Astronomy & Astrophysics manuscript no. manuscript
September 6, 2016

Classification and diagnostics of Galah dataset with t-SNE reduction of spectral information

G. Traven[1], G. Matijevič, T. Zwitter[1], M. Žerjal[1], and the Galah team

Faculty of Mathematics and Physics, University of Ljubljana, Jadranska 19, 1000 Ljubljana, Slovenia;
e-mail: gregor.traven@fmf.uni-lj.si

September 6, 2016

ABSTRACT

Galah is an ongoing high-resolution spectroscopic survey with a goal to disentangle the formation history of the Milky Way, using fossil remnants of disrupted star formation sites which are now dispersed around the Galaxy. It features a randomly selected magnitude limited (up to $V = 14$) sample of already 300,000 spectra in pursue of observing 1 Mio stars. We present a semi-automated classification scheme which identifies different types of peculiar spectral morphologies, in an effort to discover and flag potentially problematic spectra and thus help to preserve the integrity of the surveys results. To this end we employ a recently developed dimensionality reduction technique t-SNE (t-distributed Stochastic Neighbour Embedding), which enables representing the complex spectral morphology in a low dimensional projection map while still preserving the properties of the local neighbourhoods of spectra. The distribution of normal single stars in the two dimensional space reveals correlation with stellar temperature, surface gravity and metallicity, while the peculiar and problematic spectra with very diverse spectral features are grouped together in separate areas, allowing their detection and classification. We find that the majority of the Galah dataset represents normal spectra (167, 928), whereas 41, 605 spectra pertaining to 38, 633 stars are distributed into 10 classification categories: *hot stars*, *cool metal-poor giants*, *molecular abs. bands*, *binary stars*, *Hα/Hβ emission*, *Hα/Hβ emission superimposed on absorption*, *Hα/Hβ P-cygni*, *Hα/Hβ inverted P-cygni*, *lithium absorption*, and *problematic*. Spectra with assigned categories are presented in the catalogue together with supplementary information on the observed objects. The larger group of normal spectra enables a detailed analysis with existing stellar models, while some peculiar categories represent candidates for follow-up observations and population studies of the short-lived phases of stellar evolution detected in this randomly selected sample of stars.

1. Introduction

In recent times, the technology of optical-fibre spectrographs has enabled very efficient large-scale automated spectroscopic surveys. With the ability of observing up to several hundred stars simultaneously, it is now possible to obtain large numbers of spectra of high quality in a reasonable amount of time (Watson 1987). Surveys such as the Radial Velocity Experiment (Steinmetz et al. 2006), the ongoing Gaia-ESO Survey (Gilmore et al. 2012), Galah project (De Silva et al. 2015), and Gaia mission (Prusti 2012) with its future follow-up projects WEAVE (Dalton et al. 2012) and 4MOST (de Jong et al. 2012) are some of the leading examples of continuous production of overwhelming amounts of data.

To get a general overview of the observed spectra and learn more about the studied sample of stars in a spectroscopic survey such as Galah, it seems reasonable and necessary to attack this task in an unbiased and automated way. A common approach is to employ different numerical dimensionality reduction methods to reveal the complex morphological structure of the dataset at hand. By projecting the spectra onto a low dimensional space, it becomes feasible to grasp their inter-correlations and identify diverse morphological groups, thus constructing a classification of the whole dataset, and particularly its outstanding features. A plethora of linear and non-linear mathematical techniques have been developed in the past decades to tackle the problem of classification of complicated high-dimensional data such as spectra, and were also successfully applied in the astronomical community (Matijevič et al. 2012 and references therein).

Galah is an ongoing spectroscopic survey that aims at unveiling the Milky Way's history by studying the fossil record of ancient star formation and accretion events preserved in stellar light. The detailed knowledge of the chemical information of fossil remnants which have disrupted and are now dispersed around the Galaxy is essential to disentangling its formation history and explaining its current stellar populations. Recent studies of chemical abundances of stars in individual (undispersed) open clusters show that their abundance distributions are homogeneous to the level at which they can be measured, and their abundances are different from cluster to cluster (conversely, globular clusters may display inhomogenities, Bekki et al. (2007)). This enables the technique of chemical tagging (Freeman & Bland-Hawthorn 2002) to identify the fossil remnants of old dispersed clusters from their abundance patterns over many chemical elements. Galah will achieve this by measuring up to 30 elemental abundances from 7 independent element groups each with 5 measurable abundance levels, thereby obtaining enough independent cells (5^7) in the multi-dimensional chemical abundance space (C-space), in which stars from chemically homogeneous aggregates (e.g. disrupted open clusters) will lie in tight clumps (Freeman 2012). This level of accuracy and the amount of

elemental abundance information by far surpasses any existing single or multiple system stellar studies.

The Galah automatic pipeline is currently running without a classification processing stage. Manually scanning the observed sample, it has become obvious that there is a significant number of peculiar and otherwise problematic spectra. Although the majority belong to single stars and can be properly fit by synthetic spectra, neglecting the outliers can lead to wrong results in both radial velocities, atmospheric parameters, and especially detailed chemical abundances. Finding outliers by comparison to databases of known peculiar spectra might produce useful results, but would fail to give a reliable classification of the whole sample. We therefore aim to diagnose and classify the diverse morphologies in the Galah dataset with the goal to: (1) provide a clean sample without any peculiar or problematic spectra so the survey results can be more reliable, (2) identify any peculiar spectra that are interesting per se and merit further investigation, and (3) highlight all problematic spectra with unpredictable effects from either instrumentation or reduction stage. The method that we use to identify patterns or groups in the "feature space" to achieve the stated goals is unsupervised classification with t-SNE reduction of spectral information. The main advantage of this approach is that we are more likely to detect various unfamiliar morphological features as well as the many known and expected peculiar stars. For a very nice overview of the various classification and data mining techniques we refer the reader to Sharma et al.

The paper is organised as follows: The reduction and overview of Galah data is described in Section 2, the classification procedure with description of the employed techniques is detailed in Section 3, and the discovered classes of spectra are examined in Section 4 and 5. In Section 6 we present the structure of the catalogue with final classification results including supplementary information, and in Section 7 we briefly describe a visualisation tool nicknamed *Galah Explorer* which displays the t-SNE projection map featuring various useful functionalities. We conclude with discussion in Section 8.

2. Data and reduction overview

2.1. Galah spectra

The Galah survey was the main driver for the construction of Hermes (High Efficiency and Resolution Multi-Element Spectrograph), a new fiber-fed multi-object spectrometer on the 3.9 m Anglo-Australian Telescope (AAT). Its spectral resolution is about 28,000, and there is also an R = 45,000 mode using a slit mask. The spectrometer is fed via 400 fibres distributed over π square degrees of sky. Taking into account the GALAH magnitude limitation ($V = 14$), 400 stars can be observed simultaneously in that relatively small angle up to galactic latitude of $|b| \sim 28°$. HERMES has four simultaneous non-contiguous spectral bands centred at 4800, 5761, 6610 and 7740 Å (hereafter red, green, blue, and IR band), covering a total of about 1000 Å, including Hα and Hβ lines. The spectrograph is designed to have $\sim 10\%$ efficiency and to achieve SNR ~ 100 per resolution element at $V = 14$ in 1 hour exposure with measured RV errors < 1 kms^{-1} (Martell et al., submitted).

2.2. Reduction pipeline

All the spectra subject to our analysis are reduced by the pipeline used in the Galah survey to produce fully calibrated spectra for subsequent stellar atmospheric parameter estimation. The reduction pipeline is based on reliable Iraf routines and other readily available software. After the Iraf-based reduction, a code that provides first estimates of radial velocity and three basic atmospheric parameters is run and normalization of the entire observed spectrum is done for each star (see Section 6 in Kos et al. (2016)). For some spectra, the processing by this code can fail due to various reasons, and such cases are excluded from further consideration. Otherwise, the values of the three parameters ($T_{\rm eff}$, log g, [Fe/H]), to which we refer in the text and which are color coded in several figures, are produced by this code.

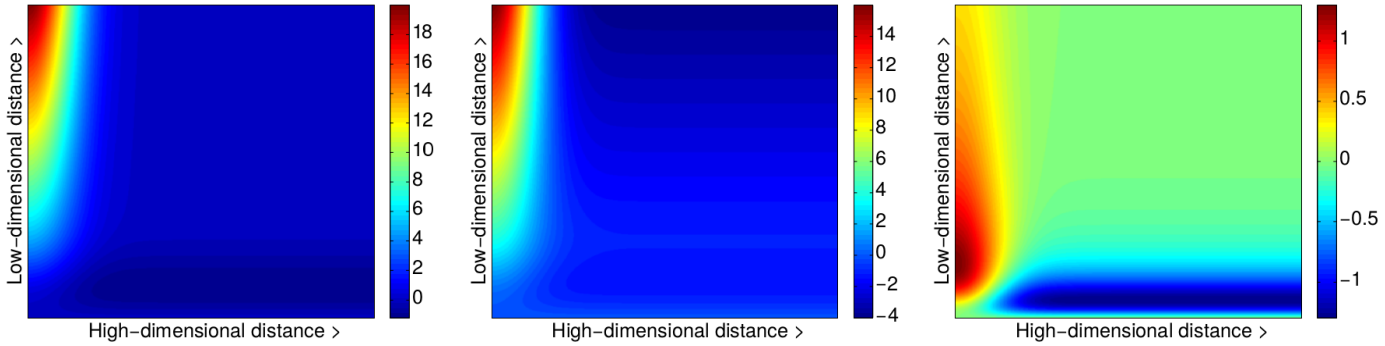
About 300,000 spectra have accumulated to date, including various calibration exposures, however we only concentrate on 210,000 spectra recorded before 30th January 2016 and reduced with the Iraf reduction pipeline version 5.1. In the future, the same study will be extended to include the upcoming spectra once they become available. For more details, we refer the reader to the thorough description of the reduction process (Kos et al. 2016). An improved process is also being developed for computation of stellar parameters using a combination of the spectral synthesis program Spectroscopy Made Easy (SME) (Valenti & Piskunov 1996; Piskunov & Valenti 2016) and the data-driven machine learning procedure of The Cannon (Ness et al. 2015), which is detailed in Martell et al. (submitted).

3. Classification

We devise a custom classification procedure which is based on two independently developed methods, the novel dimensionality reduction technique t-SNE and renowned clustering algorithm DBSCAN. Both are used more than once in an iterative approach to enable the most efficient classification and overview of our dataset. It should be noted that these purely mathematical methods are used extensively in various domains of research for unsupervised classification or clustering, and were not primarily intended for astrophysical purposes, but nevertheless they have proven to be very effective in distinguishing different morphological groups in spectral data and seem to be far more efficient than any other similar state of the art techniques. We shortly present both methods and then focus on our custom classification procedure.

3.1. t-SNE reduction of spectral information

Reduction of accumulated data is an important problem in many different domains, and deals with data of widely varying dimensionality. t-SNE (t-distributed Stochastic Neighbour Embedding) can visualize any high-dimensional data by giving each datapoint a location in a two-dimensional map. The technique is an improvement of Stochastic Neighbor Embedding (Hinton & Roweis 2002) that is much easier to optimize, and produces significantly better visualizations by reducing the tendency to crowd points together in the center of the map. t-SNE is better than existing techniques at creating a single map that reveals structure at many different scales. This is particularly important for



(a) Gradient of SNE. (b) Gradient of UNI-SNE. (c) Gradient of t-SNE.

Fig. 1: SNE gradients (from van der Maaten & Hinton (2008)) as a function of pairwise Euclidean distances ($\|\mathbf{x}_i - \mathbf{x}_j\|$ on horizontal and $\|\mathbf{y}_i - \mathbf{y}_j\|$ on vertical axes). Positive (red) values of the gradient represent an attraction between the points in Q space whereas negative values (blue) represent a repulsion.

high-dimensional data that lie on several different, but related, low-dimensional manifolds, such as SB1 spectra of multiple spectral types shifted by different radial velocities. For an illustration of the superior performance of t-SNE on a wide variety of data sets and comparison with many other non-parametric visualization techniques, including Sammon mapping, Isomap, and Locally Linear Embedding, we refer the reader to the original paper (van der Maaten & Hinton 2008), while we give a brief introduction of the technique in the following paragraphs.

t-SNE interprets the overall distances between datapoints in the high-dimensional space as a symmetric joint-probability distribution P. Likewise a joint-probability distribution Q is computed, that describes the similarity in the low-dimensional space. The goal is to achieve a representation, referred to as *embedding*, in the low dimensional space where Q faithfully represents P. This is achieved by optimizing the positions in the low-dimensional space to minimize the cost function C given by the Kullback-Leibler (KL) divergence between the joint-probability distributions P and Q:

$$C(P,Q) = KL(P\|Q) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}} \tag{1}$$

Given two data points $\mathbf{x}_i$ and $\mathbf{x}_j$ the probability p_{ij} models the similarity of these points in the high-dimensional space. To this extent, for each point a Gaussian kernel, P_i, is chosen whose variance σ_i is defined according to the local density in the high-dimensional space and then p_{ij} is described as follows:

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

$$\text{where} \qquad p_{j|i} = \frac{\exp(-\|a_i - a_j\|^2/2\sigma_i^2)}{\sum_{k\neq i} \exp(-\|a_i - a_k\|^2/2\sigma_i^2)} \tag{3}$$

$p_{j|i}$ can be seen as a relative measure of similarity based on the local neighborhood of a data-point $\mathbf{x}_i$. The perplexity value μ is a user-defined parameter that describes the effective number of neighbours considered for each datapoint. The value of σ_i is chosen such that for fixed μ and each i:

$$\mu = 2^{-\sum_j^N p_{j|i} \log_2 p_{j|i}} \tag{4}$$

In regions of the high-dimensional space with a higher data density, sigma tends to be smaller than in regions of the data space with lower density.

A normalized heavy-tailed *Student's t-Distribution* kernel with a single degree of freedom is used to compute the joint-probability distribution in the low-dimensional space Q, where the positions of the datapoints should be optimized. Given two low-dimensional points $\mathbf{y}_i$ and $\mathbf{y}_j$, the probability q_{ij} that describes their similarity is given by:

$$q_{ij} = ((1 + \|\mathbf{y}_i - \mathbf{y}_j\|^2)Z)^{-1} \tag{5}$$

$$\text{with} \qquad Z = \sum_{k\neq 1}^{N} \sum_{k\neq l}^{N} (1 + \|\mathbf{y}_k - \mathbf{y}_l\|^2)^{-1} \tag{6}$$

The heavy tails of the *Student's t-Distribution* allow dissimilar datapoints $\mathbf{x}_i$ and $\mathbf{x}_j$ to be modeled by low-dimensional counterparts $\mathbf{y}_i$ and $\mathbf{y}_y$ that are too far apart. This is desirable because it creates more space to accurately model the small pairwise distances, so the local structure in space Q.

The gradient of the Kullback-Leibler divergence between P and Q is used to minimize the non-convex cost function C (see Equation (1)). It indicates the change in position of the low-dimensional points for each step of the gradient descent and is given by:

$$\frac{\partial C}{\partial \mathbf{y}_i} = 4 \sum_{i=1}^{N} (F_i^{\text{attr}} - F_i^{\text{rep}}) \tag{7}$$

$$= 4 \sum_{i=1}^{N} \Big(\sum_{j\neq i}^{N} p_{ij} q_{ij} Z(\mathbf{y}_i - \mathbf{y}_j) - \sum_{j\neq i}^{N} q_{ij}^2 Z(\mathbf{y}_i - \mathbf{y}_j) \Big) \tag{8}$$

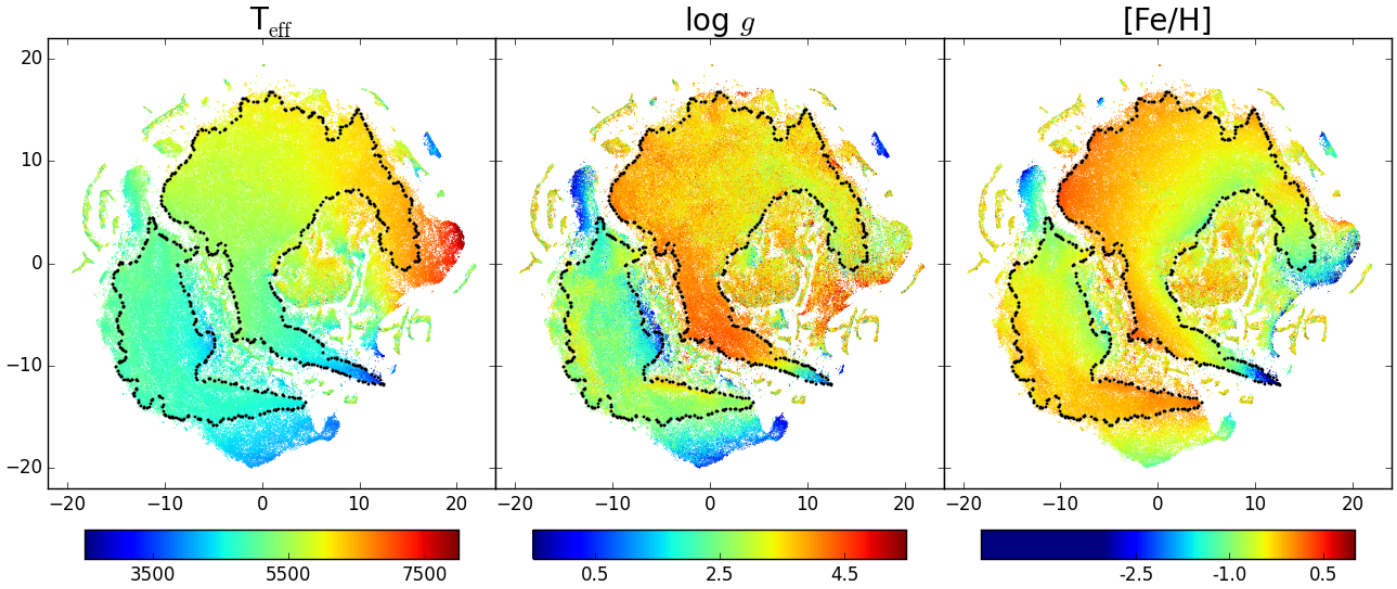


Fig. 2: The first t-SNE projection of the whole working set containing 209,533 datapoints (spectra). The three panels feature T_{eff}, log g, and [Fe/H] values for spectra, measured by the Galah reduction pipeline. The two areas encircled by black points are the two largest collections of the most appropriate DBSCAN mode ($Eps = 0.2$, $MinPts = 30$) for large scale cluster detection, which were removed in the third step of our classification procedure. The outer borders of the two collections are dotted, indicating the approximate shape which is not smooth and can be patchy also on the inside. A relatively smooth distribution of parameters is clearly seen in the enclosed areas. Altogether DBSCAN defines 235 collections in this mode, which are not marked here. All axes are arbitrary.

The gradient descent can be seen as an N-body simulation, where each data-point exerts an attractive and a repulsive force on all the other points ($F_i^{\text{attr}}, F_i^{\text{rep}}$). Figure 1 compares gradients for three different variants of SNE as a function of their pairwise Euclidean distances. The three panels demonstrate two main advantages of t-SNE gradient over the other two. First, t-SNE gradient strongly repels datapoints that are close in space Q but very dissimilar in space P. The other two also have such repulsion but its effect is minimal compared to the strong attractions elsewhere, which we can see by the scale on the vertical axes. Second, although t-SNE introduces strong repulsions between dissimilar datapoints that are modelled by small pairwise distances, these repulsions do not go to infinity as in the case of UNI-SNE gradient.

In this work, we are dealing with over 200,000 spectra which are basically datapoints in space P of dimensionality $\sim 13,600$ (all pixels from four spectral bands). The high computational complexity introduced by employing t-SNE on our growing dataset requires that we make use of the Barnes-Hut t-SNE (van der Maaten 2013), an evolution of the t-SNE algorithm that introduces different approximations to reduce the computational cost from $O(N^2)$ to $O(N \log(N))$ and the memory complexity from $O(N^2)$ to $O(N)$.

The perplexity is set to 30, a value that has generally proven to be most effective for our purpose. Smaller values of this input parameter produce sparser projection maps with denser collections of points and larger values produce more evenly covered projection space but with less pronounced separations between distinct groups. To obtain the low-dimensional embedding of our dataset, the t-SNE procedure boils down to: (1) converting Euclidean distances in space P to pairwise similarities (often computationally the most intensive part), (2) sampling map points randomly from an isotropic Gaussian with small variance that is centered around the origin, and (3) initialising the gradient descent with a fixed number of iterations (usually 1000).

3.2. DBSCAN clustering

Density-based spatial clustering of applications with noise (DBSCAN) is a data clustering algorithm relying on a density-based notion of clusters (collections) and designed to discover any arbitrary shape of collections of points in some space. It groups together those points that are closely packed (points with many nearby neighbours), marking as outliers those that lie alone in low-density regions (whose nearest neighbours are too far away). DBSCAN is one of the most common clustering algorithms and also one of the most cited in scientific literature.

There are two input parameters to DBSCAN method that have to be set by the user, *minPts* and *Eps*. Furthermore, the points are classified as core points, (density-)reachable points and outliers. The definitions are as follows:

- A point p is a core point if at least *minPts* points are within distance *Eps* of it (including p). Those points are said to be directly reachable from p. By definition, no points are directly reachable from a non-core point.
- A point q is reachable from p if there is a path p_1, ..., p_n with $p_1 = p$ and $p_n = q$, where each p_{i+1} is directly reachable from p_i (all the points on the path must be core points, with the possible exception of q).

– All points not reachable from any other point are outliers.

To find a collection of points, DBSCAN starts with an arbitrary point p and retrieves all points density-reachable from p wrt. Eps and $MinPts$. If p is a core point, this procedure yields a collection. If p is a border point, no points are density-reachable from p and DBSCAN visits the next point.

In this study, DBSCAN is employed merely as a tool for automatic detection of distinct collections in the projection map produced by t-SNE, without any bias apart from the manual selection of Eps and $MinPts$. By definition of the t-SNE projection map, datapoints that are similar to each other should be closely packed together, thereby, in DBSCAN's terminology, forming a collection which can be detected and labelled.

3.3. Classification procedure

In the process of finding the most productive and feasible way of classifying Galah spectra, we have established a procedure which is a combination of automatic and manual processing and inspection of our data. The current Galah reduction pipeline includes several stages (see Section 2) and we only retain those spectra that pass radial velocity determination and normalisation, which are in our analysis the two key properties for optimal use of t-SNE dimensionality reduction. In principle, we could use the whole dataset of spectra before any kind of reduction, but in that case the most important features driving our low-dimensional embedding (projection map) would come from missing wavelength calibration, flux normalisation, radial velocity determination, etc., in essence we would be concerned with properties, that are not scientifically intriguing and can be reliably enough accounted for with standard reduction procedures. After this selection, the number of spectra left is $209,533$, and we join together the four spectral bands to produce the so-called datapoints of our working set.

Taking the whole practically usable range of red, green, blue, and IR band, the number of normalised flux values, which are basically pixels or original dimensionality of datapoints, amounts to $13,600$ per datapoint (spectrum). At such high dimensionality multiplied by $209,533$ datapoints, the computational cost despite the Barnes-Hut implementation of t-SNE is relatively overwhelming. To ease the whole process, in terms of memory consumption and computation time, we make use of the following scheme:

1. **First t-SNE projection**: the first projection is computed for the whole working set ($209,533$ spectra). We use only the following wavelength ranges: $4850-4880$ Å, $5750-5780$ Å, $6550-6580$ Å, and $7730-7760$ Å. These ranges are selected so as to include the more interesting and diagnostic parts of each spectral band, with equal contribution from all of them, amounting to 2400 as dimensionality of datapoints. The map resulting from this first projection can be used as it is, for this is the most basic and objective clustering of all datapoints from our working set. However, a significant portion of spectral information is missing due to our cut in wavelength range. There is also a practical caveat, in that although the projection map is fairly homogeneously populated, it is also very dense, making the smaller scale clustering, the one we are most interested in, more difficult to recognise. The next steps alleviate these issues.
2. **DBSCAN large scale cluster detection**: DBSCAN input parameters are set in a way that a few large collections of datapoints are defined across the projection map. This is done in order to select those collections that presumably contain only the "normal" spectra. In our experience, well behaved spectra are usually clustered together in one or a few large areas of the map where the atmospheric parameters ($T_{\rm eff}$, $\log g$, [Fe/H]) are continuously distributed.
3. **Select and filter out collections of "normal" spectra**: It is here where the manual interaction is most important, since we are rejecting (to our analysis) less interesting datapoints and even if being very careful, we can unwillingly discard some of the desired spectra from further consideration. The map of the first t-SNE projection with DBSCAN clustering on large scale is shown in Figure 2. The two largest collections in orange and red color, amounting to 137155 datapoints, are rejected for containing presumably only "normal stars", and 76938 remaining datapoints are considered in the next steps.
4. **Second t-SNE projection**: the second projection is computed for the subsample of the working set (76938 spectra), result of the previous step. The projection map is shown in Figure 3, and serves as the final basis for our selection and analysis of peculiar spectra. Some "normal" spectra are still present in this map, but the largest portion should belong to all the peculiar objects that we are interested in. Their small scale structure, which was hidden in the first t-SNE projection, is now reflected in the large scale structure, and also more easily discernible due to the overall fewer datapoints in the map, as evident from comparing Figures 2 and 3.
5. **DBSCAN small scale cluster detection**: DBSCAN input parameters are set in a way that the defined collections correspond to relatively smaller and dense regions in the map which represent distinct morphological classes of spectra. In our experience, there is not a single parameter set for DBSCAN that would allow us to properly select the various collections, so many sets of parameters are tried and the corresponding DBSCAN results (hereafter DBSCAN modes) with different sizes and shapes of collections are available for inspection in the next step.
6. **Select relevant/categorical collections and assign classification categories/flags**: The final step involves manual overview of individual spectra in different collections with the help of our visualisation tool presented in Section 7. The goal is to find the best collection from different DBSCAN modes that fully encompasses the manually examined spectra belonging to a distinct classification category. Some outliers in terms of a chosen category will usually be present, so the final results should be regarded as a list of candidate members of some chosen classification category, where some will be probable and others possible members, but we do not make this distinction here as it is out of scope of this study. Furthermore, the selected collections from different DBSCAN modes might sometimes overlap, so one spectrum might be assigned to two categories, in which case it is labelled as a possible member of both.

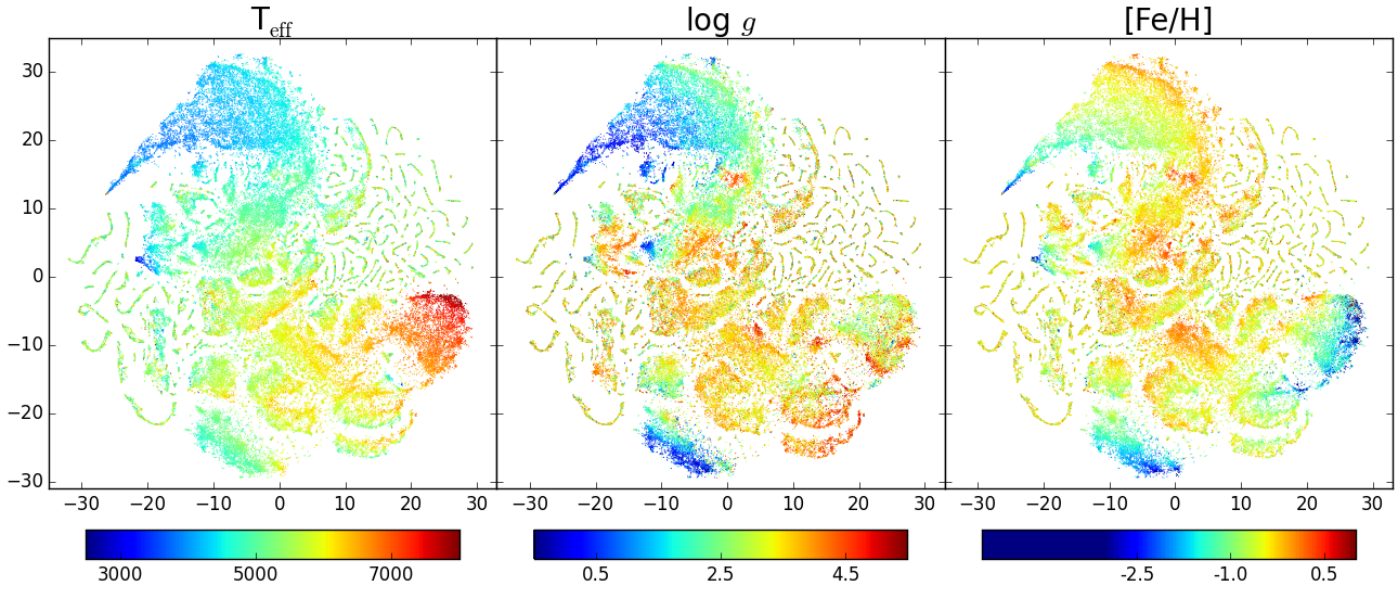


Fig. 3: Same as Figure 2 but for the second t-SNE projection map of the filtered working set containing 76938 datapoints (spectra), those that are not part of the two large collections marked in Figure 2.

4. Morphological classes of spectra

Table 1 lists 6 distinct classification categories that were defined using the classification procedure described in Section 3.3. This classification is not limited strictly to peculiar objects having spectra without a counterpart in the library of synthetic spectra, although they remain the principal motivation for this work, but is rather a search for any coherent group in the projection map, from which anyone can pinpoint their category of interest. They span from larger collections of points (spectra) like the *Hot stars* category to smaller ones that mostly contain problematic spectra, having usually one, albeit very prominent feature (e.g. a strong emission spike). The projection map that was used to search for and define the 6 general classification categories is presented in Figure 3. The overall distribution of parameters $T_{\rm eff}$, log g, and [Fe/H] in the three panels indicates their importance in feature space, where $T_{\rm eff}$ is evidently the leading one with a gradient over the whole projection map, followed by [Fe/H] and log g with dominating influence in the distribution of points inside larger and well separated collections. The collections that represent distinct classification categories are marked in Figure 4, and for some of them, there is a strong feature present in the spectra, hence the three main stellar parameters can be well mixed.

For each object from the sample of spectra with assigned classification category, a search by coordinates inside 1 arcsec radius is performed on the SIMBAD database and the most common SIMBAD *Main types* are listed along with SIMBAD *Other types* in Table 1. In the following paragraphs, classification categories are described individually, with several issues related to observations and reduction joined under category *problematic*.

4.1. Hot stars

A large collection at the right part of the map in Figure 3 contains mostly early type stars, with temperatures well above solar, characterised predominantly by widened wings of Hα and Hβ absorption lines. We observe a smooth transition of temperatures inside this collection, ranging from about 6500 up to 8000 K. The distribution of metallicity is also very smooth, along an axis perpendicular to temperature, while the surface gravity is more patchy, with dwarfs more clustered in some parts and giants more dispersed throughout the collection. A few examples of spectra with different metallicity and temperatures are shown in Figure 5.

4.2. Cool metal-poor giants

The collection in the bottom part of the projection map features mostly late type stars with a measured metallicity well below solar value ($-4.5 <$ [Fe/H] < -0.5, see Figure 6). The distribution of surface gravity is clearly seen, with a majority of stars being giants, and having temperatures mostly in the range from 4000 K to a little above solar temperature. The available records from SIMBAD support these claims, with 99 stars classified as *Red Giant Branch star*, 15 as *Possible Red Giant Branch star*, and 10 as *Variable Star of RR Lyr type*.

4.3. Stars with molecular absorption bands

The upper left part of the map in Figure 3 contains a region of the projection space that is populated by spectra with strong molecular absorption bands. It is well isolated on one end but still connected to the rest of the late type stars on the other end. The measured temperatures in this collection are mostly expectedly low, but not unquestionable, as is nicely demonstrated by the very tip of this area on

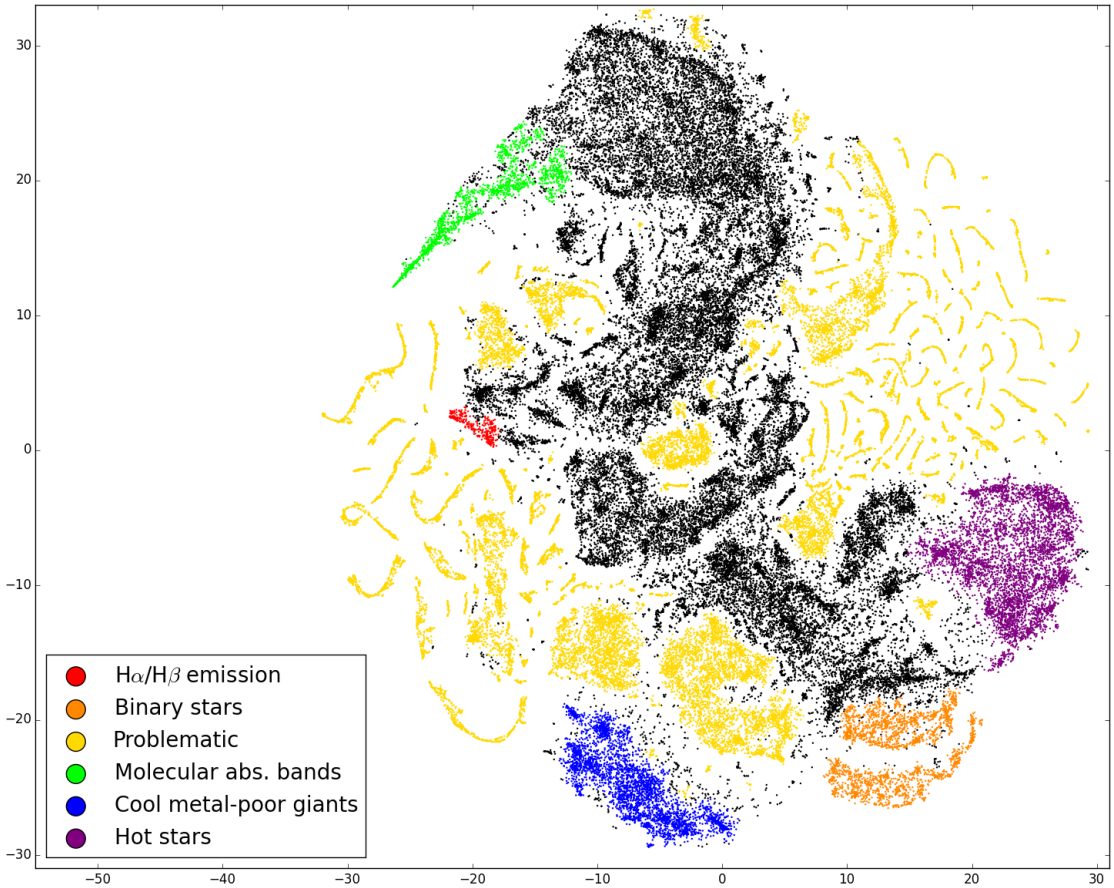


Fig. 4: The result of the classification procedure, based on the projection map in Figure 3. Collections of distinct types (categories) of spectra are flagged. All axis are arbitrary.

the left side, where we find the strongest absorption bands, while the temperatures derived for some of these spectra are much too high (above 6500 K). The progression of intensity of absorption bands nicely follows the direction from this extreme end to the larger region of late type spectra (from top to bottom on panels in Figure 7). The surface gravities and metallicities in this whole collection are also not to be trusted fully, due to the well known problem of producing reliable synthetic templates for such stars.

4.4. Binary stars

Multiple stars of which the majority represents SB2 systems are found in the bottom right part of the projection map, clustered in two well separated collections. The main and most obvious difference between them is the position of the stronger of the two components in terms of the equivalent width of absorption lines. For the collection below, the stronger component is positioned blueward and for the one above redward. Although the distinction is physically almost irrelevant, it is evidently morphologically important. Following the arc-like shape of the two collections from left to right, the spectra show progressively larger separation of the two components, from almost blended double lines to those separated by as much as 150 km/s (from bottom to top in Figure 8, the topmost spectrum belonging to a W Uma star). We have direct and indirect indications of binarity in Table 1, although only for a handful of stars, the other candidates being currently unknown for their binary nature according to the SIMBAD database. The same search by coordinates as performed on SIMBAD reveals 6 systems in Pourbaix et al. (2004) and 8 systems in Mason et al. (2001) catalogues. By visual inspection, some SB3 candidates and W UMa type SB2s are also a part of this collection, although they are not isolated enough in the projection map to be labelled separately.

The binary fraction in this collection represents around 1% of the investigated Galah spectra. However the true number of such objects is doubtlessly larger, as there are many factors hindering their detection: SB2 suffering from line blending, exclusion of potential candidates in the third step of our classification procedure or blocking of their detection by some more prominent *problematic* feature. The simulation from Matijevič et al. (2010) performed for SB2 analysis of RAVE spectra found that the detection rate should be fairly high ($\sim$ 80%) for systems with orbital periods shorter than $\approx$ 100 days. The limiting line separation $\triangle v_{orb} \approx 50$ kms^{-1} for RAVE (near-IR, SNR $\sim$ 45, $R \sim 7500$) should decrease for Galah due to the higher resolution and signal-to-noise ratio of spectra and so greatly improve the detection of longer period systems. Indeed, the smallest separations among the detected binaries in this collection are $\triangle v_{orb} \approx 15$ kms^{-1}.

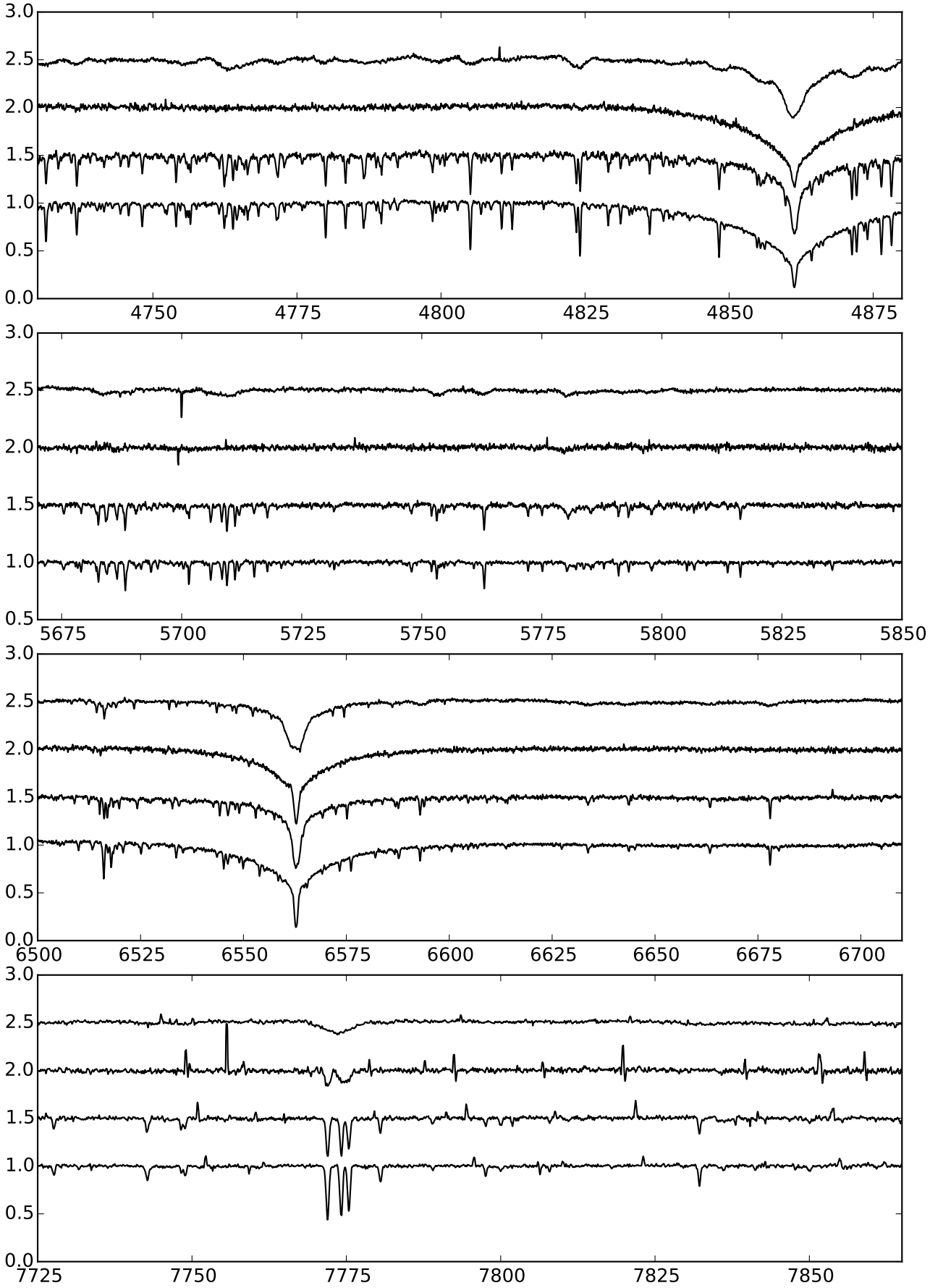


Fig. 5: Examples of *hot stars* category. Vertical spacing between spectra in each panel is adjusted for clarity. From bottom to top in each panel, spectra are identified by J2000 coordinates with APASS V magnitude as listed in Table 2.

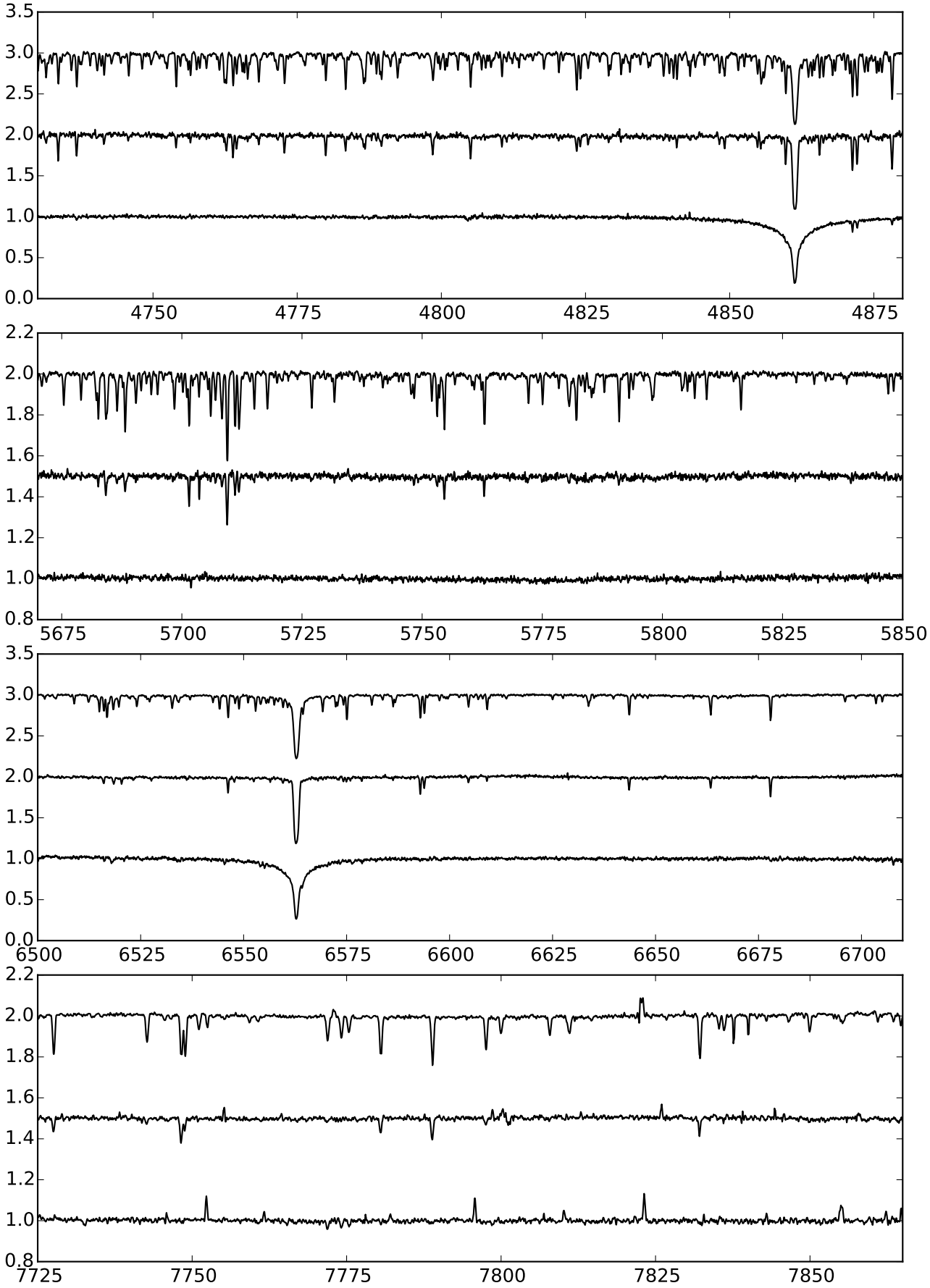


Fig. 6: Same as Figure 5 but for *cool metal-poor giants*.

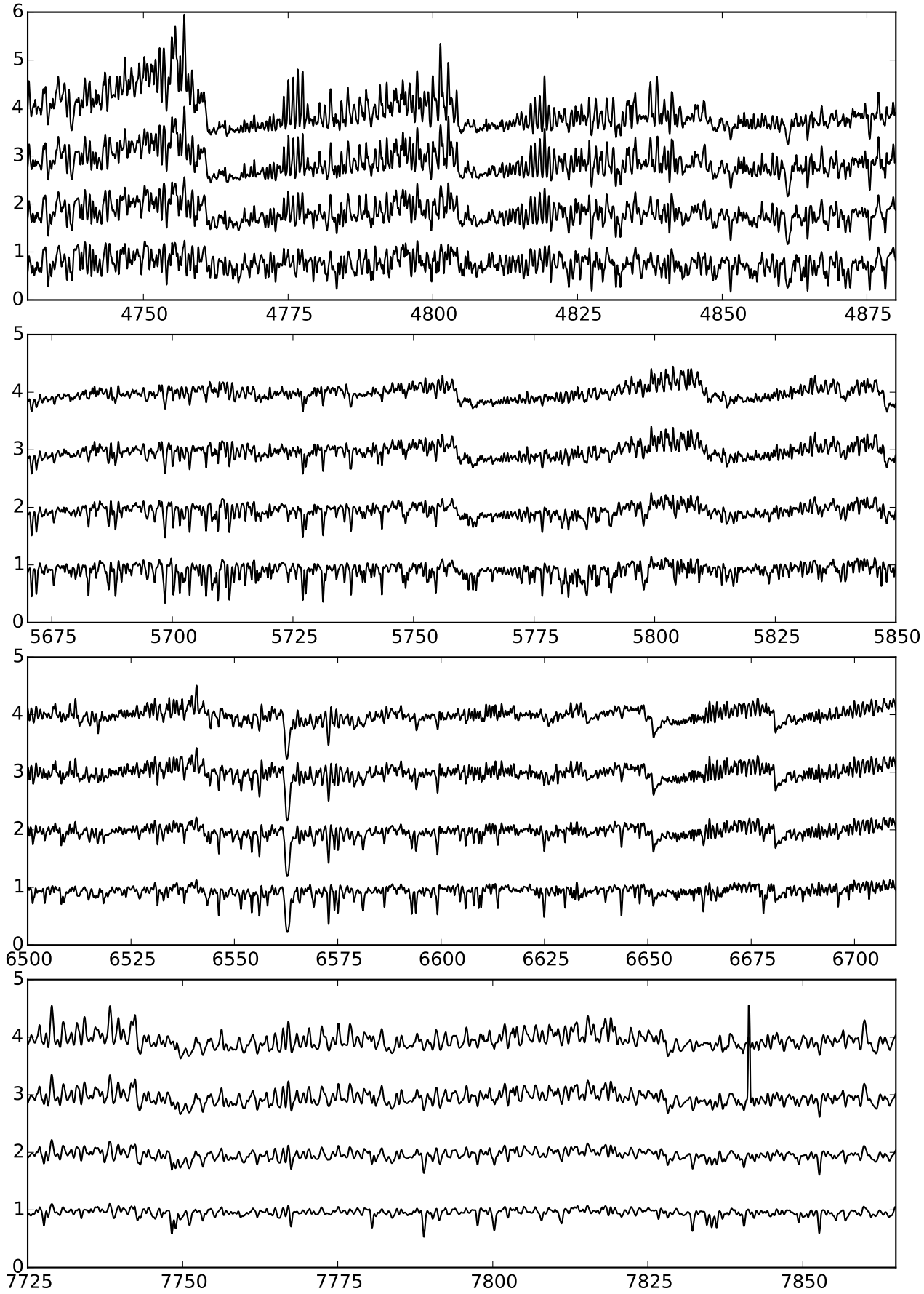


Fig. 7: Same as Figure 5 but for *molecular absorption bands*.

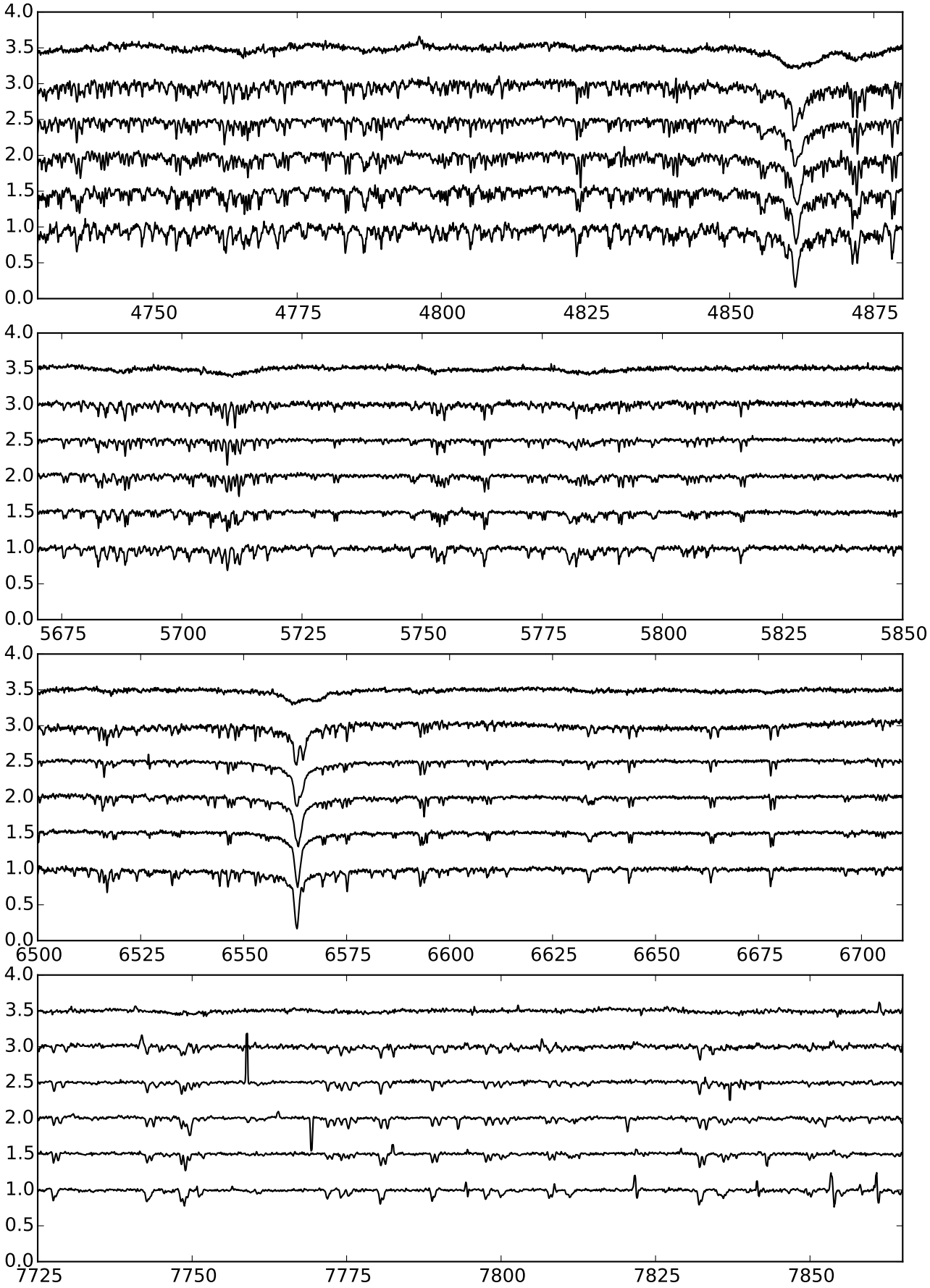


Fig. 8: Same as Figure 5 but for *binary stars*.

4.5. Hα/Hβ emission

Emission-type stars often feature diverse emission profiles in Hα and Hβ lines, indicative of young stars, cataclysmic variables, symbiotic stars, stars with massive outflows or inflows, and to many other types of active objects. The shapes of emission profiles can be described by meaningful morphological and possibly also physical categories, as demonstrated by (Traven et al. 2015). In this collection, the diverse profiles (double peaks, emission superimposed on absorption, P-Cygni, and others) are presented together, as they are relatively few and also not cleanly separated in the projection map. The Hα emission line is mostly present and often accompanied by a similar profile shape of Hβ line. In some cases, molecular absorption bands and the lithium absorption line are clearly visible, all together indicative of cooler, younger and active stars, supported also by SIMBAD classes in Table 1. Some examples of spectra in this category are displayed in Figure 9.

4.6. Problematic

This collection is very diverse as it assembles together spectra which were in some peculiar and generally unwanted way affected in either of the stages from observations to reduction. *Emission spikes* (6768 spectra) are most often present in IR band and sometimes, but less pronounced, in the red band. The left and right part of the map, shaped by low density snake-like collections, represent spectra with one *strong emission spike* (12161) in IR band. A reduction issue in the form of an *oscillating continuum* (2914) in the red band is also very numerous. The *negative flux* (2192) is most often present in the IR band, followed by the blue band and less often in the red band. There is one more quite interesting but less frequent reduction effect in the IR band, in the form of very *low continuum* (54) which is either at ~ 0.3 or close to and below zero level, often accompanied by strong oscillating features. The described subcategories follow each other in Figure 10 from top to bottom in each panel, with the zero level spectrum in IR band overlapping the previous one in the bottom panel (flux level 6). A larger part of spectra in this category is in all aspects well behaved apart from the described issues, and their automatic detection without eye inspection is very helpful for the iterative development and improvement of our reduction pipeline (Kos et al. 2016).

5. Specific search for young/active stars

We present additional classification results based on a more specific projection map, in contrast to the general one presented in Section 4. These results follow the exact same procedure as explained in Section 3.3, but with different t-SNE input parameters and input spectral ranges. The motivation for this classification channel is the search for stars in their early phases of evolution (Žerjal et al. 2013) for which features in Hα, Hβ, and ^{7}Li spectral lines can be diagnostic of their activity (Soderblom 2010; Jeffries 2014). Perplexity is set to 50 and the spectral ranges $4841 - 4881$ Å (Hβ) and $6543 - 6583$ Å (Hα) are selected for the first t-SNE projection of the whole working set, while perplexity of 15 and only narrow spectral ranges $(4859 - 4863, 6561 - 6565, 6706 - 6710$ Å) around the three diagnostic lines are selected for the second t-SNE projection of the filtered working set. Other variations of perplexity and spectral ranges were tried, but this one produced the most useful projection map.

Compared to the general classification from the previous section, we find additional candidates in categories of *binary stars* (649), *problematic* spectra with oscillating continuum (1034), and *Hα/Hβ emission* (1083). With this projection map, we are able to partition the latter category and identify four distinct morphological subtypes (see Figures 11 and 12): *Hα/Hβ emission*, *Hα/Hβ emission superimposed on absorption*, *Hα/Hβ P-cygni*, and *Hα/Hβ inverted P-cygni*, all indicative of diverse underlying physical processes (Traven et al. (2015) and references therein). The *Hα/Hβ emission* category is a counterpart to the one presented in the previous section, and contains diverse multicomponent profiles of Hα/Hβ emission lines, that were not cleanly separated in the projection map, as in the case of the latter three categories. It is possible that some of the emission profiles are a consequence of reduction issues instead of intrinsic properties of stars and their environment, but such an evaluation is out of scope of the current work, and will be addressed in future classification studies.

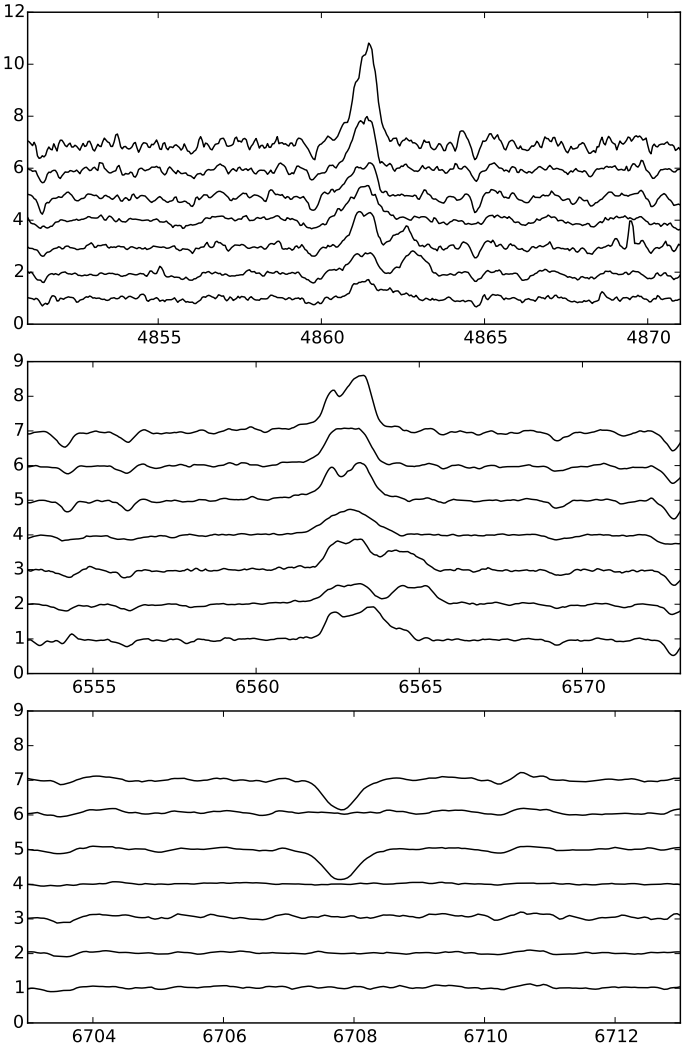


Fig. 11: Same as Figure 5 but based on results from the specific search for young/active stars. Examples for *Hα/Hβ emission* category are displayed.

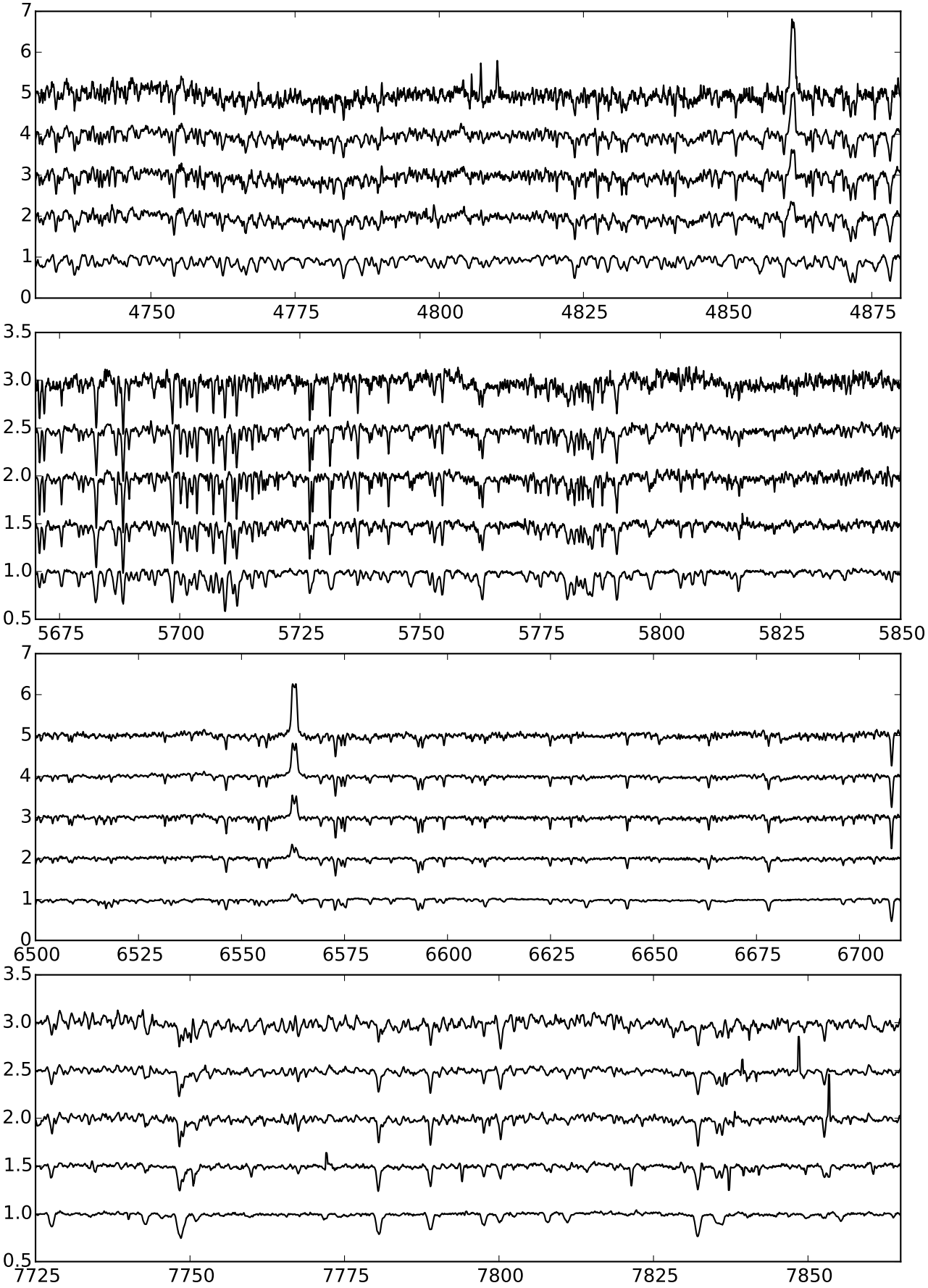


Fig. 9: Same as Figure 5 but for *Hα/Hβ emission*.

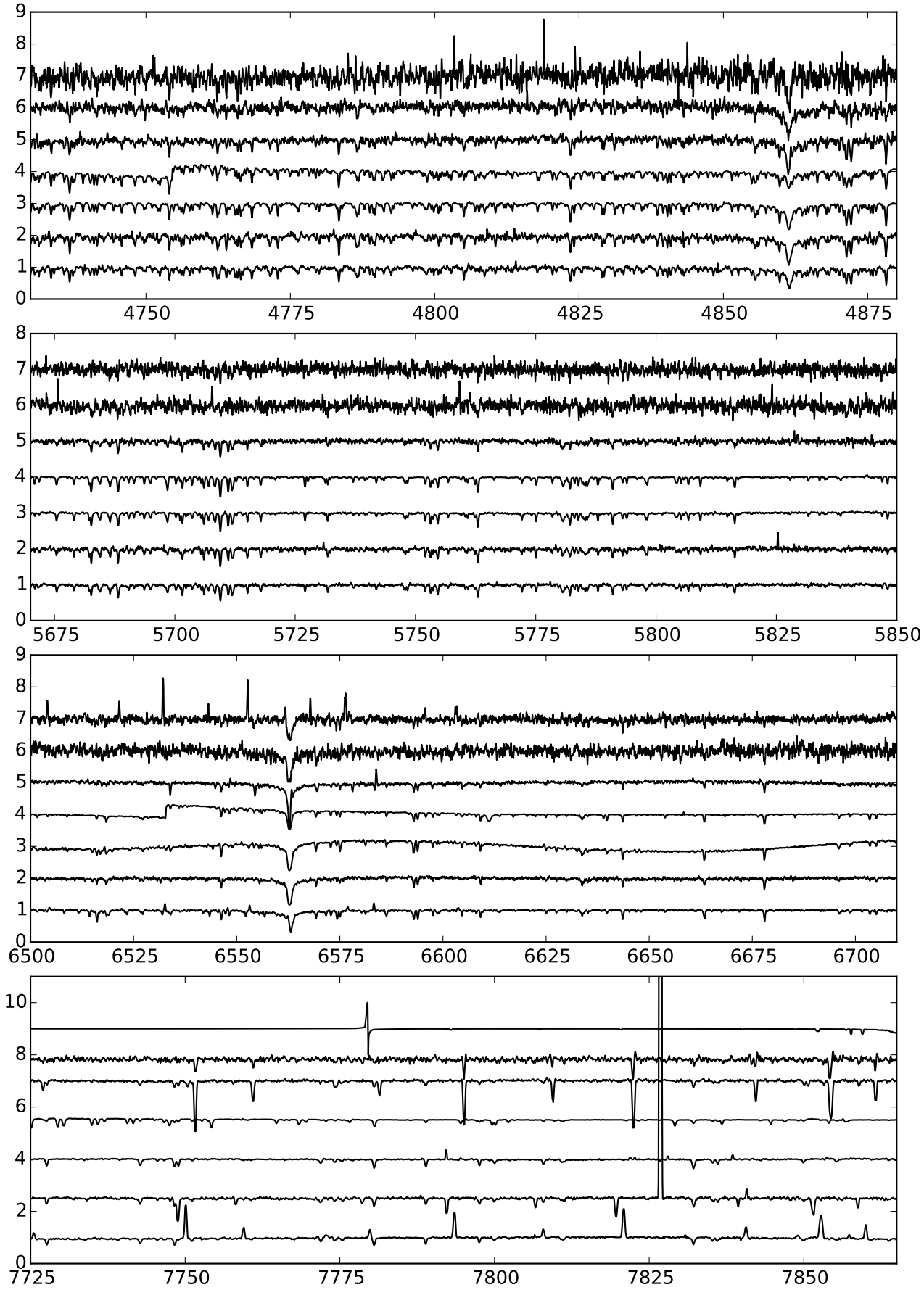


Fig. 10: Same as Figure 5 but for *problematic spectra*. Subcategories of spectra in the panels from bottom to top: emission spikes, strong emission spike, oscillating continuum, oscillating continuum, negative flux, low continuum, and low continuum.

Table 1: Classification categories based on the general projection map (see Section 4). The columns give the classification category, number of classified spectra, and most common SIMBAD *main types* and *other types*. SIMBAD defines a *main type* for each astronomical object in its database, and several *other types* generally inferred from its identifiers. For the last two columns, only first five most common types are listed here, excluding the less interesting type *Star*. Full table is available in electronic version.

Category	N	Main type	Other types
Hot stars	5101	Star in Cluster (34), Variable Star of delta Sct type (18), Variable Star of RR Lyr type (12), Variable Star (11), Eclipsing binary of Algol type (detached) (10)	Infra-Red source (1913), Star in Cluster (52), Variable Star (47), Double or multiple star (26), Variable Star of delta Sct type (24)
Cool metal-poor giants	3532	Star in Cluster (202), Red Giant Branch star (99), High proper-motion Star (16), Possible Red Giant Branch star (15), Variable Star of RR Lyr type (10)	Infra-Red source (548), Star in Cluster (384), Red Giant Branch star (109), Possible Red Giant Branch star (52), High proper-motion Star (19)
Molecular abs. bands	1728	Star in Cluster (5), Variable Star (4), Semi-regular pulsating Star (3), Red Giant Branch star (1), S Star (1)	Infra-Red source (636), Variable Star (10), Star in Cluster (6), Semi-regular pulsating Star (3), Possible Red supergiant star (2)
Binary stars	2229	Star in Cluster (10), Spectroscopic binary (4), Eclipsing binary of Algol type (detached) (2), Double or multiple star (2), Eclipsing binary (2)	Infra-Red source (325), Star in Cluster (16), Double or multiple star (12), Spectroscopic binary (8), Rotationally variable Star (5)
Hα/Hβ emission	348	High proper-motion Star (10), Pre-main sequence Star (8), Rotationally variable Star (7), T Tau-type Star (7), Star in Cluster (4)	Infra-Red source (69), X-ray source (27), High proper-motion Star (21), Pre-main sequence Star (20), Variable Star (16)
Problematic[1]	26251	Star in Cluster (345), Red Giant Branch star (34), High proper-motion Star (19), Variable Star of RR Lyr type (10), Variable Star (8)	Infra-Red source (1953), Star in Cluster (405), Red Giant Branch star (40), High proper-motion Star (23), Variable Star (23)

A new category *lithium absorption* is defined to account for spectra which display different equivalent widths of the ^{7}Li line, from weak to very strong absorptions, as displayed in Figure 13. Significant ^{7}Li absorption sometimes accompanies spectra in Hα/Hβ emission categories from the previous paragraph, as evident from Figure 11.

The classification categories presented in this section are listed in Table 3, along with SIMBAD classes which indicate the connection between youth and activity of stars and the observed Hα/Hβ multicomponent profiles and prominent lithium absorption.

6. Catalogue

The final classification results are gathered in the catalogue, whose contents are described in Table 4. For all 41605 spectra assigned with at least one classification category from either Section 4 or 5, we give the coordinates of corresponding targets, their APASS (Henden et al. 2012; Munari et al. 2014) V magnitude, classification category, and supplementary information from SIMBAD, VizieR, and ADS databases.

As the coordinates of stars are the most reliable search parameters, we used them to retrieve information from SIMBAD, VizieR, and ADS on-line databases. Epoch 2000.0 coordinates of our targets are not identical with those from the catalogues, so we adopt a search radius of 1 arcsec where applicable. The results of the search in VizieR catalogues are retrieved based on the following wavelength ranges: Gamma-ray, X-ray, EUV, UV, Optical, IR, and Radio. In the catalogue, we list the number of VizieR tables in which a match is found. References from the literature (ADS) should serve as possible additional information about objects of interest, they are not to be taken as reliable sources of the characteristics of a certain object.

Some of the 38633 stars (with 41605 spectra) in our catalogue are known to be of peculiar type and are already discussed in the literature or listed in different sources. SIMBAD yields a match for 5998 targets, VizieR expectedly finds at least one match in at least one of its catalogues for practically all unique targets (38632) while 1639 targets are matched successfully with references from the ADS database.

The electronic version of the catalogue will be made publicly available at the CDS.

7. Visualisation - t-SNE Explorer

The t-SNE or *Galah Explorer* is an interactive web application, visualising the "feature" based distribution of spectra in the t-SNE projection map. The basic view contains:

t-SNE map in the central part such as those in Figures 2 and 3. The map is split into hexagons which are color coded based on average values of parameters of contained datapoints.

Large hexagonal frame displaying datapoints of a selected hexagon from the map, where each datapoint is color coded depending on the selection of available numerical or descriptive (e.g. classification) parameters.

[1] A large fraction of such spectra are recoverable (see text).

Table 3: Same as Table 1 but for the specific projection map produced in the search for young/active stars (see Section 5). Categories *binary stars*, *Hα/Hβ emission*, and *problematic* are already defined in Section 4, while the other four are uniquely described in Section 5.

Category	N	Main type	Other types
Binary stars	1730	Star in Cluster (11), Variable Star of RR Lyr type (4), Eclipsing binary of Algol type (detached) (3), Spectroscopic binary (2), Rotationally variable Star (1)	Infra-Red source (244), Star in Cluster (16), Variable Star (9), Spectroscopic binary (7), Double or multiple star (6)
Hα/Hβ emission	212	Pre-main sequence Star (5), Rotationally variable Star (3), High proper-motion Star (3), Flare Star (3), Star in Cluster (2)	Infra-Red source (40), X-ray source (20), Pre-main sequence Star (9), High proper-motion Star (8), Variable Star (8)
Hα/Hβ emission superimposed on absorption	549	Star in Cluster (5)	Infra-Red source (8), Star in Cluster (5)
Hα/Hβ P-cygni	85	Variable Star (1)	Infra-Red source (3), Variable Star (1), Rotationally variable Star (1)
Hα/Hβ inverted P-cygni	412		Infra-Red source (1)
Lithium absorption	414	Rotationally variable Star (9), T Tau-type Star (9), Pre-main sequence Star (7), Variable of BY Dra type (5), Variable of RS CVn type (2)	Infra-Red source (130), X-ray source (32), Variable Star (25), Pre-main sequence Star (22), T Tau-type Star (20)
Problematic	2767	Star in Cluster (109), Possible Red Giant Branch star (1), Spectroscopic binary (1), High proper-motion Star (1)	Infra-Red source (346), Star in Cluster (115), Double or multiple star (6), X-ray source (2), Spectroscopic binary (2)

Table 4: Description of the content for the catalogue with 41605 entries (spectra) whose full table will be available at the CDS.

Label	**Unit**	**Description**
DATEOBS		Date and time of the observation
RA	°	RA (J2000)
DEC	°	DEC (J2000)
Class_cat_general		General classification category as given in Section 4
Class_cat_specific		Specific classification category as given in Section 5
SIMBAD_main_id		Main ID of the source in SIMBAD
SIMBAD_angular_distance	arcsec	Angular distance of Galah target to the source in SIMBAD
SIMBAD_main_type		SIMBAD *main type*
SIMBAD_other_types		SIMBAD *other types*
VizieR_n_Radio		Number of VizieR tables for Radio wavelength range in which Galah target has a match
VizieR_n_IR		As VizieR_n_Radio but for IR wavelength range
VizieR_n_optical		As VizieR_n_Radio but for optical wavelength range
VizieR_n_UV		As VizieR_n_Radio but for UV wavelength range
VizieR_n_EUV		As VizieR_n_Radio but for EUV wavelength range
VizieR_n_Xray		As VizieR_n_Radio but for X-ray wavelength range
VizieR_n_Gammaray		As VizieR_n_Radio but for Gamma-ray wavelength range
ADS_literature		A comma-separated list of articles (title and bibcode)

List of parameters which can be chosen along with individual values for all of them displayed for the currently selected datapoint. Any supplementary information on the corresponding object (star) can also be displayed together with a link to Simbad and Vizier matches.

Plotting area with four panels corresponding to four Galah spectral bands, where the median and the dispersion of normalised fluxes of all spectra of the currently selected hexagon is displayed, over-plotted with the currently selected datapoint (spectrum).

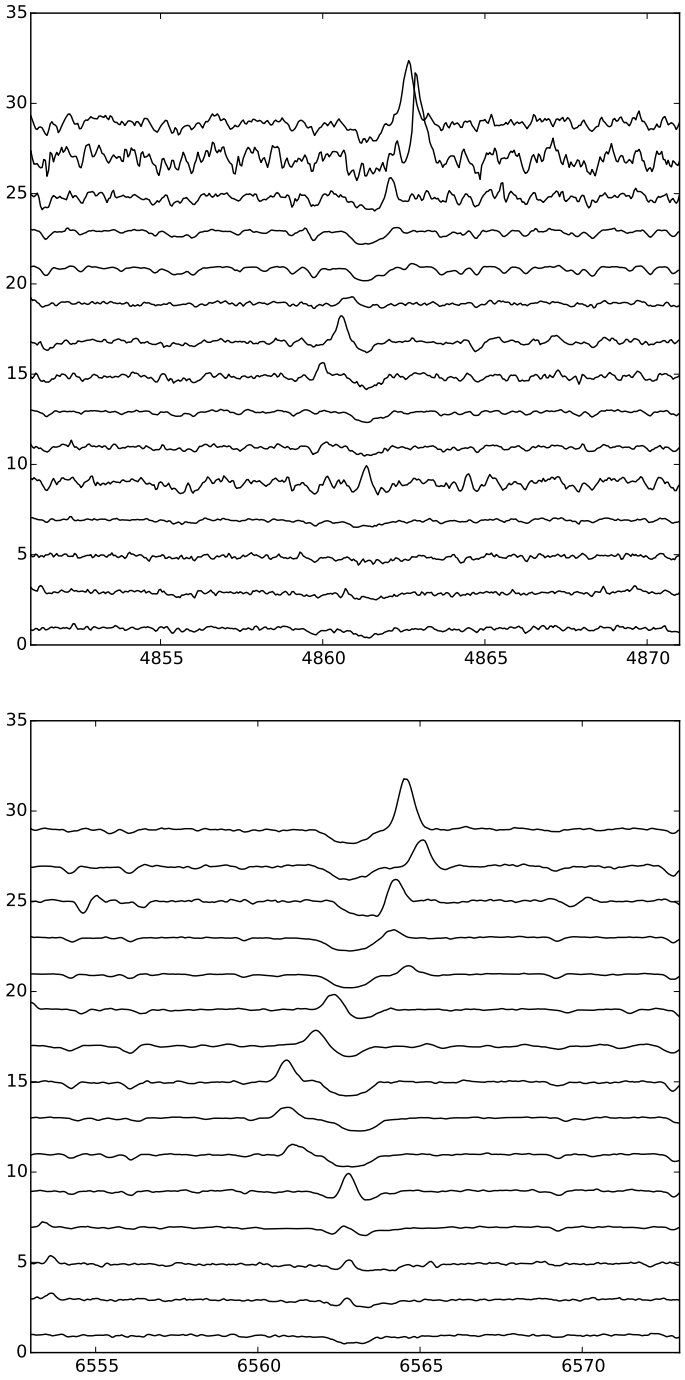


Fig. 12: Same as Figure 11 but for *Hα/Hβ emission superimposed on absorption*, *Hα/Hβ inverted P-cygni*, and *Hα/Hβ P-cygni* categories. From bottom to top in panels, each category features 5 examples of spectra.

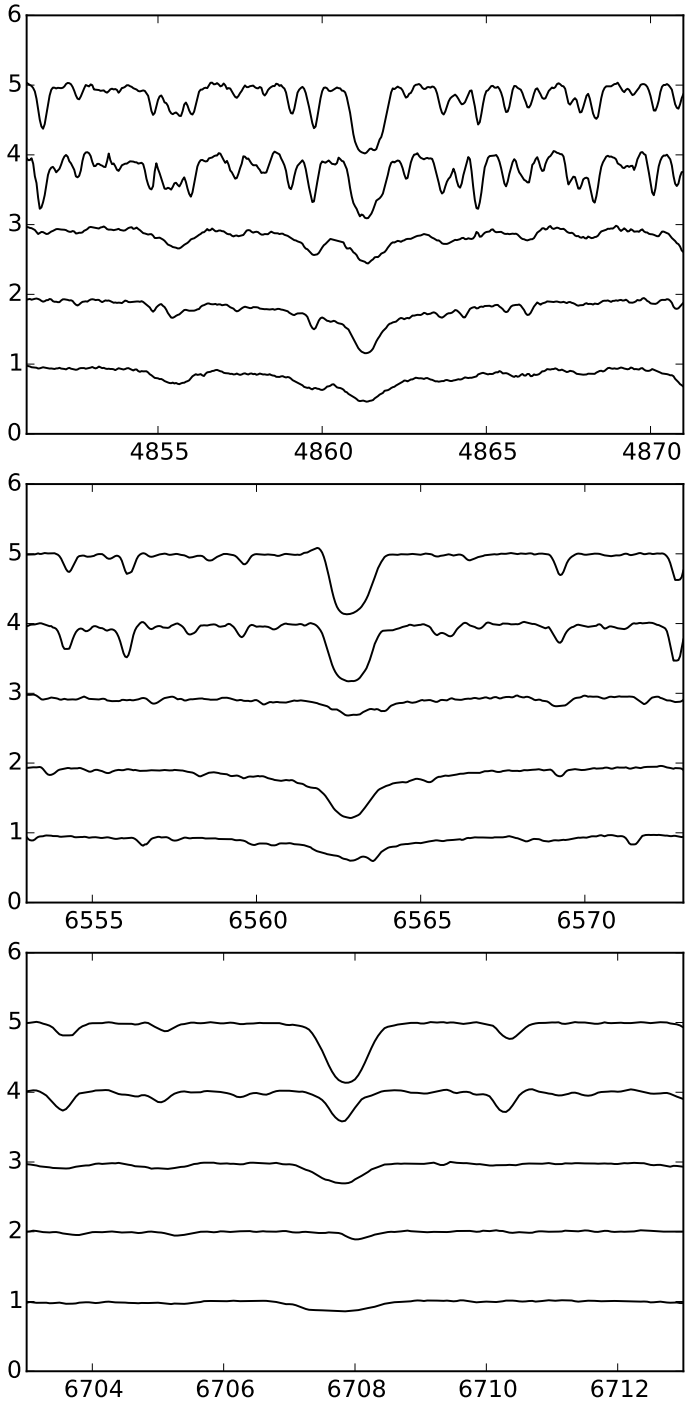


Fig. 13: Same as Figure 11 but for *lithium absorption*.

Search fields where the user can search by Galah identifier or other parameter value or label which is available in the Galah database of reduced spectra, whereby immediately selecting and displaying the matching object.

The presented segments of the *Galah Explorer* enable the user to locate specific areas of interest in the map, featuring characteristic values of parameters. One can also go the other way by looking for a specific object using e.g. its unique identifier, and once selected inspecting its morphological vicinity (parent hexagon) with its neighbouring spectra, using statistical plots to evaluate their similarity.

The (briefly) described functionality offers a very powerful and useful way of overviewing any kind of a dataset, locating and exploring its inherent structure in feature space and detecting all kinds of outliers. Many different projection maps can be incorporated into the *Galah Explorer* together with different DBSCAN modes for an efficient selection and identification of distinct morphological collections of spectra. The tool is currently available on the Galah official website at `http://galah-survey.org`.

8. Discussion

We have demonstrated that t-SNE can be used as an efficient tool for discovery of diverse spectral features and classification of stellar spectra. By projecting the Galah dataset onto a two dimensional space it is able to preserve and visually reveal its complex morphological structure. Although not tested on our sample of spectra, it was shown by van der Maaten & Hinton (2008) that t-SNE is far superior in its domain, putting emphasis on (1) modeling dissimilar datapoints by means of large pairwise distances, and (2) modeling similar datapoints by means of small pairwise distances, which is not obvious for other non-linear dimensionality reduction techniques and even less so for the linear ones.

The complexity of spectral morphologies in principle increases with increasing wavelength range, therefore Galah with its four spectral bands in this respect surpasses many other spectroscopic surveys. Consequently, the task of classification is more difficult because spectral features can appear differently and from different effects in each band (wavelength-dependent markers of physical processes, reduction issues, hardware malfunction, different optical paths, etc.).

Our classification procedure can accept any arbitrary spectral ranges selected by the user, which (1) enables us to put emphasis on particular physics we are interested in, and (2) removes the possibly unwanted influence from strong features in other parts of the spectrum, which can impede classification of the desired types of objects. Both points were demonstrated with the specific projection in Section 5, selecting only narrow regions of Hα, Hβ, and ^{7}Li for the search of young, active stars. Many detections of such objects were not possible with the first projection as the full spectral information along with strong problematic features in the e.g. IR band clouded those in other bands. For the same reason, we might be missing some interesting morphological categories with weaker characteristic features hidden by stronger ones. Therefore, the classification presented here is not absolutely representative of the whole dataset, however, the most prominent features are recognised, amounting to 10 distinct classification categories listed in Tables 1 and 3, containing altogether 41605 spectra (38633 unique targets). The specific projection map yielded new candidates for three already defined categories from the general map, while also providing four new categories, validating the principle that one can try different t-SNE set-ups to select the best (or several) projection maps for classification purposes.

Using more than one projection map, possibly produced by different input selected spectral regions, or simply using different DBSCAN modes, it can happen that the same spectrum is assigned to more than one classification category, due to the previously discussed reasons. Additional factor contributing to such cases are morphologically similar features like double lines from binary stars or emission superimposed on absorption, which might be located close in the projection space, with a possible overlap region. Spectra with more than one classification category can be easily identified in the catalogue (Table 4), having both classification fields (general as well as specific) non-empty.

The novel dimensionality reduction technique t-SNE is capable of representing astronomical spectra in a low dimension space where their morphology and hidden features can be efficiently discovered and studied. This was shown with a very effective classification of the largest astronomical high-resoluton spectroscopic dataset so far, comprising 209, 533 spectra where each one contains 13, 600 values of flux. All data products along with the t-SNE Explorer will be publicly available in the coming data releases. The source code of the employed methods is freely available on-line, and our custom procedure for classification can be readily adjusted and applied to different spectroscopic or other astronomical datasets. The knowledge and experiences gained in this work will facilitate further investigation and understanding of the Galah dataset with newly arrived spectra and enable focused studies of distinct categories of objects (e.g. binary stars).

Although this work has made use of the external sources, it does not depend on them and they serve mostly to support this proof of concept for classification of a wide variety of astronomical data.

Acknowledgements. This research has made use of the SIMBAD database (Wenger et al. 2000) and of the VizieR catalogue access tool operated at CDS, Strasbourg, France. The original description of the VizieR service was published in A&AS 143, 23.

References

Bekki, K., Campbell, S. W., Lattanzio, J. C., & Norris, J. E. 2007, MNRAS, 377, 335
Dalton, G., et al. 2012, in Proc. SPIE, Vol. 8446, Ground-based and Airborne Instrumentation for Astronomy IV, 84460P
de Jong, R. S., et al. 2012, in Proc. SPIE, Vol. 8446, Ground-based and Airborne Instrumentation for Astronomy IV, 84460T
De Silva, G. M., et al. 2015, MNRAS, 449, 2604
Freeman, K., & Bland-Hawthorn, J. 2002, ARA&A, 40, 487
Freeman, K. C. 2012, in Astronomical Society of the Pacific Conference Series, Vol. 458, Galactic Archaeology: Near-Field Cosmology and the Formation of the Milky Way, ed. W. Aoki, M. Ishigaki, T. Suda, T. Tsujimoto, & N. Arimoto, 393
Gilmore, G., et al. 2012, The Messenger, 147, 25
Henden, A. A., Levine, S. E., Terrell, D., Smith, T. C., & Welch, D. 2012, Journal of the American Association of Variable Star Observers (JAAVSO), 40, 430
Hinton, G., & Roweis, S. 2002, in Advances in Neural Information Processing Systems 15 (MIT Press), 833–840
Jeffries, R. D. 2014, in EAS Publications Series, Vol. 65, EAS Publications Series, 289–325
Kos, J., et al. 2016, MNRAS
Mason, B. D., Wycoff, G. L., Hartkopf, W. I., Douglass, G. G., & Worley, C. E. 2001, AJ, 122, 3466
Matijevič, G., et al. 2010, AJ, 140, 184
Munari, U., Henden, A., Frigo, A., & Dallaporta, S. 2014, Journal of Astronomical Data, 20
Ness, M., Hogg, D. W., Rix, H.-W., Ho, A. Y. Q., & Zasowski, G. 2015, ApJ, 808, 16
Piskunov, N., & Valenti, J. A. 2016, ArXiv e-prints
Pourbaix, D., et al. 2004, A&A, 424, 727
Prusti, T. 2012, Astronomische Nachrichten, 333, 453
Soderblom, D. R. 2010, ARA&A, 48, 581
Steinmetz, M., et al. 2006, AJ, 132, 1645
Traven, G., et al. 2015, A&A, 581, A52
Žerjal, M., et al. 2013, ApJ, 776, 127
Valenti, J. A., & Piskunov, N. 1996, A&AS, 118, 595
van der Maaten, L. 2013, ArXiv e-prints
van der Maaten, L., & Hinton, G. 2008, Journal of Machine Learning Research, 9, 2579
Watson, F. G. 1987, PhD thesis, Edinburgh Univ. (Scotland).
Wenger, M., et al. 2000, A&AS, 143, 9

Table 2: List of spectra plotted in Figures 5−13 representing distinct classification categories. The columns give coordinates, APASS V magnitude where available, figure number, and sequential spectrum number (bottom to top) in panels for each figure.

RA (J2000)	DEC (J2000)	V	N fig.	N spec.
12h 40m 59.46s	-45° 27′ 14.1″	12.2	5	1
10h 59m 01.3s	-47° 34′ 19.9″		5	2
12h 59m 40.48s	-45° 53′ 04.4″		5	3
09h 00m 57.77s	-27° 09′ 39.8″		5	4
21h 03m 00.31s	-49° 26′ 34.6″		6	1
21h 40m 02.12s	-23° 13′ 00.2″	13.0	6	2
01h 29m 21.27s	+06° 47′ 25.4″	10.9	6	3
19h 22m 58.05s	-22° 54′ 40.2″	10.1	7	1
16h 52m 28s	-24° 21′ 45.4″	10.5	7	2
16h 13m 22.82s	-17° 29′ 54.7″	11.9	7	3
13h 23m 49.52s	-49° 14′ 21.1″		7	4
09h 43m 55.26s	-76° 28′ 55.5″	13.4	8	1
18h 39m 12.45s	-53° 58′ 07.6″	13.3	8	2
06h 13m 09.26s	-54° 49′ 54.4″	12.6	8	3
11h 34m 04.8s	-39° 17′ 39.5″	12.7	8	4
06h 30m 58.37s	-31° 49′ 29.7″	13.6	8	5
14h 05m 59.68s	-45° 18′ 20.7″		8	6
16h 13m 29.29s	-23° 11′ 07.6″	11.7	9	1
02h 04m 32.8s	-74° 55′ 28.6″	13.2	9	2
12h 20m 52.77s	-49° 11′ 05.7″		9	3
12h 43m 04.53s	-49° 31′ 11″	13.1	9	4
16h 01m 22.34s	-19° 37′ 22.3″	14.5	9	5
07h 41m 05.11s	-55° 26′ 32.54″	13.8	10	1
19h 42m 39.47s	-45° 08′ 14.4″	13.9	10	2
09h 56m 53.13s	-70° 57′ 57.8″		10	3
22h 20m 59.59s	-08° 59′ 13.3″		10	4
16h 01m 54s	-22° 12′ 18.9″	14.3	10	5
12h 41m 51.89s	-36° 50′ 37.3″		10	6
18h 23m 16.32s	-34° 01′ 27.5″		10	7
07h 37m 19.419s	-55° 31′ 44.49″	14.5	11	1
20h 57m 51.85s	-33° 52′ 38.4″	13.8	11	2
20h 08m 37.69s	-41° 26′ 45.8″		11	3
11h 14m 59.85s	-33° 22′ 27.8″	13.7	11	4
16h 09m 39.7s	-22° 00′ 46.6″	14.0	11	5
13h 04m 09s	-44° 49′ 18.7″	14.2	11	6
16h 03m 46.95s	-22° 45′ 24.8″	14.0	11	7
21h 28m 30.21s	-40° 14′ 30.5″	13.9	12	1
07h 50m 47.825s	-58° 43′ 17.31″	13.9	12	2
07h 45m 32.447s	-58° 50′ 26.46″		12	3
07h 35m 59.296s	-55° 03′ 09.68″	14.0	12	4
16h 29m 41.31s	-27° 27′ 27.4″	14.6	12	5
07h 33m 39.016s	-55° 34′ 54.83″	14.7	12	6
07h 41m 48.343s	-55° 24′ 54.13″		12	7
16h 27m 09.9s	-28° 21′ 58″		12	8
16h 28m 43.44s	-28° 20′ 24.1″	14.5	12	9
07h 42m 50.476s	-55° 26′ 22.95″	14.9	12	10
15h 59m 08.89s	-26° 03′ 19.6″	13.5	12	11
16h 09m 09.88s	-22° 10′ 13.2″	13.9	12	12
16h 27m 50.82s	-28° 58′ 28.8″		12	13
16h 28m 30.01s	-27° 16′ 57.8″	15.3	12	14
16h 28m 51.24s	-27° 53′ 15.8″	14.6	12	15
03h 44m 03.539s	+24° 30′ 15.14″	10.9	13	1
16h 16m 03.04s	-28° 02′ 46.4″	11.2	13	2
06h 22m 39.33s	-55° 07′ 13.1″	13.2	13	3
15h 38m 51.73s	-39° 54′ 46.3″	13.5	13	4
12h 09m 56.1s	-30° 50′ 05.8″	13.0	13	5