i/\text{Fe}]$ are correlated linearly; (2) The errors of $[X_i/\text{Fe}]$ are Gaussian-distributed. The first assumption is reasonable because $[X_i/\text{Fe}]$ are abundances in log scale– the assumption holds if $N_{X_i} \propto N_{X_j}^m$, for all $m \in \mathbb{R}$, where N_X is the abundance of element X. Furthermore, our study samples are always restricted to a small metallicity abundance range. To justify the second assumption, we compared the [Fe/H] of stars in common of RAVE Survey (Steinmetz et al. 2006) and Geneva-Copenhagen Survey (Nordström et al. 2004). We found that the differences can be approximated by a Gaussian distribution. This suggests that the abundance errors in log scale could be Gaussian-distributed.

$m \in [2, 5]$, $\Delta \in [0.3, 0.7]$. The simulation showed that a reasonable cutoff for identifying real principal components in the PCA analysis is about 85%.

In summary, assuming a cosmic spread of $\Delta = 0.3$–0.7, measurement uncertainty of 0.1 dex and a flat manifold (which can be tilted), the rank of the eigenvalue corresponding to the cumulative percentage of about 85% gives a robust estimate of the independent dimensions of the chemical space. The remaining variance of the data cloud comes from measurement uncertainty. If the measurement uncertainty is less than 0.1 dex, we should take the cut off larger than 85% and vice versa. In this study we will perform PCA analysis on the random variables $[X/\text{Fe}]$[10], where X can include Al, Sc (light odd-Z elements); Mg, Si, Ca, Ti (α-elements), V, Mn, Cr, Co, Ni, Zn (Fe-peak elements) and Y, Zr, Ba, La, Nd, Eu (n-capture elements). Among n-capture elements, we consider Y and Zr as light s-process (ls) elements, Ba and La as heavy s-process (hs) elements and Nd, Eu as mostly r-process elements (refer to Arlandini et al. 1999; Burris et al. 2000). This choice of elements is determined by those in the available abundance surveys (see Table 1). The *HERMES* survey will include some other elements, and their contribution is discussed later.

It is important to note that the assumption $\Delta = 0.3$–0.7 might not hold for some elements, such as Ni. Our simulations show that if the intrinsic cosmic scatter for *all* $[X/\text{Fe}]$ is 0.3 dex or less, the simulated noise will dominate over the intrinsic cosmic scatter and the the estimated cut-off will start to drop significantly (about 70%). However, this will not qualitatively alter our conclusion: for n-capture elements subspace, the cosmic scatters of n-capture elements are believed to be larger than 0.3 dex. For the all elements space, we only interpreted the first four eigenvectors, for which the cumulative eigenvalues percentage is about 75%, cf. Section 5.

4.2 Estimate of intrinsic correlation

For typical samples of stellar abundances with typical measuring errors, we showed in Section 4.1.3 that we can take as real the eigenvectors or principal components contributing to the first 85% or so of the cumulative percentages for the ranked eigenvalues. Now we address a different issue. We may need to compare two samples which have different measurement uncertainties. For example, when comparing the principal components for abundances in dwarf galaxies and the Milky Way, it is important to find a way to correct for the different measuring uncertainties.

We can do this by estimating via simulations the intrinsic correlations for each sample, i.e. the values of the correlations if the measurement were perfect, without uncertainties. Then we can directly compare the principal components for each sample. Having reduced the contribution of noise to the correlations, the level of the cumulative percentages for the real ranked eigenvalues is now larger than 85%, and close

---

[10] We chose to work in $[X/\text{Fe}]$ space because all elements are highly correlated in $[X/\text{H}]$, i.e. if we were to perform PCA on $[X/\text{H}]$, the dominant dimension will consume more than 80% of the variance for all cases.

to 100%, though not exactly 100% because of the residual uncertainties in the noise reduction process.

In the simulations, given two correlated elements $\mathcal{X}_1$ and $\mathcal{X}_2$, first we need to find the best fit line to $[\mathcal{X}_1/\text{Fe}]$ and $[\mathcal{X}_2/\text{Fe}]$ to create the mock data set. Instead of deriving the best fit line using weighted or bi-weighted *linear* least squares, we searched for the best fit line by weighted *total* least squares. For more details, refer to Krystek & Anton (2007), and Appendix C.

This is crucial to make sure that the best fit line is symmetric in fitting y to x and fitting x to y and therefore minimize the differences between estimating in the forward and reverse directions. Let (x_k, y_k) be the k-th data point, and $u_{x,k}$ and $u_{y,k}$ be the corresponding measurement uncertainties. In the special case $u_{x,k} = u_{y,k} = \sigma$, minimizing

$$\chi^2 = \frac{1}{n-2} \sum_{k=1}^{n} \left[\frac{(x_k - X_k)^2}{u_{x,k}^2} + \frac{(y_k - Y_k)^2}{u_{y,k}^2} \right] \qquad (1)$$

is equivalent to minimizing the weighted orthogonal distance of the measured points (x_k, y_k) to the fitting line $y = ax + b$ where (X_k, Y_k) is the orthogonal projection of the data point (x_k, y_k) to the line.

In the simulation, we start with perfectly correlated mock data points lying on the best fit line, as shown in Fig. 4, the *red open circles*. The *black filled circles* are the observed data. We then add simulated cosmic scatter on the *red open circles*. The updated mock data points are now denoted as *blue open circles*. We then add simulated measurement uncertainty (known amplitude) to the *blue open circles* and the updated mock data points are denoted with *orange crosses*. The mock data are adjusted until the mock data with simulated noise (cosmic scatter and measurement uncertainty) have the same scatter as the observed data, i.e. the program is iterated until the *orange crosses* have the same correlation as the *black filled circles*. The average of Pearson's correlation of the *blue open circles* from forward and reverse fitting is used as the estimated intrinsic correlation. For every element pair $\mathcal{X}_1$ and $\mathcal{X}_2$, we performed 100 Monte Carlo simulations and adopted the median as the best estimate. This procedure works well for element pairs with measured correlations > 0.1 for which the correlations are probably real. Some element pairs have smaller correlations, and the procedure is not so obvious.

The goal of much of our analysis is to compare the eigenvectors and dimensionality for various domains of [Fe/H] and subsamples like the dwarf galaxies and star clusters. When the correlation between an element pair is small, its sign may not be correct because of noise. The Monte Carlo procedure described above will not change the sign of the correlation so cannot account for an error in sign of a small correlation. We therefore also explored the possibility that the true sign of the correlation is different from the one observed and then estimated what effect this has on the dimensionality; it is always very small. For the results reported below, we were guided by the larger stellar samples in adopting the sign of a small correlation for the Monte Carlo modeling.

5 ANALYSIS RESULTS

In this section, we show the PCA analysis results of n-capture elements alone and then all elements, including light
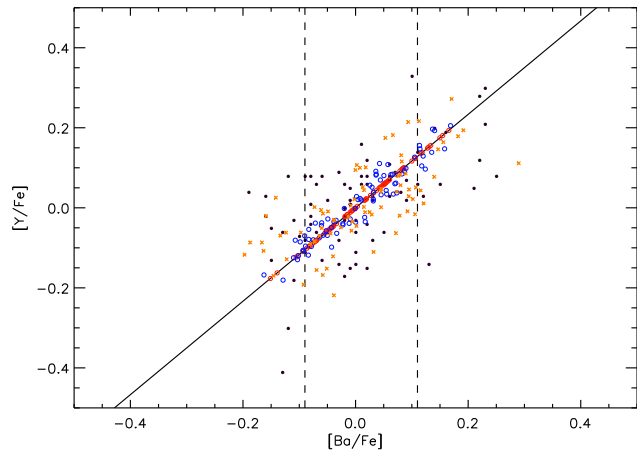

Figure 4. A example of simulated mock data to calculate the intrinsic correlation between [Y/Fe] and [Ba/Fe] for the high metallicity sample, in metallicity range $-1.2 < [\text{Fe/H}] < -0.6$. Here we assume a measurement uncertainty of 0.05 dex for both [Y/Fe] and [Ba/Fe]. The symbols are explained in the text. The *black solid line* is the best fit line calculated using weighted total square. The *black dashed vertical lines* are the 1σ of the observed data points distribution, assuming Gaussian distribution.

odd-Z, α-, Fe-peak and n-capture elements. We discuss stars of low metallicity and high metallicity separately, and then consider open clusters, dwarf galaxies and globular clusters.

5.1 Low metallicity stars

We first study the dimensionality of the low metallicity stars using the ranked-eigenvalues cumulative percentages. Fig. 5 shows the results of the Barklem's sample (solid line) and First Stars Survey sample (dotted line) considering firstly the n-capture elements (Y, Zr, Ba, Nd, and Eu) alone. The vertical dashed line shows the principal component dimensions that corresponds to 85% of the cumulative percentage. Both samples showed that the n-capture elements at low metallicity have a strong component that makes up almost all of the variances.

We then examine the individual components. Fig. 6 shows the composition of the first two eigenvectors (first two principal components), i.e. the first two directions that are orthogonal and account for most of the variance in the normalized mean shifted $\mathcal{C}$-space. The upper plot and lower plot show the results of the Barklem sample and First Star Survey sample respectively. The colour is coded as a rough grouping to differentiate families of elements from different major production sources: *dark red* represents light s-process elements, *red* stands for heavy s-process elements, and mostly r-process elements are coded as *orange*. For simplicity, on the x-axis we denote [X/Fe] as X in all such figures in this paper. Since the eigenvectors are normalized, the quadratic sum of the five [X/Fe] abundances contribution for each principal component equals 1.

In the first component, we see that all elements have the same sign. This illustrates that this dominant principal component is pointing in the positive diagonal direction in this 5-dimensional $\mathcal{C}$-subspace. We can infer that the first com-
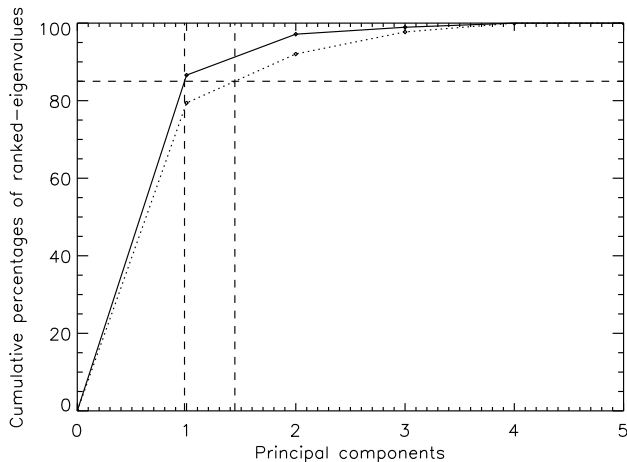
Figure 5. This ranked-eigenvalues cumulative percentages of *n*-capture elements (Y, Zr, Ba, Nd, and Eu) for the low metallicity samples– Barklem et al. (2005) sample (solid line) and First Stars Survey sample (dotted line).

ponent represents a mechanism that produce all 5 elements simultaneously. This suggests that all of these *n*-capture elements are mostly produced by a dominant primary process at low redshift, which could be the *r*-process (Truran et al. 2002; Wanajo & Ishimaru 2006).

The second eigenvector points in the anti-diagonal direction of the light *s*-process elements (Y, Zr) and the heavier *n*-capture elements (Ba, Nd, Eu), i.e. this component represents a mechanism that produces mainly light *s*-process elements or mainly heavy *n*-capture elements. Fig. 7 helps to resolve the nature of this mechanism: we see that the light *s*-process elements are over-abundant at low metallicity, and therefore the former interpretation is preferred. This second principal component might correspond to the contribution of LEPP (Travaglio et al. 2004; Qian & Wasserburg 2007; Izutani et al. 2009). However, as shown in the ranked-eigenvalues cumulative percentages plot (cf. Fig. 5), the contribution of this component is small relative to the first principal component. We discuss this further in Section 6.

To show the relative importance of the two components as a function of metallicity, we performed the PCA analysis on stars in [Fe/H] bins of 0.7 dex. We calculated the ratio of the second component eigenvalue to the sum of the eigenvalues for the first two components to estimate the contribution of the second component to the overall *n*-capture element yield. The results is shown in the upper panel of Table 2. The same calculation was repeated for the First Stars Survey sample and is shown in lower panel of Table 2. The fractional contribution of the second component appears to decrease with increasing metallicity. We have performed Kendall's τ and Spearman's ρ tests: both tests showed that this trend is significant ($< 5\%$ probability of a false positive correlation).

Note that La is not included in the analysis of *n*-capture elements, in order to compare the low metallicity results with those for high metallicity. La is not measured in our high metallicity Reddy's sample. However, we have tested our results by adding La or restricting ourselves to a smaller subspace. In either case, it does not alter the results quali-
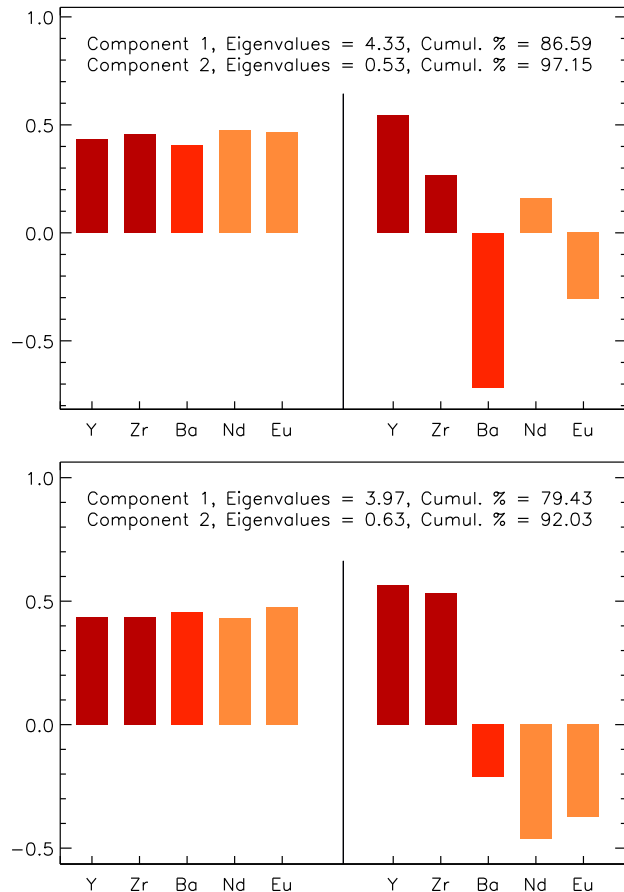

Figure 6. The composition of the first two *n*-capture elements normalized principal components, i.e. the eigenvector components of the *n*-capture elements correlation matrix, using the low metallicity samples: Barklem's sample (upper plot) and First Stars Survey sample (lower plot). The horizontal axis shows the 5 n-capture elements and the vertical axis shows the component of the eigenvector in the direction of each of the 5 elements.. See the text for the colour code.

Table 2. Fraction of the second component contribution to the *n*-capture elements yield: Barklem's sample (upper panel), the First Stars Survey sample (lower panel).

[Fe/H] range	Fraction(%)
$(-3.4, -2.7)$	10.15
$(-3.3, -2.6)$	10.09
$(-3.2, -2.5)$	8.41
$(-3.1, -2.4)$	7.88
$(-3.0, -2.3)$	7.09
$(-2.9, -2.2)$	5.62

[Fe/H] range	Fraction(%)
$(-3.5, -2.8)$	18.10
$(-3.4, -2.7)$	16.78
$(-3.3, -2.6)$	13.67
$(-3.2, -2.5)$	10.96

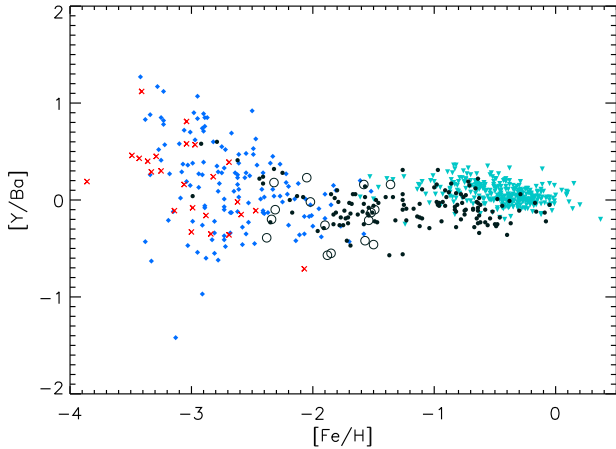**Figure 7.** Y (*ls*) to Ba (*hs*) ratios as a function of [Fe/H] for the Milky Way halo/disk stars from First Stars Survey (*red crosses*), Barklem et al. (2005) (*blue diamonds*), Fulbright (2000, 2002) (*black filled circles*), Reddy et al. (2003, 2006) (*cyan triangles*) and globular clusters from Pritzl et al. (2005) (*black open circles*) All samples assume 1D-LTE stellar model.

tatively. The rest of the principal components are probably due to measurement uncertainty since the first two components already account for $> 90\%$ of the data cloud variance.

Now we calculate the principal components for the whole set of elements available for the low metallicity stars. The principal components of PCA analysis on 17 elements (Al, Sc, Mg, Ca, Ti, V, Cr, Mn, Co, Ni, Zn, Y, Zr, Ba, La, Nd, and Eu) for Barklem's sample are shown in Fig. 8. The colour is coded to differentiate categories of elements: the *n*-capture elements are coded as before, light odd-Z elements are colour-coded with *dark brown*, α-elements with *blue*, Fe-peak elements (beside Cr & Mn) with *green*, and Cr & Mn with *black*. If we adopt a 85% cutoff for the ranked-eigenvalues cumulative percentages, as shown in Fig. 9, we find about 6 independent dimensions for the 17-dimensional $\mathcal{C}$-space. The number of dimensions is not altered if we apply a +0.6 dex NLTE correction for [Al/Fe], as shown in Fig. 9.

The first four principal components of this 17-dimensional $\mathcal{C}$-space can be summarized as follows:

• The first component shows contributions primarily from *all n*-capture elements and also contribution from α-elements. This suggests that the production sites of the *r*-process also produce significant amounts of α-elements, which may be consistent with core-collapse supernovae as the *r*-process site rather than neutron star mergers. The first component could be the contribution from core-collapse supernovae that involve the *r*-process.

• The second component shows an anticorrelation of α-elements with Fe-peak elements and *n*-capture elements. This suggests that there is a mechanism that produces α-elements but does not produce Fe-peak and *n*-capture elements. This component may correspond to the contribution from 'normal' core-collapse supernovae that do not involve the *r*-process
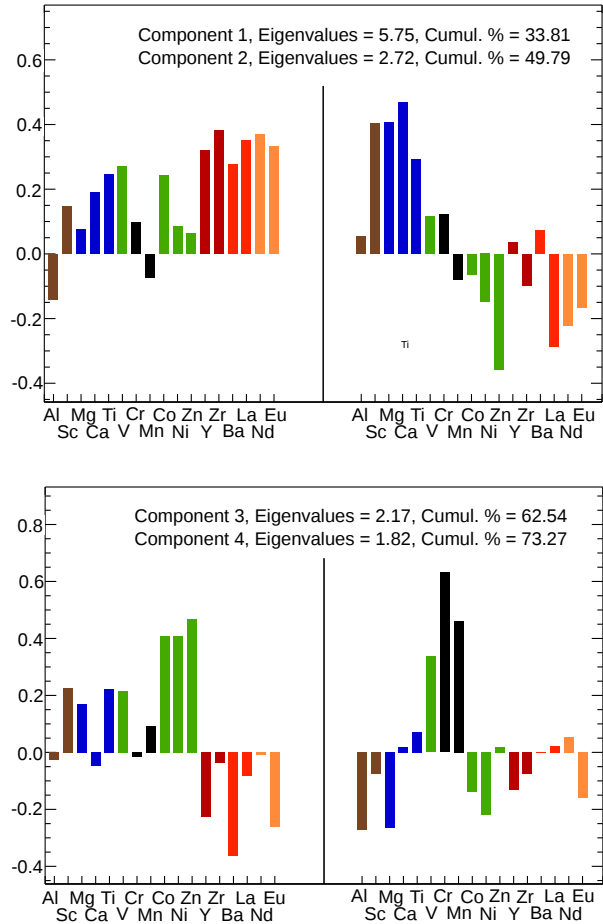
• The third component shows an anticorrelation of α-

Figure 8. The normalized principal components of 17 elements for the low metallicity sample (Barklem et al. 2005). [Al/Fe] is corrected for NLTE effect with +0.6 dex. The upper plot and lower plot show the first two principal components and the third and fourth principal components, respectively.
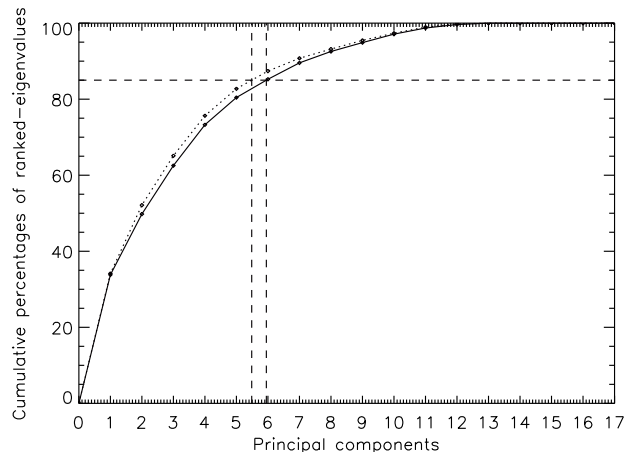

Figure 9. The ranked-eigenvalues cumulative percentages of 17 elements for the low metallicity sample (Barklem et al. 2005). The *solid line* and *dotted line* show the results with and without the NLTE correction of [Al/Fe], respectively.
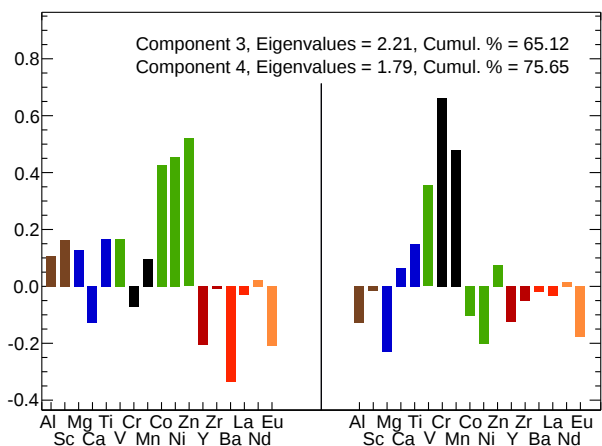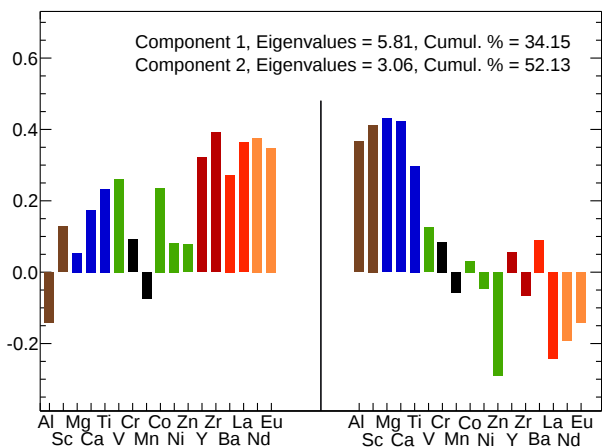
Figure 10. The same as Fig. 8 but without the NLTE correction for [Al/Fe].

elements and Fe-peak elements with the n-capture elements. This suggests the production of *both* α-elements and Fe-peak elements. This component may correspond to the contribution of hypernovae (e.g. Kobayashi et al. 2006) where significant amounts of Co, Ni, Zn are produced with high explosion energy.

• The fourth component shows a strong contribution from Cr and Mn. Different from heavier Fe peak elements, Cr and Mn are synthesized in the incomplete Si-burning region, which is the outer region of the complete Si-burning region. We discuss this further in Section 6.

Note that with the NLTE correction, Al does not contribute to the first four components. If we do not apply NLTE correction for Al, Al contributes to the second component as shown in Fig. 10, but the results for other elements remain unaltered. We will discuss this further in Section 6.

5.2   High metallicity stars

Fig. 11 and Fig. 12 show the PCA analysis results for the n-capture elements (Y, Zr, Ba, Nd, Eu) in the Reddy high metallicity sample. This results show two marked differences from the low metallicity sample (cf. Fig. 5, Fig. 6): (1) The first principal component contributes far less (56%) to the overall variance compared with low metallicity sample (79
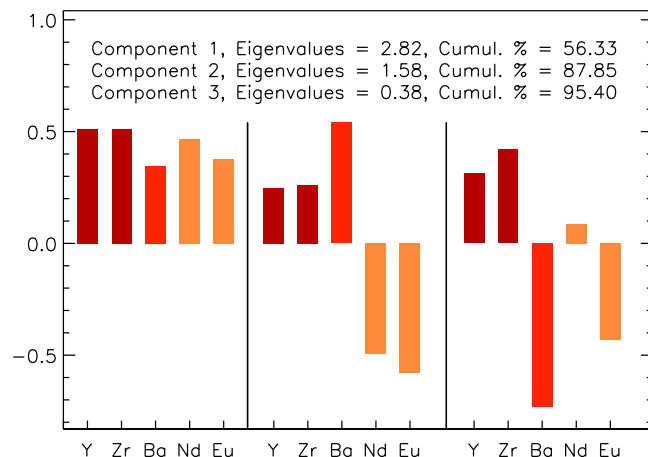

Figure 11. The composition of the first three normalized principal components for the n-capture elements for the high metallicity sample.
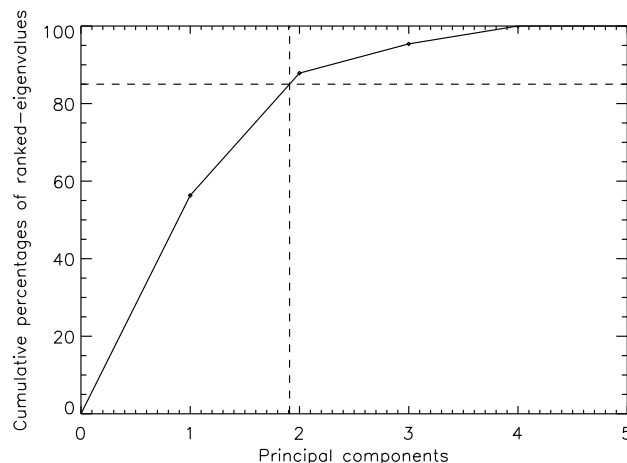

Figure 12. The ranked-eigenvalues cumulative percentages for the n-capture elements for the high-metallicity sample.

– 87%). It is clear that at high metallicity, there is more than one mechanism contributing to the n-capture elements; (2) At high metallicity, the second component is made up of both light s-process elements (Y and Zr) and heavy s-process elements (Ba), instead of light s-process elements alone as in the low metallicity sample anticorrelating with mostly r-process elements (Nd and Eu). This component may correspond to the s-process in low-mass AGB stars. The s-process in low-mass AGB stars, operating in metal-rich environment, is expected to produce both light s-process and heavy s-process elements. The third component might be the remnant of LEPP contribution at low metallicity, or it might just be noise, since the first two components already account for 88% of the total variance.

Fig. 13 shows the first four principal components from the PCA analysis of all 17 elements (Al, Sc, Mg, Si, Ca, Ti, V, Cr, Mn, Co, Ni, Zn, Y, Zr, Ba, Nd, and Eu) which are available for the high metallicity sample. The colour coding
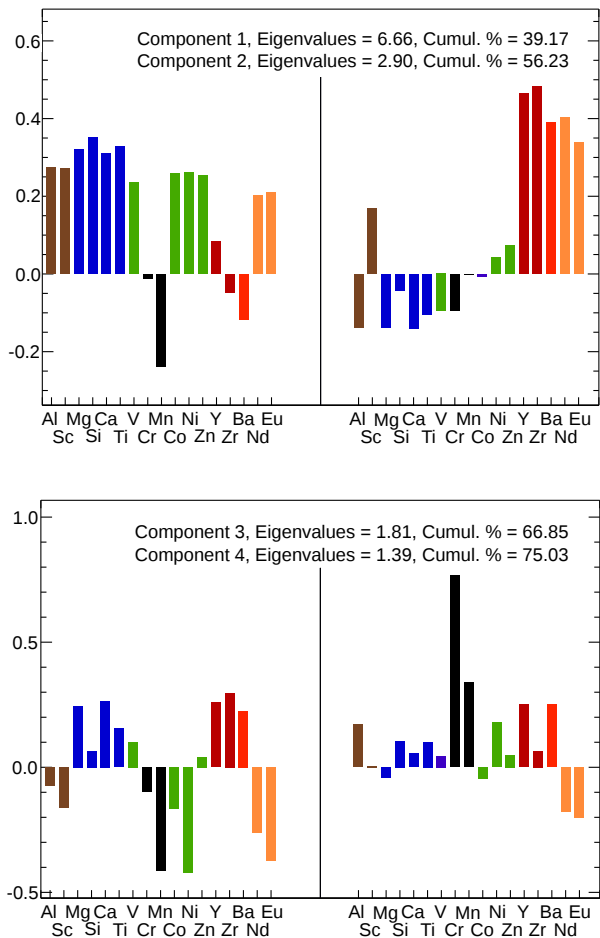
Figure 13. The normalized principal components of 17 elements for the high metallicity sample. The upper and lower plot show the first two principal components and the third and fourth principal components, respectively.

Figure 14. The ranked-eigenvalues cumulative percentages of 17 elements for the high metallicity sample.

is the same as in Fig. 8. The results are summarized as follows:

• The first component is a strong component, accounts for about 40% of the total variance, and is made up of α-elements and Fe-peak elements (except for Mn and Cr) as well as light odd-Z element Al (1D-LTE-based). Note that NLTE correction is supposed to be negligible at high metallicity. We recall that in the low metallicity sample, Al was absent from the first few principal components when NLTE-corrected. This component may correspond to the well-mixed ISM that is affected by many generations of core-collapse supernovae. The deficit of Mn suggests that this component does not include the contribution from SNe Ia.

• The second component is made up mainly of all of the n-capture elements, which could be the contribution from r-process. However, unlike the first principal component of the low metallicity sample which also includes all of the n-capture elements, the α-elements contribution are not visible in this second component for the high metallicity sample. We discuss this in more detail in Section 6.

• The third component shows anticorrelation between light s-process, heavy s-process, α-elements and the other
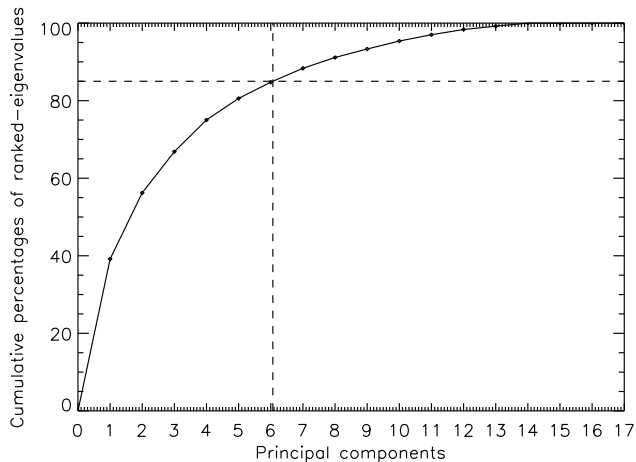
elements. This could correspond to the contribution of AGB stars ejecta.

• The fourth component is again dominated by Cr and Mn. We will discuss this further in Section 6.

Adopting a 85% cutoff for the ranked-eigenvalues cumulative percentages, as shown in Fig. 14, we again have 6 independent dimensions in the 17 dimensional $\mathcal{C}$-space, as for the low-metallicity sample. This is not totally unexpected. Although dimensions are lost by the homogenization of ISM at higher metallicity (Karlsson & Gustafsson 2001; Bland-Hawthorn et al. 2010), there are also processes such as the AGB stars and SNe Ia that will contribute only at higher metallicity. This extra 'birth-imprint' appears to compensate for the loss of dimensions due to homogenization.

5.3 Open clusters

Now we turn to the open clusters, which cover a larger volume of the Galaxy than the stellar samples discussed so far. We do not make a separate study of their n-capture element subspace because there are no Eu data in the open cluster compilation. Eu is potentially important to separate the r-process and s-process contributions. Instead, we analyze the subspace of α-elements: Mg, Si, Ca, Ti. We then analyze the all-elements space, including the 13 elements: Al, Sc, Mg, Si, Ca, Ti, V, Cr, Co, Ni, Y, Ba, and Nd. We compare the results for the open cluster compilation with the Reddy solar neighbourhood sample by restricting the metallicity range of the solar neighbourhood sample to [Fe/H] > −0.5 to match the range for the open clusters.

The ranked eigenvalues shown as the *solid lines* in Fig. 15 and Fig. 16 suggests that the $\mathcal{C}$-space for the open clusters has a higher dimensionality (about 1 more dimension) than for the nearby metal-rich field stars, both for the restricted α-element subspace and for the 13-element space. However, systematic differences between authors may contribute extra scatter to the open clusters abundances and the extra dimension could be spurious. To show that this is *not* the case, we have performed the intrinsic correlation
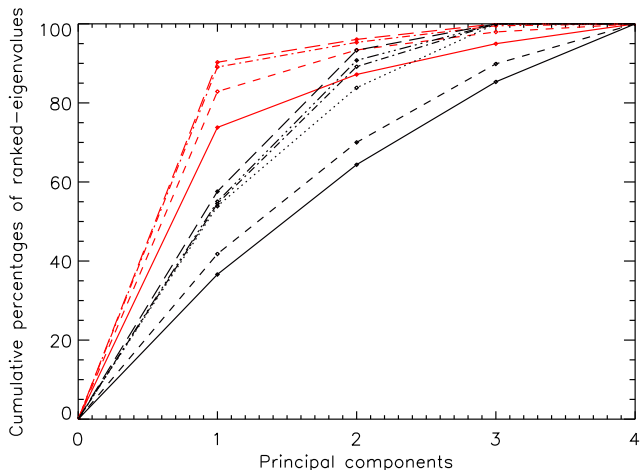
Figure 15. The ranked-eigenvalues cumulative percentages of the α-elements, comparing open clusters (*black lines*) with solar neighbourhood stars (*red lines*). The solid lines show the observed correlation. The dashed/broken lines show the cumulative percentages derived from the estimated intrinsic correlation matrix, assuming measurement uncertainty of 0.05, 0.07, 0.10, 0.15, and 0.20 dex respectively for open clusters, and 0.02, 0.03, and 0.10 dex respectively for solar neighbourhood stars.

Figure 16. The same as Fig. 15 but for 13 elements, and assuming measurement uncertainty of 0.05, 0.07, and 0.20 dex respectively for open clusters, and 0.03, 0.05, and 0.10 dex respectively for solar neighbourhood stars.

estimates as described in Section 4.2. The results are over-plotted in Fig. 15 and Fig. 16.

For example in Fig. 15, the *black dashed/broken lines* show the ranked-eigenvalues cumulative percentages derived from the estimated intrinsic correlation for the open clusters sample, assuming that the measurement uncertainty is 0.05 (short-dashed), 0.07, 0.1, 0.15, and 0.2 (long-dashed) dex, respectively. The *red dashed/broken lines* show the cumulative percentages derived from the estimated intrinsic correlation for the solar neighbourhood sample, assuming a measurement uncertainty of 0.02, 0.03 and 0.1 dex, respectively. Recall that for the estimated intrinsic correlation matrix, by definition, there is no residual contribution from measurement uncertainty. Therefore the cumulative percentages cutoff should be close to 100% instead of 85%.

The results show that our previous interpretation holds even if we make extreme assumptions about the measurement uncertainty of the two samples, such as assuming measurement uncertainty and systematic differences of 0.2 dex for open clusters and a very small measuring uncertainty of 0.03 dex for the solar neighbourhood stellar sample. The result is robust for the 4-dimensional α-element subspace. This is consistent with results in the literature that the α-element distribution may not be the same within and outside the solar neighborhood: for example the enhanced ratios of [Mg/Fe] and [Ti/Fe], but not for [Si/Fe] and [Ca/Fe] reported by Yong et al. (2005) at larger galactocentric distances. For the larger 13-dimensional $\mathcal{C}$-space, the result seems secure but is a little less robust because of the concerns described in Section 4.2 about deriving the intrinsic correlations for weakly correlated element pairs.

Our results are consistent with De Silva, Freeman & Bland-Hawthorn (2009) and Roederer (2009) that the larger regions of the Galaxy covered by the open cluster sample may show more independent variations of element abundances than the immediate solar
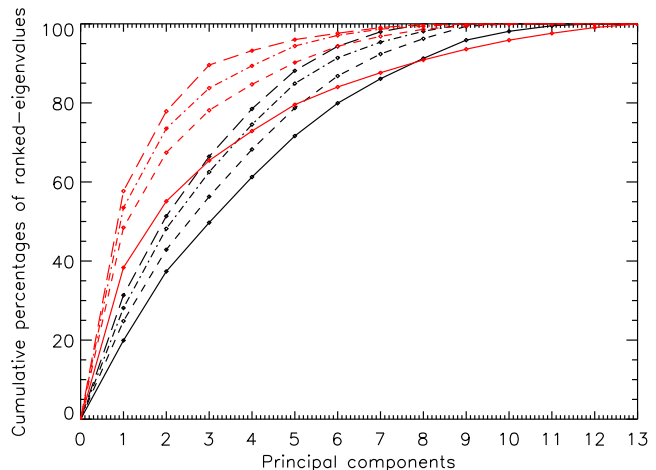
neighbourhood, even though we cannot readily identify the nature of the extra dimension of their $\mathcal{C}$-space. Although the solar neighbourhood appears chemically well mixed, as shown by the relative tightness of say the [α/Fe]–[Fe/H] relation, this may not prevail over larger volumes of the Galaxy.

5.4 Satellite galaxies

Now we investigate two samples of stars, one for the Fornax dSph galaxy and one for the LMC. These are of interest because they represent different evolutionary environments from the Milky Way, and also because older versions of these galaxies may be similar to the systems which merged with the Milky Way (see Searle & Zinn 1978).

In the previous section, we showed that the $\mathcal{C}$-space dimensionality might vary as a function of the survey volume. Therefore, robust comparisons can only be made if the survey volumes are about the same. Since our homogeneous satellite galaxy samples are restricted to a small survey volume $(1 - 2$ kpc in diameter[11]$)$, it seems appropriate to compare the $\mathcal{C}$-space results for the satellite galaxies and the solar neighbourhood, for which the surveys cover a similar volume. For each comparison, we restricted the solar neighbourhood stars to the same metallicity range as Fornax $(-1.2 < [\text{Fe/H}] < -0.6)$ and the LMC $(-1.0 < [\text{Fe/H}] < -0.5)$.

First we did a PCA analysis on the *n*-capture elements (Y, Ba, Nd, and Eu). The results is shown in Fig. 17. Fornax appears to show more $\mathcal{C}$-space dimensions than the Milky Way's solar neighbourhood from the observed correlation (solid lines). However, the extra dimension could be spurious if the measurement uncertainties for the dSph galaxy are larger. To examine this possibility, we estimated the in-

---

[11] Letarte's study covered 25′of the centre of Fornax (about 1 kpc, e.g. Rizzi et al. 2007). The survey of Pompeia et al. (2008) covered 2 kpc of the LMC inner disk
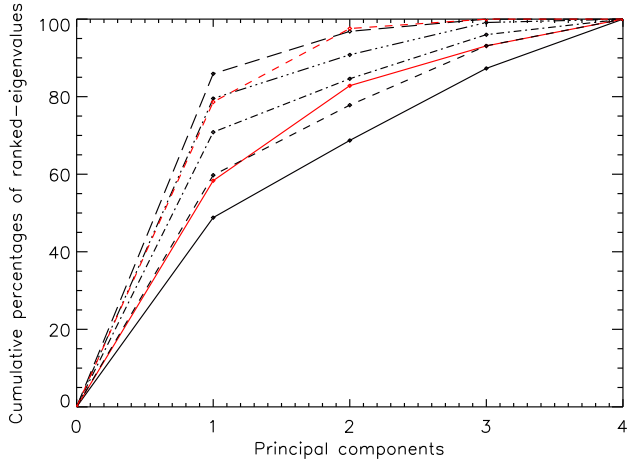
Figure 17. The *n*-capture element ranked-eigenvalues cumulative percentages comparing the Fornax dSph galaxy (*black lines*) with its solar neighbourhood counterparts (*red lines*) in the same metallicity range. The solid lines show the observed correlation. The dashed/broken lines show the cumulative percentages derived from the estimated intrinsic correlation matrix, assuming measurement uncertainty of 0.1, 0.13, 0.15, and 0.2 dex respectively for the dSph, and 0.07 dex for the solar neighbourhood counterparts.
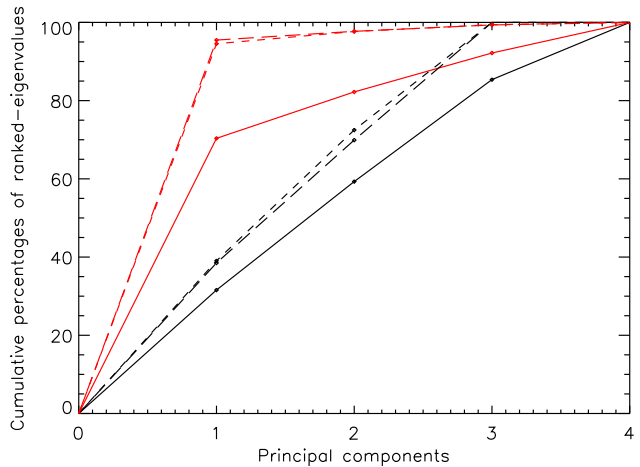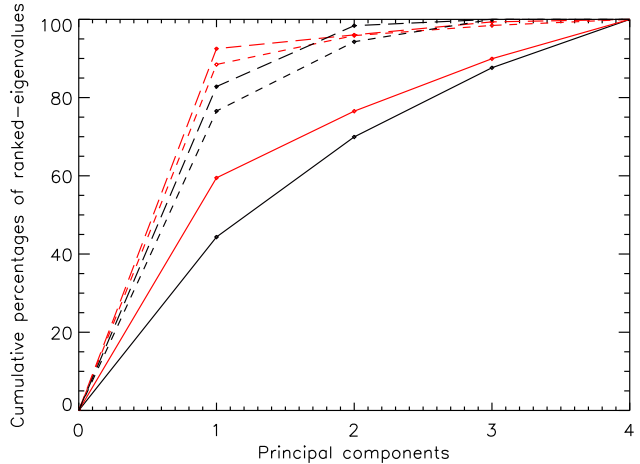

Figure 18. The composition of the first three normalized principal components for the *n*-capture elements of the Fornax dSph galaxy.

Figure 19. The ranked-eigenvalues cumulative percentages for the α-elements for the dwarf galaxies (*black lines*) and the solar neighbourhood (*red lines*) in the same metallicity range. The results for Fornax and the LMC are shown in the *upper plot* and *lower plot*, respectively. The solid lines show the observed correlation. The dashed/broken lines show the cumulative percentages derived from the estimated intrinsic correlation matrix, assuming measurement uncertainty of 0.1, 0.15 dex respectively for the dwarf galaxies and 0.05, 0.1 dex respectively for the solar neighbourhood counterparts.

trinsic correlation as before and the results is over-plotted in Fig. 17.

Note that Reddy et al. (2003, 2006) reported measurement uncertainty of $0.07 - 0.1$ dex for *n*-capture element abundances and Letarte et al. (2010) reported uncertainty of $0.15 - 0.25$ dex. In the comparison with Fornax, we took the measurement uncertainty σ for the solar neighbourhood sample to be 0.07 dex; if we were to choose $\sigma = 0.1$ dex, the PCA results for the solar neighbourhood sample in Fig. 17 would further deviate from the Fornax sample.

The simulation shows that, if we assume that the measurement uncertainty for Fornax to be $0.1 - 0.15$ dex, then the Fornax $\mathcal{C}$-space has more independent dimensions than the solar neighbourhood in the same metallicity range. How-

ever, we cannot exclude the possibility that the extra dimension is due to larger uncertainty for the Fornax sample. If we assume that the measurement uncertainty of the Fornax sample is larger than 0.2 dex, then Fornax's dimensionality is the about the same as the solar neighbourhood.

On the other hand, there is evidence that Letarte et al. (2010) may have overestimated their measurement uncertainty. For example, in their fig. 15, the [Eu/Fe]–[Fe/H] and [Nd/Fe]–[Fe/H] relations are quite tight, and the scatter of the points appears at odds with the error bar shown. Thus, we contend that the extra apparent dimension of the $\mathcal{C}$-space for Fornax is probably not due to a larger measurement uncertainty.

The principal components of the PCA analysis for the Fornax *n*-capture elements (Y, Ba, La, Nd, and Eu) are shown in Fig. 18. The first component consists of all elements in the same sign as before. This component represents a mechanism that produces all *n*-capture elements simulta-
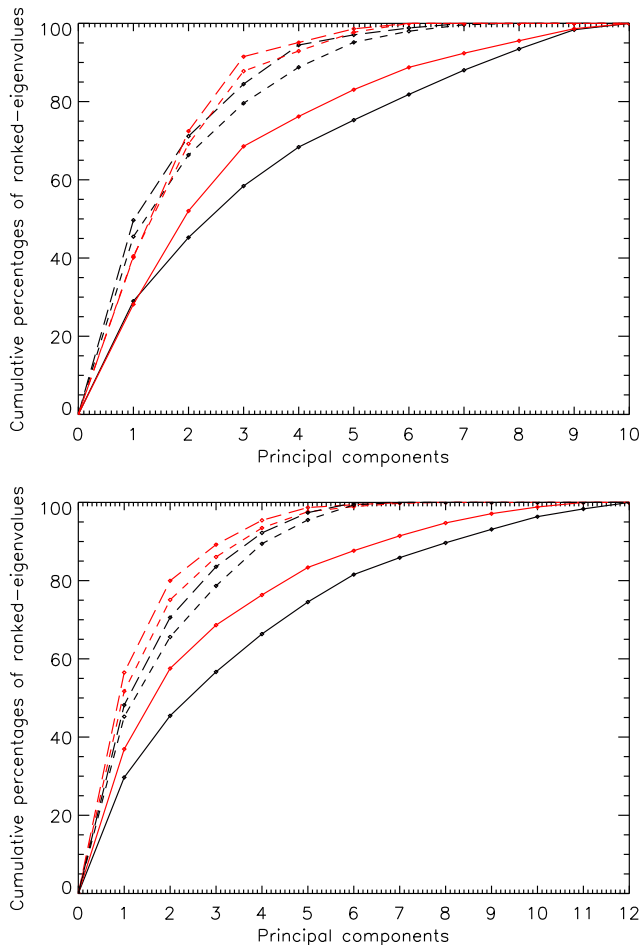
Figure 20. The same as Fig. 19 but for all elements in common.

5.5   Globular clusters

Now we study the Galactic globular clusters, which are all old objects and cover a large range in Galacto-centric radius. The formation process of the globular clusters remains poorly understood (for a review, see Gratton, Sneden & Carretta 2004), and we are interested to know what can be learned from their $\mathcal{C}$-space properties. We use the Pritzl et al. (2005) compilation which gives the mean abundances for each globular cluster.

We compared n-capture elements (Y, Ba, La, and Eu) for globular clusters with solar neighbourhood stars from Burris et al. (2000), and compared α-elements (Mg, Si, Ca, and Ti) and then all elements in common (Mg, Si, Ca, Ti, Y, Ba, and Eu) for globular clusters with solar neighbourhood stars from Fulbright (2000, 2002), restricting all samples to the same metallicity range: $-2.5 < [\text{Fe/H}] < -1.0$.

The ranked-eigenvalues cumulative percentages for the n-capture elements are plotted in the upper panel of Fig. 21. At first sight, globular clusters (*black solid line*) seem to have more $\mathcal{C}$-space dimensions than the solar neighbourhood (*red solid line*). As we have cautioned above, however, the sample of globular clusters is not homogeneous but rather a compilation from different authors. Therefore, we would expect more systematic uncertainty for the globular clusters sample. We estimated the intrinsic correlation with the same method used in the open clusters and dwarf galaxies study, and the results is over-plotted in the lower panel of Fig. 21. The results show that if we assume a 0.1 dex measurement uncertainty for solar neighbourhood stars and 0.2 dex measurement uncertainty for globular clusters, the independent dimensions of the n-capture element space are about the same. On the other hand, when we examined the dimensionality for the α-element subspace and the whole set of available elements in common to globular cluster and solar neighbourhood samples, the dimensionality of the globular cluster space seems to be larger than for the solar neighbourhood, as shown in the middle and lower panel of Fig. 21.

The sample of globular clusters is small (33 clusters in the metallicity range $-2.5 < [\text{Fe/H}] < -1$) and the impact of measurement uncertainty is significant. This can be seen in Fig. 21 in which the estimated intrinsic correlation for globular clusters shows a significant decrease of their independent $\mathcal{C}$-space dimensions. This illustrates that the sample size is crucial to determine a robust $\mathcal{C}$-space dimensionality. We also note that our study is somewhat limited in its scope: for example, we recall the Na/O anticorrelation which is seen in globular cluster stars (Carretta et al. 2009; Gratton et al. 2010) but not among the field stars, and note that Na and O are not part of this study.

In summary, within the limitations of the comparison, the globular clusters appear to have slightly larger chemical space dimensionality than the solar neighbourhood stars in the same metallicity range. It is difficult to know whether this result should be expected, because we do not yet have a good perspective on what globular clusters are and how (and where) their chemical evolution took place. A larger and homogeneous abundances survey covering more elements of globular clusters would be very desirable.

neously which probably involves the r-process. The second component shows an enhancement in light s-process elements (Y in this case) which could be the remnant of LEPP. And the third component shows enhancement in heavy s-process elements (Ba and La), which may be the contribution of the s-process in AGB stars operating at metal-poor environment. We postpone further discussion to Section 6.

For the LMC sample, Pompeia et al. (2008) did not measure Eu. As Eu is a crucial element to distinguish between pure r-process and s-process, we did not study the $\mathcal{C}$-subspace of the LMC n-capture elements. Instead, together with the Fornax dSph galaxy, we studied the ranked-eigenvalues cumulative percentages of α-elements (Mg, Si, Ca, Ti) and all elements measured in common: (1) for the LMC and the solar neighbourhood– Sc, Mg, Si, Ca, Ti, V, Cr, Co, Ni, Y, Zr, and Ba; (2) for Fornax and the solar neighbourhood– Mg, Si, Ca, Ti, Cr, Ni, Y, Ba, Nd, and Eu. The results are plotted in Fig. 19 and Fig. 20. Both plots show that dwarf galaxies have slightly more independent $\mathcal{C}$-space dimensions than the solar neighbourhood, even though we adopt a larger measurement uncertainty for the dwarf galaxies. This is consistent with our previous discussion and results. However, to draw a firmer conclusion, dwarf galaxy samples with measurement uncertainty smaller than in the current samples are urgently needed.
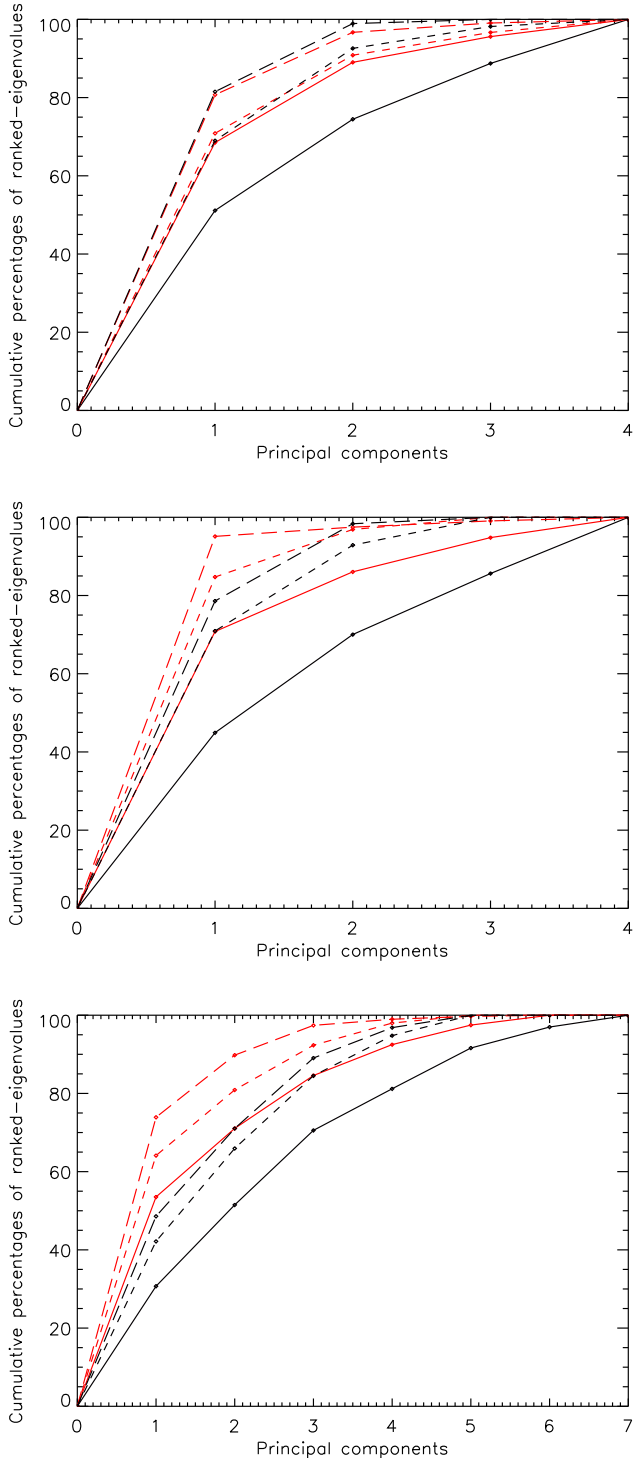
Figure 21. The ranked-eigenvalues cumulative percentages for the globular clusters (*black lines*) and the solar neighbourhood stars (*red lines*) in the same metallicity range: *n*-capture elements (*upper panel*), α-elements (*middle panel*) and all elements in common (*lower panel*). The solid lines show the observed correlation. The dashed/broken lines show the cumulative percentages derived from the estimated intrinsic correlation matrix, assuming measurement uncertainty of 0.1, 0.2 dex respectively for globular clusters, and 0.05, 0.1 dex respectively for the solar neighbourhood stars.

6   DISCUSSION

6.1   The *n*-capture elements subspace

6.1.1   *The r-process contribution*

Several authors including Honda et al. (2004a); Ivans et al. (2006) and Hayek et al. (2009) noted that the *r*-process production ratio for the *n*-capture elements in [X/Fe] is almost constant despite the large scatter in the abundance of each element. Therefore, in the $\mathcal{C}$-space, if the major contribution is from the *r*-process, one would expect the data cloud to form a tight straight line in the *n*-capture element space. It is pleasing to see that, at low metallicity where the *r*-process should be dominant (Truran 1981), the PCA shows a strong first component made up of all elements positively correlated. This *r*-process component contributed $\gtrsim 80\%$ of the data cloud variance for the metal-poor stars. This agrees with other conclusions from the literature.

6.1.2   *The overabundance of light s-process elements*

Observations have also reported more scatter in the ratio of light *s*-process elements to heavy *s*-process elements such as [Y/Ba] (Burris et al. 2000; Norris et al. 2001; Ivans et al. 2003; Aoki et al. 2005; Honda et al. 2007) at [Fe/H] < -3. As shown in Fig. 7, the light *s*-process elements seem to be overproduced at lower metallicity relative to the heavier elements. Various mechanisms have been proposed to explain this observation including the weak *r*-process (e.g. Izutani et al. 2009) and the CPR (Qian & Wasserburg 2007). Our PCA method clearly show the existence of this unknown source. For the low metallicity sample (Fig. 6), the second principal component is made up of light *s*-process elements anticorrelated with heavier *n*-capture elements. This suggests a second mechanism which either overproduces light *s*-process elements or overproduces heavier *n*-capture elements. From Fig. 7, the former interpretation is preferred.

Also, as shown in Table 2, the second primary source for light *s*-process elements contributes more as the metallicity decreases. In Fig. 5, the First Stars Survey sample shows slightly more dimensions than the Barklem's sample. This could be interpreted as a larger contribution from the LEPP since the First Stars Survey sample has a lower mean metallicity. This decreasing trend of LEPP with increasing metallicity suggests that the culprit for LEPP is restricted to low metallicity and/or massive progenitors. If this is due to the mass dependence, this is consistent with the proposed LEPP candidates– the weak *r*-process, CPR, collapsars or the truncated *r*-process. The relative contribution of the LEPP becomes smaller at higher metallicity and the production of *n*-capture elements is dominated by the main *r*-process and the *s*-process. The results in Table 2 also illustrates one of the strengths of the PCA method compared with the traditional [X/Fe]–[Fe/H] plots– one can quantify the contributions of different mechanisms.

6.1.3   *Low-mass AGB contribution*

At high metallicity, one would expect a contribution from low-mass AGB stars with initial mass range of $1.5 - 3$ M$_\odot$. Unlike LEPP, the *s*-process in low-mass AGB stars produces both light *s*-process elements and heavy *s*-process elements

such as Ba (see Herwig 2005; Käppeler et al. 2011), but not so much of the mostly r-process elements such as Nd and Eu relative to the r-process contribution (Arlandini et al. 1999).

In our PCA analysis, we find that the relative contribution from the first component (i.e. the r-process) decreases at higher metallicity (Fig. 5 and Fig. 12). We also find that the second component is made up of both light and heavy s-process elements (Fig. 11). Both are consistent with the s-process in low-mass AGB stars, which is dominant at high metallicity.

To show that this transition between high metallicity and low metallicity is real, we have also analyzed the intermediate metallicity sample taken from Burris et al. (2000). We separate the the sample into two extreme metallicity bin: $-2.7 < $ [Fe/H] $ < -2$ and $-1.5 < $ [Fe/H] $ < -1$. The former shows results that resembles Fig. 6 and the latter shows results that is similar to Fig. 11. This further confirms that the transition is indeed happening and occurs around $-2 \lesssim$ [Fe/H] $\lesssim -1.5$.

6.2 Satellite galaxies

In Section 5.4, we presented results for the dwarf galaxies samples in comparison with the solar neighbourhood stars in the same metallicity range. Recall that in this metallicity range, many stars in the solar neighbourhood are likely to be thick disk stars, and the belief is that the thick disk had a brief and intense star formation history (SFH) (see also Chiappini, Matteucci & Romano 2001; Kobayashi et al. 2006). On the other hand, the SFH is believed to be slower for dwarf galaxies (e.g. Weisz et al. 2011).

In our study, our PCA method results are consistent with the speculation that dwarf galaxies have slower SFH. We showed that dwarf galaxies seem to have more independent $\mathcal{C}$-space dimensions than the solar neighbourhood. This implies that the ISM in dwarf galaxies is impacted by a higher level of stochastic enrichment: the less vigorous history of enrichment events in the dwarf galaxies may allow more chemical traces to persist.

Now we will discuss the principal components for n-capture elements. Our PCA analysis shows two peculiarities of the n-capture elements in the dwarf galaxies. Firstly, the second principal component gives a significant contribution to the data cloud variance. We interpret the second principal component as the LEPP contribution (because it has a large contribution from light s-process elements). In contrast, for the solar neighbourhood stars in the same metallicity range, the LEPP contribution is hardly visible and only contributions from the r-process and the s-process in low-mass AGB stars are seen. This is consistent with the view that the SFH in dwarf galaxies is slower and therefore retains more chemical substructure from the previous generation of ejecta.

Secondly, the third principal component in this analysis was not seen before. It has heavy s-process elements (Ba, La) anticorrelated with the other n-capture elements, including the light s-process elements. Recall that previously at high metallicity, for n-capture elements in the Milky Way's solar neighbourhood, the s-process in low-mass AGB stars tends to produce both light and heavy s-process elements. If the third component represents the AGB contribution, one would expect it to have the same sign for the light and
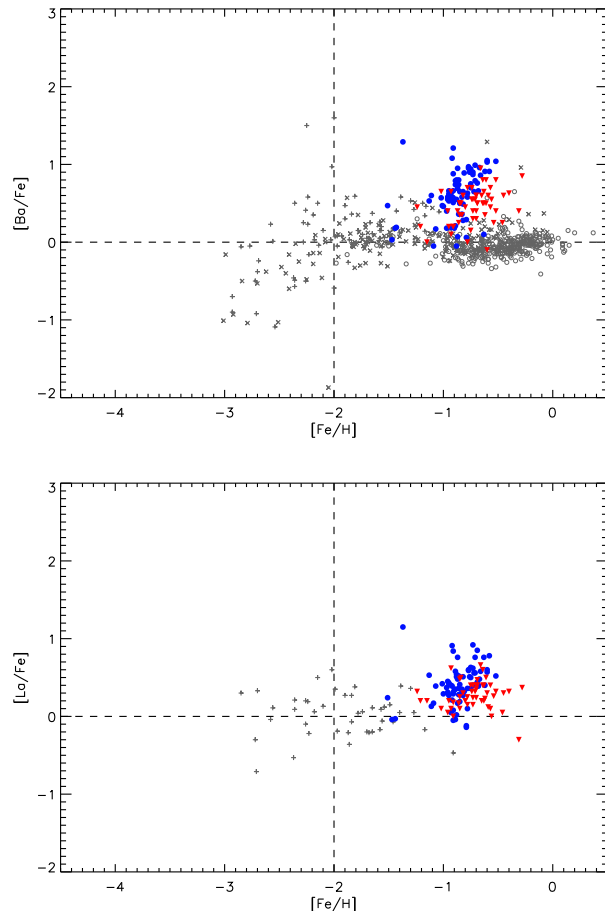

Figure 22. The [hs/Fe]–[Fe/H] of the solar neighbourhood, Fornax and the LMC stars. The symbols in *grey* represent the solar neighbourhood compilation from Burris et al. (2000) (*plus*), Fulbright (2000, 2002) (*crosses*) and Reddy et al. (2003, 2006) (*open circles*). The *blue filled circles* are Fornax stars from Letarte et al. (2010). The *red filled triangles* are the LMC stars from Pompeia et al. (2008). The two dwarf galaxies are overabundant in heavy s-process elements, compared to solar neighbourhood stars.

heavy s-process elements. However, this is not the case for the dwarf galaxies. This result is not totally unexpected. As shown in Fig. 22 and Fig. 23, observations for dwarf galaxies have shown over-abundances of [hs/Fe] at high metallicity but similar abundances or even under-abundances for the [ls/Fe].

This could be explained by the argument that the slower SFH in these galaxies allows low-mass AGB stars to contribute at lower metallicity: AGB stars working in a metal-poor environment have higher ratio of neutrons to Fe-peak seeds and will therefore preferentially produce heavier s-process elements (cf. Section 2).

If this is the case, we should see some similarities between solar neighbourhood CEMP-s stars and dwarf galaxies. It has been proposed that CEMP-s stars have suffered from AGB binary mass transfer at low metallicity. AGB stars which operate at low metallicity should, by the same argument, preferentially produce hs than ls. As CEMP-s stars are rare objects, we use compilation from the SAGA
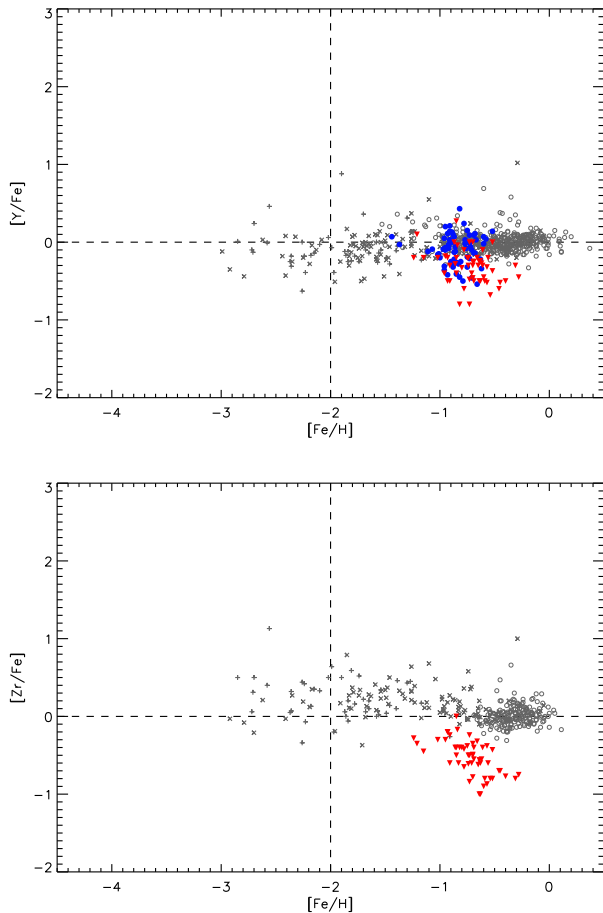
Figure 23. The [*ls*/Fe]–[Fe/H] of the solar neighbourhood, Fornax and the LMC stars. The symbols are the same as in Fig. 22. The two dwarf galaxies have similar or lower abundances of light *s*-process elements compared to solar neighbourhood stars.

Figure 24. The Y (*ls*) to Ba (*hs*) ratios as a function of [Fe/H] for solar neighbourhood not-CEMP stars, CEMP-s stars and dwarf galaxies stars. The solar neighbourhood not-CEMP stars symbols are the same as in Fig. 7. CEMP-s stars are plotted with *black open diamonds*. Fornax stars are plotted with *black crosses* and the LMC stars are plotted with *black open squares*.

database (Suda et al. 2008), and select stars with [C/Fe] > 1 and [Ba/Fe] > 1 (criteria from Beers & Christlieb 2005). For those stars with multiple measurements, we take the measurement from the study with the highest resolution and signal-to-noise ratio (for more details, see Suda et al. 2011). In Fig. 24, we show the [Y/Ba] for the solar neighbourhood not-CEMP stars, CEMP-s stars and dwarf galaxies stars. This figure is consistent with our speculation. Both solar neighbourhood CEMP-s stars and dwarf galaxies stars show lower [Y/Ba] than the solar neighbourhood not-CEMP stars. This suggests that *n*-capture elements for both CEMP-s stars and dwarf galaxies could have been produced by the *s*-process in AGB stars operating in a metal-poor environment.

Alternatively, the 'normal' abundances of light *s*-process element for dwarf galaxies can be explained if dwarf galaxies do not form many massive stars during their formation. This leads to a smaller contribution from LEPP and smaller production of light *s*-process elements in the formation phase of these galaxies. The under-production of light *s*-process elements from LEPP is then compensated by the onset of low-mass AGB star at low metallicities. The smaller LEPP contribution at low metallicity is not necessarily in-
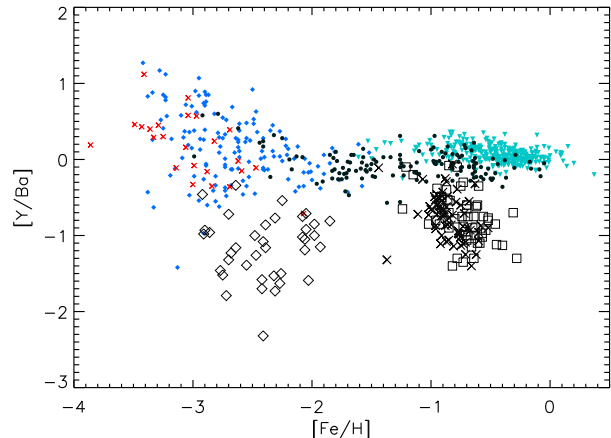
consistent with the PCA analysis results showing that LEPP contribution is still visible at the Fornax metallicity range (−1.2 < [Fe/H] < −0.8), as shown in Fig. 18, because there are also less *r*-process events to homogenize the ejecta.

6.3   All elements

6.3.1   Low metallicity

The interpretation of the PCA analysis of all elements at low metallicity as shown in Fig. 8 can be summarized as follows:

The first principal component could be the contribution of the *r*-process since it involves substantial production of all of our *n*-capture elements. There is also substantial *α*-element production together with the *r*-process production. This supports the view that the *r*-process occurs in core-collapse supernovae rather than neutron star mergers. It is generally agreed that the *α*-elements at low metallicity are produced in core-collapse supernovae. If this interpretation is correct, among *α*-elements, the composition of the first principal component as shown in Fig. 8 suggests that core-collapse supernovae that involve the *r*-process produce more heavier *α*-elements (Ca and Ti) than light *α*-elements (Mg). In the yields from Kobayashi et al. (2006), heavier *α*-elements are produced more in massive progenitors and/or with higher explosion energy. Our PCA results may suggest that the *r*-process is associated with more massive progenitors and/or high explosion energy.

The second and third principal components should be discussed together. The second component shows anticorrelation between *α*-elements and Fe-peak elements whereas the third component shows the production of *both α*-elements and Fe-peak elements. We suggest that this could be tentatively explained by the hypernovae-supernovae model (Umeda & Nomoto 2002, 2005; Kobayashi et al. 2006). The 'normal' supernovae produce more *α*-elements than iron (see also Woosley & Weaver 1995; Chieffi & Limongi

2002). However, the amounts are constrained from Galactic chemical evolution models (e.g. Timmes et al. 1995; Kobayashi et al. 2006). Theoretically, since the formation of neutron stars or blackholes is uncertain, the remnant mass is determined by a mass-cut. For low-mass supernovae ($\lesssim 20$ M$_\odot$), the ejected Fe mass is constrained to be about $0.07 M_\odot$ from the light-curves and spectra of individual supernovae. For massive supernovae ($\gtrsim 20$ M$_\odot$), the observational data is limited, but the significant production of Fe (together with α-elements) is supported by the light-curve and spectra. With higher explosion energy ($\gtrsim 10 \times 10^{51}$ erg), it is possible to produce enough amounts of heavy Fe-peak elements including Zn and Co to fit the observed abundances ratios (Kobayashi et al. 2006). Our PCA analysis seems to well-match this scenario: the third component could be interpreted as the contribution of hypernovae and the second component could be interpreted as the contribution from 'normal' supernovae.

We examine this idea by comparing the First Stars Survey sample which has a lower mean metallicity with the Barklem's sample. Since hypernovae are preferentially produced by more massive progenitors and therefore slightly shorter life span, Kobayashi & Nakasato (2011a) introduced metallicity dependence of the hypernovae fraction and argued that the contribution of hypernovae is larger at lower metallicity. However, the fraction of 'normal' supernovae and hypernovae is uncertain. To examine these ideas, we analyze the $\mathcal{C}$-subspace of Mg, Ca, Ti, Co, Ni, Zn. In this subspace, we interpret the component that has positive contribution from all elements to represent the hypernovae contribution and the component that shows anticorrelation between the α-elements (Mg, Ca, Ti) and heavy Fe-peak elements (Co, Ni, Zn) to be the usual supernovae contribution. With these assumptions, we found that for the First Stars Survey sample, the hypernovae component makes up 58.5% of the two contributions, whereas for the higher-metallicity Barklem's sample, the ratio decreases to 52.8%.

The α-elements contribution appears in all of the first three components. However the second and third components do not show a visible contribution from all n-capture elements. This might imply that, although r-process production is accompanied by α-elements yield, the reverse is not always true. If the r-process occurs in core collapse supernovae, then perhaps it can be triggered only if certain conditions or progenitor mass range are satisfied, as suggested by previous work (Wanajo & Ishimaru 2006).

The Al abundance was corrected by +0.6 dex for NLTE effects. If this correction is not made, the results remains largely unaltered, except that Al is contributing to the second component. Al is an light odd-Z element which comes from ^{22}Ne, and thus the yields depends on the metallicity of progenitor stars (e.g Kobayashi et al. 2006); see also Arnett (1971). Fig. 25 shows that, relative to light even-Z elements such as Mg, light odd-Z elements such as Na and Al are indeed metallicity-dependent and are suppressed at low metallicity. This shows that the production of light odd-Z elements is indeed metallicity-dependent and is suppressed at low metallcity. Therefore, although our NLTE-correction for Al is crude and does not depend on the stellar parameters, the correction is in the correct direction and is required to give a sensible PCA result– i.e. Al should not be con-
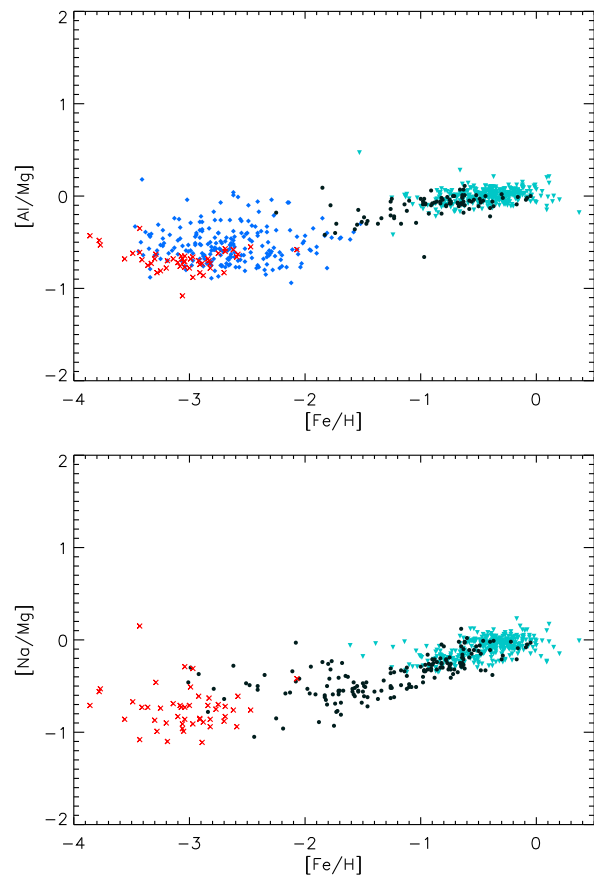

Figure 25. These plot show the ratio of light odd-Z elements (Al and Na) to light even-Z elements/α-elements (Mg) as a function of [Fe/H] of solar neighbourhood halo and disk stars. The symbols are the same as in Fig. 7. Al, Na and Mg abundances for First Stars Survey are NLTE-corrected according to Andrievsky et al. (2007, 2008, 2010) in these two plots. [Al/Fe] for Barklem's sample are NLTE-corrected by +0.6 dex. Both plots show an odd–even effect, i.e. light odd-Z elements are suppressed at low metallicity.

tributing to the core-collapse supernovae component at low metallicity.

The fourth Cr–Mn component could be interpreted as the contribution from incomplete Si-burning, favoring the idea that the production of Cr and Mn relies on a different mass cut region as proposed by Nakamura et al. (1999). If this interpretation is correct, this might suggest that the production ratio of elements synthesized in the incomplete Si-burning region are strongly correlated among themselves, but their production ratios are decoupled from the production of other elements such as Zn, Co, and Ni that are synthesized in the complete Si-burning region. This component should however be viewed with caution. Only Cr I is detected in Barklem's sample, and both Cr I and Mn are suspected to require significant NLTE correction (Bergemann & Gehren 2008; Bergemann & Cescutti 2010b). Therefore this dimension could be spurious. Further discussion of this component requires the availability of a large homogeneous NLTE-corrected data set.

6.3.2   High metallicity

Stars in this metallicity range have ages up to about 10 Gyrs, corresponding to more than a thousand generations of core-collapse supernovae (refer to fig. 8 in Kobayashi & Nomoto 2009). The PCA results for the high metallicity sample shown in Fig. 13 can be summarized as follows:

The first component is a strong component (40% of the total variance) which consists of all light odd-Z elements, α-elements and Fe-peak elements. This could be interpreted as the loss of chemical dimensions after many generations of short-time-scale core-collapse supernovae as shown in Karlsson & Gustafsson (2001) and Bland-Hawthorn et al. (2010). After many generations of enrichment, elements that can be formed by short-time-scale core-collapse supernovae would have been well mixed in the ISM. In short, the first component is just a measure of the metal-richness of the gas reservoirs from which the stars formed. Much information was lost after the first few generations of supernovae.

It is also interesting that Al contributes to the first component. This is different from the low metallicity sample for which the Al does not contribute. This is not unexpected. As discussed, Al is expected to be suppressed at low metallicity, but starts to contribute as metallicity increases.

The second component can be interpreted as an r-process contribution, since all n-capture elements contribute to this component in the same sign. Unlike the low-metallicity case, α-elements do not seem to contribute to this component. As discussed above, this may be due to the small parameter space of core-collapse supernovae that involve the r-process. As massive stars become fewer at high metallicity, most of the core-collapse supernovae will preferentially only produce α-elements but not r-process elements. These α-elements, as they continue to be produced and homogenized within the ISM, will only contribute to the hyperplane orthogonal to the n-capture elements. If this is the case, the PCA results will show a strong first component that includes supernovae ejecta and leaves the occasionally-produced n-captured elements to form a second component. This is also consistent with the results that the s-process in low-mass AGB stars starts to dominate n-capture elements production at high metallicity. Alternatively, the PCA results at high metallicity may imply that there is another r-process site that does not produce α-elements, such as neutron star mergers (e.g. Freiburghaus et al. 1999b; Rosswog et al. 1999)

The third component includes the contribution from light and heavy s-process elements, and as discussed in the n-capture elements section, could be interpreted as the low-mass AGB contribution. The appearance of α-elements constribution is unclear, but could be due to the envelope mass loss releasing the natal gas with heavier elements during the AGB phase.

6.4   Wider region of survey

As shown in Fig. 16, among the 12 elements that we investigated, compared to solar neighbourhood disk stars, we find about 1 more dimension for open clusters which were sampled in a larger Galactic volume. The extra dimension for a larger Galactic volume is not surprising. Although stars from the solar neighbourhood could come from different regions of the galaxies via radial mixing, stars from open clusters far from the solar radius are unlikely to have undergone the same mixing processes as the solar neighbourhood stars. The larger Galactic volume ($6 \lesssim r_G \lesssim 20$ kpc for open clusters, compared to $7.5 \lesssim r_G \lesssim 8.5$ kpc for the solar neighbourhood) is likely to retain more variation in $\mathcal{C}$-space.

6.5   K and Cu; *APOGEE*; the Ca-triplet region

In this section, we briefly discuss the effects of adding K and Cu to the list of elements, the dimensionality of the *APOGEE* $\mathcal{C}$-space, and the dimensionality of the $\mathcal{C}$-space accessible from observations of the Ca-triplet region of the spectrum.

We did not study the contribution from K and Cu, because these elements are absent from most of our samples except for Reddy's high metallicity sample. The *HERMES* bands are designed to include these two elements. We found that adding K and Cu to the high metallicity sample only slightly increases the effective dimensionality– from 6 (for 17 elements) to about 6.5 (for 19 elements)– with K and Cu contributing equally. However, we suspect that K and Cu may add more dimensions at low metallicity because we can still distinguish contributions from different types of core-collapse supernovae. Cu is an interesting element; it is an odd Z elements preferably produced with hypernovae (Kobayashi et al. 2006) and may have significant contribution from the weak s-process (like Ga & Ge, Pignatari et al. 2010). K may be affected by aspherical explosions (Maeda & Nomoto 2003) and also by the neutrino luminosity (Kobayashi et al. 2011d). The neutrino luminosity and weak s-process contribution could add promising extra dimensions at low metallicity.

In this study, we focused on the *HERMES* $\mathcal{C}$-space. The *APOGEE* high resolution near-infrared survey will measure the elements C, N, O, Na, Al, K, S, Mg, Si, Ca, Ti, V, Cr, Mn, Co, Ni but not n-capture elements, and we now estimate the dimensionality of this $\mathcal{C}$-space. We exclude C, N, O, Na as before because they may be subject to internal mixing. First, we considered the 9-dimensional $\mathcal{C}$-space (Al, Mg, Ca, Ti, V, Cr, Mn, Co, Ni) measured by both Barklem's low metallicity sample and Reddy's high metallicity sample. We found that the low metallicity sample has about 4.5 independent dimensions, whereas the high metallicity sample has 4 independent dimensions. This is consistent with our previous observation that n-capture elements contribute 2 independent dimensions at high metallicity– from r-process and s-process in low-mass AGB stars, but at low metallicity, n-capture elements only contribute $1 - 1.5$ dimensions.

In addition to these 9 dimensions, Reddy's sample also included Si, S and K. We found that adding Si alone does not add to the effective dimensionality, as Si is strongly correlated with the other α-elements. However, we found that including S and K will add around 1 additional dimension to the $\mathcal{C}$-space. Finally we summarize the effective dimensions of the $\mathcal{C}$-space (and subspaces) of various environments in Table 3.

The Ca-triplet region is widely used for large-scale spectroscopic surveys, so we also estimated the dimensionality of the $\mathcal{C}$-space defined by this region of the spectrum. For the metal-rich stars, the elements Al, Mg, Si, Ca, Ti and Ni define a space with 2 independent dimensions. For the

Table 3. Summary of effective dimensions in various environments and $\mathcal{C}$-subspace.

Environment	Elements considered	Eff. dims.	No. of elements
	Neutron-Capture Elements		
Solar neighbourhood halo stars	Y,Zr,Ba,La,Nd,Eu	$1-2$	6
Solar neighbourhood disk stars	Y,Zr,Ba,La,Nd,Eu	2	6
Fornax dSph galaxy	Y,Ba,La,Nd,Eu	3	5
	All Elements (HERMES)		
Solar neighbourhood halo stars	Al,Sc,Mg,Ca,Ti,V,Cr,Mn,Co,Ni,Zn,Y,Zr,Ba,La,Nd,Eu	6	17
Solar neighbourhood disk stars	Al,Sc,Mg,Si,Ca,Ti,V,Cr,Mn,Co,Ni,Zn,Y,Zr,Ba,Nd,Eu	6	17
Solar neighbourhood disk stars	Al,K,Sc,Mg,Si,Ca,Ti,V,Cr,Mn,Co,Ni,Cu,Zn,Y,Zr,Ba,Nd,Eu	$6-7$	19
MW open clusters	Al,Sc,Mg,Si,Ca,Ti,V,Cr,Co,Ni,Y,Ba,Nd	$6-7$	13
MW globular clusters	Mg,Si,Ca,Ti,Y,Ba,Eu	4	7
Fornax dSph galaxy	Mg,Si,Ca,Ti,Cr,Ni,Y,Ba,Nd,Eu	6	10
The Large Magellanic Cloud	Sc,Mg,Si,Ca,Ti,V,Cr,Co,Ni,Y,Zr,Ba	$6-7$	12
	All Elements (APOGEE)		
Solar neighbourhood halo stars	Al,Mg,Ca,Ti,V,Cr,Mn,Co,Ni	$4-5$	9
Solar neighbourhood disk stars	Al,Mg,Ca,Ti,V,Cr,Mn,Co,Ni	4	9
Solar neighbourhood disk stars	Al,K,S,Mg,Si,Ca,Ti,V,Cr,Mn,Co,Ni	5	12

* We assume measurement uncertainty of 0.1 dex and adopt 85% cut-off for the ranked-eigenvalues cumulative percentages to derive the effective dimensions. This might not hold for open/globular clusters and dwarf galaxies as they have higher measurement/systematic uncertainty. Therefore the dimensions for open/globular clusters and dwarf galaxies in this table should be regarded as upper limit

metal-poor stars, the elements Al, Mg, Ca, Ti and Ni define a space with 3 independent dimensions. Note that this estimate excludes possible extra dimensions from the light elements C, N and O.

7 CONCLUSION

In this paper, we used Principal Component Analysis to study and interpret stellar element abundances. We discussed a way to deal with the non-semi-definite positivity of the correlation matrix due to the incompleteness of the data.

We illustrated the power of this method by confirming the tight yield ratio of n-capture elements from the r-process, which confirms the universality of the r-process. We also traced the over-abundances of light s-process elements at low metallicity, and found that the relative contribution of the mechanism causing this over-abundances is decreasing with increasing metallicity. The site of LEPP is likely to be associated with massive stars, which is consistent with weak r-process and charged particle reactions. The method is also able to trace the s-process in low-mass AGB stars at high metallicity. The transition can be clearly shown using data with intermediate metallicity and occurs at $-2 \lesssim$ [Fe/H] $\lesssim -1.5$.

Our analysis also suggests that r-process production sites are accompanied by α-element production, favouring core-collapse supernovae as the r-process site. This can be seen at low metallicity but is no longer visible at high metallicity, which may be due to the small parameter space of the r-process site, or the inclusion of another site such as neutron star mergers that does not produce α-elements.

The analysis indicates two types of core-collapse supernovae: one produces mainly α-elements, the other produces both α-elements and Fe-peak elements with a large enhance-

ment of heavy Fe-peak elements. This is consistent with hypernovae. We find that the contribution of hypernovae is larger at lower metallicity, which may be important to understand the physics of hypernovae. We also discussed the Cr and Mn production from the incomplete Si burning region, and the metallicity dependence of odd-Z element (Al in our analysis), but these may be affected by NLTE effect.

To estimate the dimensionality of the C-space available to HERMES, we chose to work in [X/Fe] space, because all of the elements are highly correlated with [Fe/H]. We based our estimate of the dimensionality of [X/Fe] space on the simulations described in Section 4.1.3, indicating that the eigenvectors contributing to the first 85% of the cumulative sum of the ranked eigenvectors in [X/Fe] space are probably real. Due to the paucity of data and the potential contribution from internal mixing, we did not include Li, C, N, O and Na in our study. The PCA analysis of 17 elements shows that the chemical [X/Fe] space has about 6 dimensions both at high metallicity and low metallicity in the solar neighbourhood. We expect to have about 1 to 2 further independent dimensions of the [X/Fe] space from these elements that we excluded. Adding K and Cu for the HERMES sample provides about another half of a dimension for the [X/Fe] space. Our ultimate goal here is to evaluate the dimensionality of the 25-element HERMES space. Including the further dimension from [Fe/H] itself would give about 8 to 9 independent dimensions for the HERMES chemical [X/Fe]+[Fe/H] space.

Although the number of $\mathcal{C}$-space dimensions is similar at high and low metallicity, the interpretations of the principal components are very different in these two cases. For example, at high metallicity, we have an extra 'birth-imprint' from low-mass AGB stars. This extra contribution compensates the dimension loss due to homogenization of the core-collapse supernovae ejecta.

Our analysis indicates that dwarf galaxies retain more

chemical inhomogeneity than the Milky Way disk. The analysis also suggests that the *s*-process in AGB stars of dwarf galaxies preferentially produce heavier *n*-capture elements which is similar to the solar neighbourhood CEMP-s stars. Both of these effects may be produced by AGB stars working in metal-poor environment. These findings are consistent with the view that dwarf galaxies have had a slower SFH than the Milky Way disk.

The $\mathcal{C}$-space for Galactic globular clusters and open clusters, which span a large Galactic volume, appear to define a slightly higher-dimensional $\mathcal{C}$ space than that for the solar neighbourhood. A large homogeneous sample of abundances in globular clusters and open clusters is highly desirable to confirm this conclusion.

We showed that PCA can shed some insight on the underlying nucleosynthesis mechanisms. At low metallicity where the chemical space is less homogenized, this method could be exploited to put constraints on the number of mechanisms or, more precisely, on the number of parameters that govern Galactic chemical evolution in different environments.

For our future observational work in chemical tagging, the PCA method could combine with group-finding algorithms (Agrawal et al. 1998; Sharma & Johnston 2009) to increase the efficiency of the algorithms for finding substructure in chemical space. The group-finding algorithms are sensitive to redundant dimensions which decrease the density of the clusters in chemical space. We expect that our PCA method can reduce the effective dimensionality of the chemical space without compromising any useful information.

ACKNOWLEDGMENTS

REFERENCES

Agrawal R., Gehrke J., Gunopulos D., Raghavan P., 1998, Automatic Subspace Clustering of High Dimensional Data for Data Mining Applications
Akerman C. J., Carigi L., Nissen P. E., Pettini M., Asplund M., 2004, A&A, 414, 931
Allende Prieto C., Beers T. C., Wilhelm R., Newberg H. J., Rockosi C. M., Yanny B., Lee Y. S., 2006, ApJ, 636, 804
Allende Prieto C. et al., 2008, Astronomische Nachrichten, 329, 1018
Anders E., Grevesse N., 1989, Geochim. Cosmochim. Acta, 53, 197
Andreuzzi G., Bragaglia A., Tosi M., Marconi G., 2011, MNRAS, 412, 1265
Andrievsky S. M., Spite M., Korotin S. A., Spite F., Bonifacio P., Cayrel R., Hill V., François P., 2007, A&A, 464, 1081
Andrievsky S. M., Spite M., Korotin S. A., Spite F., Bonifacio P., Cayrel R., Hill V., François P., 2008, A&A, 481, 481
Andrievsky S. M., Spite M., Korotin S. A., Spite F., François P., Bonifacio P., Cayrel R., Hill V., 2009, A&A, 494, 1083
Andrievsky S. M., Spite M., Korotin S. A., Spite F., Bonifacio P., Cayrel R., François P., 2010, A&A, 509, A88
Andrievsky S. M., Spite F., Korotin S. A., Francçois P., Spite M., Bonifacio P., Cayrel R., Hill V., 2011, A&A, 530, A105
Aoki W., Norris J. E., Ryan S. G., Beers T. C., Ando H., 2002, ApJ, 576, L141
Aoki W. et al., 2005, ApJ, 632, 611
Aoki W., Beers T. C., Christlieb N., Norris J. E., Ryan S. G., Tsangarides S., 2007, ApJ, 655, 492
Aoki W. et al., 2008, ApJ, 678, 1351
Aoki W. et al., 2009, A&A, 502, 569
Arnett W. D., 1971, ApJ, 166, 153
Arlandini C., Käppeler F., Wisshak K., Gallino R., Lugaro M., Busso M., Straniero O., 1999, ApJ, 525, 886
Arnone E., Ryan S. G., Argast D., Norris J. E., Beers T. C., 2005, A&A, 430, 507
Asplund M., 2005b, ARA&A, 43, 481
Asplund M., Nissen P. E., Lambert D. L., Primas F., Smith V. V., 2005a, in Hill V., François P., Primas F., eds, Proc. IAU Symp. 228, p. 53
Asplund M., Grevesse N., Sauval A. J., 2005c, in Barnes T. G., Bash F. N., eds, Proc. ASP Conf. 336, p. 25
Asplund M., Lambert D. L., Nissen P. E., Primas F., Smith V. V., 2006, ApJ, 644, 229
Asplund M., Grevesse N., Sauval A. J., Scoot P., 2009, ARA&A, 47, 481
Audouze J., Silk J., 1995, ApJ, 451, L49
Banerjee P., Haxton W. C., Qian Y. Z., 2011, Phys. Rev. Lett., 106, 201104
Barklem P. S. et al., 2005, A&A, 439, 129
Barklem P. S., Belyaev A. K., Guitou M., Feautrier N., Gadéa F. X., Spielfiedel A., 2011, A&A, 530, A94
Baumueller D., Gehren T., 1997, A&A, 325, 1088
Baumueller D., Butler K., Gehren T., 1998, A&A, 338, 637
Beers T. C., Christlieb N., 2005, ARA&A, 43, 531
Bensby T., Feltzing S., Lundström I., 2003, A&A, 410, 527
Bensby T., Feltzing S., Lundström I., Ilyin T., 2005, A&A, 433, 185
Bergemann M., Cescutti G., 2010b, A&A, 522, A9
Bergemann M., Gehren T., 2008, A&A, 492, 823
Bergemann M., Pickering J. C., Gehren T., 2010a, MNRAS, 401, 1334
Bisterzo S., Gallino R., Straniero O., Aoki W., 2009 PASA, 26, 314
Bisterzo S., Gallino R., Straniero O., Cristallo S., Käppeler F., 2010, MNRAS, 404, 1529
Bland-Hawthorn J., Freeman K. C., 2004, PASA, 21, 110
Bland-Hawthorn J., Karlsson T., Sharma S., Krumholz M., Silk J., 2010, ApJ, 721, 582
Bonifacio P. et al., 2007, A&A, 462, 851
Bonifacio P. et al., 2009, A&A, 501, 519
Boyd R., Famiano M., Meyer B., Motizuki Y., Kajino T.,

Roederer I., 2011, arXiv:1111.3985

Bubar E. J., King J. R., 2010, AJ, 140, 293

Burris D. L., Pilachowski C. A., Armandroff T. E., Sneden C., Cowan J. J., Roe H., 2000, ApJ, 544, 302

Busso M., Gallino R., Wasserburg G. J., 1999, ARA&A, 37, 239

Busso M., Gallino R., Lambert D. L., Travaglio C., Smith V. V., 2001, ApJ, 557, 802

Caffau E., Bonifacio P., Faraggiana R., Sbordone L., 2005a, A&A, 436, L9

Caffau E., Bonifacio P., Faraggiana R., François P., Gratton R. G., Barbieri M., 2005b, A&A, 441, 533

Carrera R., Pancino E., 2011, A&A, 535, A30

Carretta E., Gratton E., Cohen J. G., Beers T. C., Christlieb N., 2002, AJ, 124, 481

Carretta E., Bragaglia A., Gratton R., Lucatello S., 2009, A&A, 505, 139

Cayrel R. et al., 2004, A&A, 416, 1117

Chen Y. Q., Nissen P. E., Zhao G., Asplund M., 2002, A&A, 390, 225

Chen Y. Q., Zhao G., Nissen P. E., Bai G. S., Qiu H. M., 2003, ApJ, 591, 925

Chen Y. Q., Nissen P. E., Zhao G., 2004, A&A, 425, 697

Chernoff D. F., Weinberg M. D., 1990, ApJ, 351, 121

Chiappini C., Matteucci F., Romano D., 2001, ApJ, 554, 1044

Chieffi A., Limongi M., 2002, ApJ, 577, 281

Cohen J. G. et al., 2004, ApJ, 612, 1107

Cohen J. G. et al., 2006, AJ, 132, 137

Cowan J. J. et al., 2002, ApJ, 572, 861

Cowan J. J. et al., 2005, ApJ, 627, 238

Cristallo S., Straniero O., Gallino R., Piersanti L., Domínguez I., Lederer M. T., 2009, ApJ, 696, 797

Depagne E. et al., 2002, A&A, 390, 187

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., Bland-Hawthorn J., Asplund M., Bessell M. S., 2007, AJ, 133, 694

De Silva G. M., Freeman K. C., Bland-Hawthorn J., 2009, PASA, 26, 11

Edvardsson B., Andersen J., Gustafsson B., Lambert D. L., Nissen P. E., Tomkin J., 1993, A&A, 275, 101

François P. et al., 2007, A&A, 476, 935

Frebel A., Simon J. D., Geha M., Willman B., 2010, ApJ, 708, 560

Frebel A., Kirby E. N., Simon J. D., 2010, Nature, 464, 72

Freeman K. C., Bland-Hawthorn J., 2002, ARA&A, 40, 487

Freiburghaus C., Rembges J. F., Rauscher T., Kolbe E., Thielemann F. K., Kratz K. L., Pfeiffer B., Cowan J. J., 1999a, ApJ, 516, 381

Freiburghaus C., Rosswog S., Thielemann F. K., 1999b, ApJ, 525, L121

Friel E. D., Jacobson H. R., Pilachowski C. A., 2010, AJ, 139, 1942

Fuhrmann K., 2011, MNRAS, 414, 2893

Fulbright J. P., 2000, AJ, 120, 1841

Fulbright J. P., 2002, AJ, 123, 404

Fulbright J. P., McWilliam A., Rich R. M., 2007, ApJ, 661, 1152

Gallino R., Arlandini C., Busso M., Lugaro M., Travaglio

C., Straniero O., Chieffi A., Limongi M., 1998, ApJ, 497, 388

García Pérez A. E., Asplund M., Primas F., Nissen P. E., Gustafsson B., 2006, A&A, 451, 621

Geisler D., Smith V. V., Wallerstein G., Gonzalez G., Charbonnel C., 2005, AJ, 129, 1428

Goswami A., Aoki W., Beers T. C., Christlieb N., Norris J. E., Ryan S. G., Tsangarides S., 2006, MNRAS, 372, 343

Gratton R. G., Carretta E., Eriksson K., Gustafsson B., 1999, A&A, 350, 955

Gratton R. G., Carretta E., Claudi R., Lucatello S., Barbieri M., 2003a, A&A, 404, 187

Gratton R. G., Carretta E., Desidera S., Lucatello S., Mazzei P., Barbieri M., 2003b, A&A, 406, 131

Gratton R. G., Sneden C., Carretta E., 2004, ARA&A, 42, 385

Gratton R. G., Carretta E., Bragaglia A., Lucatello S., D'Orazi V., 2010, The Messenger, 142, 28

Grevesse N., Sauval A. J., 1998, in Fröhlich C., Huber M. C. E., Solanki S. K., von Steiger R., eds, Solar Composition and Its Evolution– From Core to Corona, p. 161

Hanson R. B., Sneden C., Kraft R. P., Fulbright J., 1998, AJ, 116, 1286

Hayek W. et al., 2009, A&A, 504, 511

Heger A., Woosley S. E., 2010, ApJ, 724, 341

Herwig F., 2005, ARA&A, 43, 435

Hill V. et al., 2002, A&A, 387, 560

Hollowell D., Iben Jr. I., 1988, ApJ, 333, L25

Honda S. et al., 2004, ApJS, 152, 113

Honda S., Aoki W., Kajino T., Ando H., Beers T. C., Izumiura H., Sadakane K., Takada-Hidai M., 2004, ApJ, 607, 474

Honda S., Aoki W., Ishimaru Y., Wanajo S., 2007, ApJ, 666, 1189

Iben Jr. I., 1975, ApJ, 196, 525

Iben Jr. I., Renzini A., 1982, ApJ, 263, L23

Israelian G., Rebolo R., García López R. J., Bonifacio P., Molaro P., Basri G., Shchukina N., 2001, ApJ, 551, 833

Ivans I. I., Sneden C., James C. R., Preston G. W., Fulbright J. P., Höflich P. A., Carney B. W., Wheeler J. C., 2003, ApJ, 592, 906

Ivans I. I., Simmerer J., Sneden C., Lawler J. E., Cowan J. J., Gallino R., Bisterzo S., 2006, ApJ, 645, 613

Izutani N., Umenda H., 2010, ApJ, 720, L1

Izutani N., Umeda H., Tominaga N., 2009, ApJ, 692, 1517

Jacobson H. R., Friel E. D., Pilachowski C. A., 2011, AJ, 141, 58

Johnson J. A., Bolte M., 2002a, ApJ, 579, 616

Johnson J. A., Bolte M., 2002b, ApJ, 579, L87

Johnson J. A., Bolte M., 2004, ApJ, 605, 462

Jonsell K., Edvardsson B., Gustafsson B., Magain P., Nissen P. E., Asplund M., 2005, A&A, 440, 321

Käppeler F., Gallino R., Busso M., Picchio G., Raiteri C. M., 1990, ApJ, 354, 630

Käppeler F., Gallino R., Bisterzo S., Aoki W., 2011, Rev. Mod. Phys., 83, 157

Karakas A. I., 2010, MNRAS, 403, 1413

Karakas A. I., Lattanzio J. C., 2003, PASA, 20, 279

Karlsson T., 2006, ApJ, 641, L41

Karlsson T., Gustafsson B., 2001, A&A, 379, 461

Kirby E. N., Simon J. D., Geha M., Guhathakurta P., Frebel A., 2008, ApJ, 685, L43

Kobayashi C., Nakasato N., 2011a, ApJ, 729, 16

Kobayashi C., Nomoto K., 2009, ApJ, 707, 1466

Kobayashi C., Umeda H., Nomoto K., Tominaga N., Ohkubo T., 2006, ApJ, 653, 1145

Kobayashi C., Tominaga N., Nomoto K., 2011b, ApJ, 730, L14

Kobayashi C., Karakas A. I., Umeda H., 2011c, MNRAS, 414, 3231

Kobayashi C., Izutani N., Karakas A. I., Yoshida I., Yong D., Umeda H., 2011d, ApJ, 739, L57

Koch A., McWilliam A., Grebel E. K., Zucker D. B., Belokurov V., 2008, ApJ, 688, L13

Korn A. J., Shi J., Gehren T., 2003, A&A, 407, 691

Krystek M., Anton M., 2007, Measurement Science and Technology, 18, 3438

Lai D. K., Johnson J. A., Bolte M., Lucatello S., 2007, ApJ, 667, 1185

Lai D. K., Bolte M., Johnson J. A., Lucatello S., Heger A., Woosley S. E., 2008, ApJ, 681, 1524

Langer N., Heger A., Wellstein S., Herwig F., 1999, A&A, 346, L37

Letarte B. et al., 2010, A&A, 523, A17

Lodders K., Palme H., Gail H. P., 2009, in Trümper J. E., ed, Abundances of the Elements in the Solar System. Springer-Verlag, Berlin, p. 560

Lucatello S., Gratton R., Cohen J. G., Beers T. C., Christlieb N., Carretta E., Ramírez S., 2003, AJ, 125, 875

Maeda K., Nomoto K., 2003, ApJ, 598, 1163

Marigo P., 2001, A&A, 370, 194

Mashokina L., Gehren T., 2001, A&A, 376, 232

Masseron T., Johnson J. A., Plez B., van Eck S., Primas F., Goriely S., Jorissen A., 2010, A&A, 509, A93

Mateo M. L., 1998, ARA&A, 36, 435

McClure R. D., 1984, ApJ, 280, L31

McWilliam A., 1998, AJ, 115, 1640

McWilliam A., Rich R. M., 1994, ApJS, 91, 749

McWilliam A., Preston G. W., Sneden C., Searle L., 1995, AJ, 109, 2757

McWilliam A., Rich R. M., Smecker-Hane T. A., 2003, ApJ, 592, L21

Meynet G., Maeder A., 2002, A&A, 390, 561

Meynet G., Ekström S., Maeder A., 2006, A&A, 447, 623

Mikolaitis Š, Tautvaišienė G., Gratton R., Bragaglia A., Carretta E., 2010, MNRAS, 407, 1866

Mikolaitis Š, Tautvaišienė G., Gratton R., Bragaglia A., Carretta E., 2011, MNRAS, 413, 2199

Nakamura T., Umeda H., Nomoto K., Thielemann F. K., Burrows A., 1999, ApJ, 517, 193

Neves V., Santos N. C., Sousa S. G., Correia A. C. M., Israelian G., 2009, A&A, 497, 563

Nissen P. E., Schuster W. J., 1997, A&A, 326, 751

Nissen P. E., Schuster W. J., 2010, A&A, 511, L10

Nissen P. E., Primas F., Asplund M., Lambert D. L., 2002, A&A, 390, 235

Nissen P. E., Chen Y. Q., Asplund M., Pettini M., 2004, A&A, 415, 993

Nissen P. E., Asplund M., Fabbian D., Kerber F., Käufl H. U., Pettini M., 2007a, The Messenger, 128, 38

Nissen P. E., Akerman C., Asplund M., Fabbian D., Kerber F., Käufl H. U., Pettini M., 2007b, A&A, 469, 319

Nomoto K., Maeda K., Mazzali P. A., Umeda H., Deng J., Iwamoto K., 2004, in Fryer C. L., ed, Stellar Collapse.

Astrophysics and Space Science, Kluwer, p. 277

Nordström B. et al., 2004, A&A, 418, 989

Norris J. E., Ryan S. G., Beers T. C., Deliyannis C. P., 1997, ApJ, 485, 370

Norris J. E., Ryan S. G., Beers T. C., 2001, ApJ, 561, 1034

Odenkirchen M., et al. 2003, AJ, 126, 2385

Pancino E., Carrera R., Rossetti E., Gallart C., 2010, A&A, 511, A56

Pignatari M., Gallino R., Heil M., Wiescher M., Käppeler F., Herwig F., Bisterzo S., 2010, ApJ, 710, 1557

Pompeia L. et al., 2008, A&A, 480, 379

Prantzos N., Hashimoto M., Nomoto K., 1990, A&A, 234, 211

Preston G. W., Sneden C., 2000, AJ, 120, 1014

Preston G. W., Sneden C., Thompson I. B., Shectman S. A., Burley G. S., 2006, AJ, 132, 85

Pritzl B. J., Venn K. A., Irwin M., 2005, AJ, 130, 2140

Prochaska J. X., Naumov S. O., Carney B. W., McWilliam A., Wolfe A. M., 2000, AJ, 120, 2513

Pruet J., Thompson T. A., Hoffman R. D., 2004, AJ, 606, 1006

Qian Y. Z., Wasserburg G. J., 2007, Phys. Rep., 442, 237

Qian Y. Z., Wasserburg G. J., 2008, ApJ, 687, 272

Rebonato R., Jäckel P., 1999, The Most General Methodology to Create a Valid Correlation Matrix for Risk Management and Option Pricing Purposes

Reddy B. E., Tomkin J., Lambert D. L., Allende Prieto C., 2003, MNRAS, 340, 304

Reddy B. E., Lambert D. L., Allende Prieto C., 2006, MNRAS, 367, 1329

Reimers D., 1977, A&A, 57, 395

Rizzi L., Held E. V., Saviane I., Tully R. B., Gullieuszik M., 2007, MNRAS, 380, 1255

Roederer I. U., 2009, AJ, 137, 272

Roederer I. U., Sneden C., 2011, AJ, 142, 22

Roederer I. U., Cowan J. J., Karakas A. I., Kratz K. L., Lugaro M., Simmerer J., Farouqi K., Sneden C., 2010, ApJ, 724, 975

Rosswog S., Liebendörfer M., Thielemann F. K., Davies M. B., Benz W., Piran T., 1999, A&A, 341, 499

Ruchti G. R. et al., 2010, ApJ, 721, L92

Ryan S. G., Norris J. E., Beers T. C., 1996, ApJ, 471, 254

Sbordone L., Bonifacio P., Buonanno R., Marconi G., Monaco L., Zaggia S., 2007, A&A, 465, 815

Sbordone L. et al., 2010, A&A, 522, A26

Searle L., Zinn R., 1978, ApJ, 225, 357

Sharma S., Johnston K. V., 2009, ApJ, 703, 1061

Shetrone M. D., Côté P., Sargent W. L. W., 2001, ApJ, 548, 592

Shetrone M. D., Venn K. A., Tolstoy E., Primas F., Hill V., Kaufer A., 2003, AJ, 125, 684

Siess L., 2006, A&A, 448, 717

Siess L., 2007, A&A, 476, 893

Siess L., 2010, A&A, 512, A10

Sivarani T. et al., 2004, A&A, 413, 1073

Sneden C., McWilliam A., Preston G. W., Cowan J. J., Burris D. L., Armosky B. J., 1996, ApJ, 467, 819

Sousa S. G. et al., 2008, A&A, 487, 373

Spite M. et al., 2005, A&A, 430, 655

Stancliffe R. J., 2009, MNRAS, 394, 1051

Stancliffe R. J., Glebbeek E., 2008, MNRAS, 389, 1828

Stancliffe R. J., Jeffery C. S., 2007, MNRAS, 375, 1280

Steinmetz M. et al., 2006, AJ, 132, 1645

Stephens A., Boesgaard A. M., 2002, AJ, 123, 1647

Suda T. et al., 2008, PASJ, 60, 1159

Suda T., Yamada S., Katsuta Y., Komiya Y., Ishizuka C., Aoki W., Fujimoto M. Y., 2011, MNRAS, 412, 843

Tafelmeyer M. et al., 2010, A&A, 524, A58

Takeda Y., Zhao G., Takada-Hidai M., Chen Y. Q., Saito Y. J., Zhang H. W., 2003, Chinese J. Astron. Astrophys., 3, 316

Timmes F. X., Woosley S. E., Weaver T. A., 1995, ApJS, 98, 617

Tolstoy E., Hill V., Tosi M., 2009, ARA&A, 47, 371

Travaglio C., Gallino R., Arnone E., Cowan J., Jordan F., Sneden C., 2004, ApJ, 601, 864

Truran J. W., 1981, A&A, 97, 391

Truran J. W., Iben Jr. I., 1977, ApJ, 216, 797

Truran J. W., Cowan J. J., Pilachowski C. A., Sneden C., 2002, PASP, 114, 1293

Tsujimoto T., Shigeyama T., 2003, ApJ, 584, L87

Umeda H., Nomoto K., 2002, ApJ, 565, 385

Umeda H., Nomoto K., 2003, Nature, 422, 871

Umeda H., Nomoto K., 2005, ApJ, 619, 427

Vassiliadis E., Wood P. R., 1993, ApJ, 413, 641

Venn K. A., Hill V. M., 2008, The Messenger, 134, 23

Venn K. A., Irwin M., Shetrone M. D., Tout C. A., Hill V., Tolstoy E., 2004, AJ, 128, 1177

Wanajo S., Ishimaru Y., 2006, Nuclear Physics A, 777, 676

Wanajo S., Nomoto K., Janka H. T., Kitaura F. S., Müller B., 2009, ApJ, 695, 208

Weisz D. R. et al., 2011, ApJ, 739, 5

Westin J., Sneden C., Gustafsson B., Cowan J. J., 2000, ApJ, 530, 783

Woosley S. E., Weaver T. A., 1995, ApJS, 101, 181

Woosley S. E., Wilson J. R., Mathews G. J., Hoffman R. D., Meyer B. S., 1994, ApJ, 433, 229

Yong D., Carney B. W., Teixera de Almeida M. L., 2005, AJ, 130, 597

APPENDIX A: PRINCIPAL COMPONENT ANALYSIS

Consider an n dimensional space with random variables $(X_1, \ldots, X_n)$. In this section, we will re-derive and show that the normalized eigenvector corresponding to the largest eigenvalue of the correlation matrix is the direction where the standardized (i.e. normalized and mean-shifted) data cloud has the largest variance. Suppose we have m data points such that the random variables value of the q-th data point is $(X_{1,q}, \ldots, X_{n,q})$. Let $\overline{X_q}$, σ_q be the mean and the standard deviation of the random variable X_q respectively, and $\mathbf{B}$ be the matrix such that the p-th column corresponds to the standardized p-th data point, i.e. $\left(\frac{X_{1,p}-\overline{X_1}}{\sigma_1}, \ldots, \frac{X_{n,p}-\overline{X_n}}{\sigma_n}\right)$. One can show that $\mathbf{C} = \mathbf{B}\,\mathbf{B}^T$ is the correlation matrix.

Our goal is to find the normalized vector $\boldsymbol{u}$ centered at $(\overline{X_1}, \ldots, \overline{X_n})$ such that the projection of the standardized data cloud on $\boldsymbol{u}$ has the maximum variance. The p-th standardized data point projection on $\boldsymbol{u}$ is $\left(\frac{X_{1,p}-\overline{X_1}}{\sigma_1}, \ldots, \frac{X_{n,p}-\overline{X_n}}{\sigma_n}\right) \cdot \boldsymbol{u}$, so $\mathbf{B}^T \boldsymbol{u}$ is the ensemble of $\boldsymbol{u}$-projection of each standardized data point. One can check

that the mean of $\mathbf{B}^T \boldsymbol{u} = 0$, and therefore the variance of the standardized data cloud projected on $\boldsymbol{u}$ is $(\boldsymbol{u}^T \mathbf{B}) \cdot (\mathbf{B}^T \boldsymbol{u})/m$. To summarize, our goal is to maximize this value under the constraint $\boldsymbol{u}^T \boldsymbol{u} = 1$. Without lost of generality, we will maximize $(\boldsymbol{u}^T \mathbf{B}) \cdot (\mathbf{B}^T \boldsymbol{u})$ instead of $(\boldsymbol{u}^T \mathbf{B}) \cdot (\mathbf{B}^T \boldsymbol{u})/m$.

We use Lagrange multiplier formalism to deal with the constraint by introducing Langrange multiplier α_1. In this formalism, the goal is to maximize the expression as shown in Equation A1

$$L_1(\boldsymbol{u}, \alpha_1) = (\mathbf{B}^T \boldsymbol{u})^2 - \alpha_1(\boldsymbol{u}^2 - 1) \tag{A1}$$

By using the necessary condition of the maximality, i.e. $\frac{\partial L_1}{\partial \boldsymbol{u}} = \frac{\partial L_1}{\partial \alpha_1} = 0$, one can show that $\boldsymbol{u}$ has to be a eigenvector of $\mathbf{C}$ and $\boldsymbol{u}^2 = 1$. Recall that $\mathbf{C}$ is symmetric and therefore orthogonally diagonalizable. Let the diagonal matrix to be $\mathbf{D} = \text{diag}(\lambda_i)$, i.e. $\boldsymbol{u}^T \mathbf{C} \boldsymbol{u} = \boldsymbol{u}^T \mathbf{P}^T \mathbf{D} \mathbf{P} \boldsymbol{u}$ for some orthonormal matrix $\mathbf{P}$. Since $\mathbf{P}$ is orthonormal and $\boldsymbol{u}^2 = 1$, it is easy to check that $(\mathbf{P}\boldsymbol{u})^2 = 1$. Let λ_1 be the largest eigenvalue. One has

$$\max_{||\boldsymbol{u}||=1} \boldsymbol{u}^T \mathbf{P}^T \mathbf{D} \mathbf{P} \boldsymbol{u} = \max_{||\boldsymbol{v}||=1} \boldsymbol{v}^T \mathbf{D} \boldsymbol{v} = \lambda_1 \tag{A2}$$

This can be attained by choosing $\boldsymbol{u}$ to be the eigenvector of $\mathbf{C}$ corresponding to the eigenvalue λ_1. And thus it is the sufficient condition. Similarly, to find the second principal component $\boldsymbol{w}$ such that it is orthogonal to the first principal component $\boldsymbol{u}$ (i.e with constraint $\boldsymbol{w}^T \cdot \boldsymbol{u} = 0$) and accounts for the largest part of the rest of the variance, it suffices to maximize the expression as shown in Equation A3:

$$L_2(\boldsymbol{w}, \beta_1, \beta_2) = (\mathbf{B}^T \boldsymbol{w})^2 - \beta_1(\boldsymbol{w}^2 - 1) - \beta_2(\boldsymbol{w}^T \cdot \boldsymbol{u}) \tag{A3}$$

By similar calculation, one can show that $\boldsymbol{w}$ being the eigenvector corresponding to the second largest eigenvalue is necessary and sufficient to maximize this expression, and so forth for the subsequent eigenvectors. Graphically, we are looking for a orthogonal transformation of the random variables space such that after the transformation, the first axis will account for the largest part of the total variance, and second axis is orthogonal to the first axis and account for the largest part of the rest of the variance. It is important to note that the variances that they account, are given by the eigenvalues of the correlation matrix as shown in Equation A2.

APPENDIX B: INCOMPLETE DATA SET

If a data set is incomplete, in principle we can still calculate the Pearson's correlation for any two random variables by using only the data points that have value for both random variables, and therefore we can construct the correlation matrix entry by entry. However the problem of this approach is obvious: since the correlation matrix $\mathbf{C}$ is not $\mathbf{BB}^T$ as before, although it is still symmetric, it might not be semi-positive-definite, i.e. it might have undesirable negative eigenvalues. Our goal is to find a semi-positive-definite matrix that is close to the correlation matrix $\mathbf{C}$. Rebonato & Jäckel (1999) suggested the following:

Let $\mathbf{S}$ to be the ensemble of eigenvectors of matrix $\mathbf{C}$, i.e. $\mathbf{C} \cdot \mathbf{S} = \mathbf{\Lambda} \cdot \mathbf{S}$, where $\mathbf{\Lambda} = \text{diag}(\lambda_i)$, and λ_i the eigenvalues. If $\mathbf{C}$ is not semi-positive-definite, it has at least one negative eigenvalue. We define the positive diagonal matrix $\mathbf{\Lambda}' \equiv$

$\text{diag}(\lambda_i')$:

$$\mathbf{\Lambda'} \quad : \quad \lambda_i' = \begin{cases} \lambda_i & \text{if } \lambda_i \geqslant 0 \\ 0 & \text{if } \lambda_i < 0 \end{cases} \tag{B1}$$

and the diagonal 'scaling' matrix $\mathbf{T} \equiv \text{diag}(t_i)$:

$$\mathbf{T} \quad : \quad t_i = \left[\sum_m s_{im}^2 \lambda_m' \right]^{-1} \tag{B2}$$

Let $\mathbf{B'} \equiv \sqrt{\mathbf{T}} \mathbf{S} \sqrt{\mathbf{\Lambda'}}$, where the square root of a diagonal matrix is defined as the square root of each of its diagonal entry. Finally we define: $\mathbf{C'} \equiv \mathbf{B'B'}^T$. One would expect $\mathbf{C'}$ to be quite close to $\mathbf{C}$ since $\mathbf{C} = \mathbf{S}^T \mathbf{\Lambda} \mathbf{S}^T$ and $\mathbf{C'} = \sqrt{\mathbf{T}} \mathbf{\Lambda'} \mathbf{S} \sqrt{\mathbf{T}}$. The lost from $\mathbf{\Lambda} \to \mathbf{\Lambda'}$ is compensated by the rescaling matrix $\mathbf{T}$. There are better ways to optimize the search of $\mathbf{C'}$ but they are mostly computational much more demanding than this method. In our case, this estimation is good enough since it gives reasonable small errors both in term of $\epsilon_1 \equiv \sum_{ij}(C_{ij} - C_{ij}')^2$ and $\epsilon_2 \equiv \sum_{i=1}^{n}(\lambda_i - \lambda_i')^2$, where λ_i' are eigenvalues of $\mathbf{C'}$.

APPENDIX C: WEIGHTED TOTAL LEAST SQUARE

This method is adopted from Krystek & Anton (2007). As discussed in Section 4.2, our goal is to minimize Equation 1. Instead of considering the best fit line $y = ax + b$ using variables a and b, Krystek & Anton (2007) suggested a change of variable $\mathbb{R}^2 \underset{(a,b) \longmapsto (\alpha, p)}{\to} (-\frac{\pi}{2}, \frac{\pi}{2}) \times \mathbb{R}_+$, where $a = \tan(\alpha)$ and $b = p/\cos(\alpha)$. They showed that in this case, Equation 1 becomes

$$\chi^2(\alpha, p) = \frac{1}{n-2} \sum_{k=1}^{n} \frac{(y_k \cos \alpha - x_k \sin \alpha - p)^2}{u_{x,k}^2 \sin^2 \alpha + u_{y,k}^2 \cos^2 \alpha} \tag{C1}$$

For our case, we assume $u_{x,k} = u_{y,k} = \sigma$, for all k. Therefore we have a very neat expression:

$$\chi^2(\alpha, p) = \frac{1}{\sigma^2(n-2)} \sum_{k=1}^{n} (y_k \cos \alpha - x_k \sin \alpha - p)^2 \tag{C2}$$

We use Truncated-Newton Method (`TNMIN.pro` in IDL, written by Craig B. Markwardt) to search for the minimal point (α_0, p_0). Furthermore, we can approximate the uncertainty of the parameters estimation using the inverse Hessian of χ^2. More explicitly:

$$\begin{pmatrix} \sigma^2(p) & \text{cov}(p, \alpha) \\ \text{cov}(\alpha, p) & \sigma^2(\alpha) \end{pmatrix} = 2 \begin{pmatrix} \chi_{pp}^2 & \chi_{p\alpha}^2 \\ \chi_{\alpha p}^2 & \chi_{\alpha\alpha}^2 \end{pmatrix}^{-1} \Bigg|_{\alpha=\alpha_0, p=p_0} \tag{C3}$$

where $\chi_{pp}^2 \equiv \frac{\partial^2 \chi^2}{\partial p^2}$, $\chi_{\alpha\alpha}^2 \equiv \frac{\partial^2 \chi^2}{\partial \alpha^2}$, $\chi_{\alpha p}^2 = \chi_{p\alpha}^2 \equiv \frac{\partial^2 \chi^2}{\partial \alpha \partial p}$