



**HAL**  
open science

# Sliced Inverse Regression: application to fundamental stellar parameters

Sarkis Kassounian, Marwan Gebran, Frederic Paletou, Victor Watson

## ► To cite this version:

Sarkis Kassounian, Marwan Gebran, Frederic Paletou, Victor Watson. Sliced Inverse Regression: application to fundamental stellar parameters. *Open Astronomy*, 2019, 28 (1), pp.68-84. <10.1515/astro-2019-0006>. <hal-02142318v2>

**HAL Id: hal-02142318**

**<https://hal.science/hal-02142318v2>**

Submitted on 3 Jul 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

## Research Article

Sarkis Kassounian, Marwan Gebran\*, Frédéric Paletou, and Victor Watson

# Sliced Inverse Regression: application to fundamental stellar parameters

<https://doi.org/10.1515/astro-2019-0006>

Received Sep 24, 2018; accepted Nov 20, 2018

**Abstract:** We present a method for deriving the stellar fundamental parameters. It is based on a regularized sliced inverse regression (RSIR). We first tested it on noisy synthetic spectra of A, F, G, and K-type stars, and inverted simultaneously their atmospheric fundamental parameters:  $T_{\text{eff}}$ ,  $\log g$ ,  $[M/H]$  and  $v \sin i$ . Different learning databases were calculated using a range of sampling in  $T_{\text{eff}}$ ,  $\log g$ ,  $v \sin i$ , and  $[M/H]$ . Combined with a principal component analysis (PCA) nearest neighbors (NN) search, the size of the learning database is reduced. A Tikhonov regularization is applied, given the ill-conditioning of SIR. For all spectral types, decreasing the size of the learning database allowed us to reach internal accuracies better than PCA-based NN-search using larger learning databases. For each analyzed parameter, we have reached internal errors that are smaller than the sampling step of the parameter. We have also applied the technique to a sample of observed FGK and A stars. For a selection of well-studied stars, the inverted parameters are in agreement with the ones derived in previous studies. The RSIR inversion technique, complemented with PCA pre-processing proves to be efficient in estimating stellar parameters of A, F, G, and K-type stars.

**Keywords:** methods: data analysis, methods: statistical, techniques: spectroscopic, stars: fundamental parameters

## 1 Introduction

Astronomical surveys, either spaceborne or ground-based, are gathering an unprecedented amount of data. One can mention the SDSS DR14 data (Abolfathi et al. 2018) that contains 154 TB of millions of spectroscopic and photometric data. The DR5 of the LAMOST survey (Cui et al. 2012) contains 9 million spectra in total. Gaia DR2 provides information about 1.3 billion stars (Katz & Brown 2017). These space and ground-based surveys quantify the size of the data the astronomical community will face in the near future.

Spectroscopic analysis is crucial for the derivation of fundamental stellar atmospheric parameters which are the effective temperature ( $T_{\text{eff}}$ ), the surface gravity ( $\log g$ ), and the metallicity ( $[M/H]$ ). In addition to these fundamentals, and because it may strongly affect the shape of the observed spectra, the projected equatorial rotational velocity,  $v \sin i$ , is also retrieved from spectroscopic information. Many authors have for long been using spectroscopic data to estimate the stellar atmospheric parameters (McWilliam 1990; Latham et al. 2002; Torres et al. 2002; Buchhave et al. 2012; Schönrich & Bergemann 2014; Dieterich et al. 2017; Fabbro et al. 2018). However in order to extract the most relevant and accurate information from high-resolution, and large bandwidths stellar spectra, still more endeavour is required.

**Corresponding Author: M. Gebran:** Department of Physics and Astronomy, Notre Dame University-Louaize, PO Box 72, Zouk Mikael, Lebanon; Email: mgebran@ndu.edu.lb

**S. Kassounian:** Department of Physics and Astronomy, Notre Dame University-Louaize, PO Box 72, Zouk Mikael, Lebanon; Email: sakokassounian@gmail.com

**F. Paletou:** Université Paul Sabatier, Observatoire Midi-Pyrénées, Cnrs, Cnes, IRAP, F-31400 Toulouse, France; Cnrs, Cnes, Institut de Recherche en Astrophysique et Planétologie, 14 av. E. Belin, F-31400 Toulouse, France; Email: frederic.paletou@univ-tlse3.fr

**V. Watson:** Université Paul Sabatier, Observatoire Midi-Pyrénées, Cnrs, Cnes, IRAP, F-31400 Toulouse, France; Cnrs, Cnes, Institut de Recherche en Astrophysique et Planétologie, 14 av. E. Belin, F-31400 Toulouse, France; Email: viktorwatson@gmail.com

Most of the traditional approaches and developed pipelines rely on standard procedures such as comparing an observed spectrum with a set of theoretical spectra (Valenti & Piskunov 1996; Morris et al. 2018). The requirement for advanced computational techniques arises from the generated large dimensionality of the data due to the wide wavelength coverage together with the high spectral resolution. Many new techniques are being developed. In Ness et al. (2015) and Casey et al. (2016), a data-driven approach is introduced (CANNON) for determining stellar labels (fundamental parameters and detailed stel-

lar abundances) from spectroscopic data. Their learning databases (LDB) are based on a subset of reference objects for which the stellar labels are known with high accuracy. Dimensionality reduction techniques are also developed and used, such as applying the Principal Component Analysis (PCA) for data reduction (see e.g. Jolliffe 1986). PCA has shown its effectiveness in inverting the fundamental stellar atmospheric parameters in several studies (Bailer-Jones et al. 1998; Re Fiorentin et al. 2007; Paletou et al. 2015 a,b; Gebran et al. 2016). Xiang et al. (2017) estimated the stellar atmospheric parameters as well as the absolute magnitudes and  $\alpha$ -elements abundances from the LAMOST spectra with a multivariate regression method based on kernel-based PCA. The LAMOST spectroscopic survey data has also been recently analyzed by Boeche et al. (2018) to invert stellar parameters and chemical abundances in which several combined approaches and techniques were compared. The authors developed a code called `SP_Ace` which utilizes nearest neighbor comparison and non-linear model fitting techniques. In Wilkinson et al. (2017), a spectral fitting code (`FIREFLY`) was developed to derive the stellar population properties of stellar systems. `FIREFLY` uses a  $\chi$ -squared minimization fitting procedure that fits stellar population models to spectroscopic data, following an iterative best-fitting process controlled by a Bayesian information criterion. Their approach is efficient to overcome the so-called “ambiguities” in the spectra. More recently, Gill et al. (2018) used wavelet decomposition to distinguish between noise, continuum trends, and stellar spectral features in the CORALIE FGK-type spectra. By calculating a subset of wavelet coefficients from the target spectrum and comparing it to those from a grid of models in a Bayesian framework, they were able to derive  $T_{\text{eff}}$ ,  $[M/H]$ , and  $v \sin i$  for these stars. Ting et al. (2018) presented `The Payne`, a general method for the precise and simultaneous determination of numerous stellar labels from observed spectra. Using a simple neural-net-like functional form and a suitable choice of training labels, `The Payne` yields a spectral flux prediction good to  $10^{-3}$  rms across a wide range of  $T_{\text{eff}}$  and  $\log g$ . Ting et al. (2018) applied this approach to the APOGEE DR14 data set and obtained precise elemental abundances of 15 chemical species. In the same context, Fabbro et al. (2018) applied a deep neural network architecture to analyse both SDSS-III APOGEE DR13 and synthetic stellar spectra. Their convolutional neural network model, `StarNet`, was able to predict precise stellar parameters when trained on APOGEE spectra or on synthetic data.

In this study, we apply techniques such as, reduction of dimensionality with PCA, and a PCA-based nearest neighbor search (Paletou et al. 2015 a,b; Gebran et al. 2016) complemented with a Regularized Sliced Inverse Regression

(Bernard-Michel et al. 2009, 2007) (RSIR) procedure in order to derive simultaneously  $T_{\text{eff}}$ ,  $\log g$ ,  $[M/H]$  and  $v \sin i$  from spectra of A, F, G, and K-type stars. Up to now, sliced inverse regression has been rarely used in astronomy (Bernard-Michel et al. 2009; Watson et al. 2017). When combined with PCA-based techniques, the derivations of the fundamental atmospheric parameters are achieved with higher accuracy compared to the sole/mere PCA-based nearest neighbor inversion (Paletou et al. 2015 a,b; Gebran et al. 2016). The mathematical description of our method is detailed in Sec. 2. Following this, Sec. 3 describes all elements used for the enhancement of the computational abilities of SIR. Sec. 4 discusses the application of the technique on synthetic spectra for A, F, G, and K-type like stars. In Sec. 5, we show the results of inversions of real stars. Discussion and conclusion are gathered in Sec. 6.

## 2 Sliced inverse regression (SIR)

SIR, originally formulated by Li (1991), is a statistical technique that reduces multivariate regression to a *lower dimension*. It finds an inverse functional relationship between the response and the predictor which are the fundamental parameters and the flux respectively. Synthetic spectra flux values,  $x_{\text{syn}}$ , are usually calculated based on the set of stellar atmospheric parameters in the form of:

$$x_{\text{syn}} = f(T_{\text{eff}}, \log g, [M/H], v \sin i). \quad (1)$$

The inverse functional relation is used to predict the parameters of the observed flux values,  $x_{\text{obs}}$ , in the form of:

$$f^{-1}(x_{\text{obs}}) = (T_{\text{eff}}, \log g, [M/H], v \sin i). \quad (2)$$

In our work, we have derived a functional relationship for each parameter in the following way:

$$Y_j = f_j^{-1}(x_{\text{obs}}), \quad (3)$$

where  $j = 1, 2, 3, 4$  for  $T_{\text{eff}}$ ,  $\log g$ ,  $[M/H]$ , and  $v \sin i$ .

### 2.1 Global covariance matrix $\Sigma$

SIR starts with the computation of the covariance matrix  $\Sigma$  of all the synthetic spectra  $x_i$  of the LDB.

First, the spectra are gathered in a matrix of dimension  $N_{\text{spectra}} \times N_{\lambda}$ , where  $N_{\lambda}$  is the number of wavelength points per spectrum and  $N_{\text{spectra}}$  is the total number of spectra in the LDB. Then, the covariance matrix  $\Sigma$ , is defined as:

$$\Sigma = \frac{1}{N_{\text{spectra}}} \sum_{i=1}^{N_{\text{spectra}}} (x_i - \bar{x}) \cdot (x_i - \bar{x})^T, \quad (4)$$

where the global mean  $\bar{x}$  is defined as:

$$\bar{x} = \frac{1}{N_{\text{spectra}}} \sum_{i=1}^{N_{\text{spectra}}} x_i, \quad (5)$$

with  $x_i$  being a row vector containing the flux values of spectrum  $i$ .

## 2.2 Intra-slices covariance matrix

In SIR, all spectra are organized based on an increasing order of the considered parameter for inversion. For example, if we are to invert  $T_{\text{eff}}$  of each star, the spectra database should be organized in increasing order of  $T_{\text{eff}}$  while having the other parameters ordered randomly. We then build-up subsets of spectra, also called “slices”, having the same value of the parameter one wishes to determine first. These slices should not overlap each other (Li 1991). Then we calculate the means  $\bar{x}_h$  of the slice of the spectra found in each slice  $S_h$  that contains  $n_h$  synthetic spectra ( $h$  being the index of each slice). For the inversion of each parameter,  $\bar{x}_h$  and  $\bar{x}$  are used to calculate the “intra-slices” covariance matrix,  $\Gamma$ :

$$\Gamma = \sum_{h=1}^H \frac{n_h}{N} (\bar{x}_h - \bar{x})(\bar{x}_h - \bar{x})^T, \quad (6)$$

where

$$\bar{x}_h = \frac{1}{n_h} \sum_{x \in S_h} x_i. \quad (7)$$

## 2.3 Dimension reduction and parameter inversion

SIR aims to build a reducing subspace that *maximizes the variance between the slices while minimizing the variance within the slices* which creates a reduced predictor versus response regressive relationship to predict the parameters of the observed stars. This is applied by the process of stacking the spectra by an increase order of similar or close valued parameters and averaging them into a single spectra and projecting them on a new subspace. These new projection will later be used predictors of the functional relationship. Since the reduced projections are formed from spectra having close parameter values, this insures a higher accuracy of regressive predictions (Watson et al. 2017). On the other hand, slicing the spectra based on non-overlapping similar parameters insures this inter-slice maximization and intra-slice minimization.

The matrix  $\Sigma^{-1}\Gamma$  is then calculated where  $\Sigma$  and  $\Gamma$  are the two previously defined matrices. One eigenvector of

$\Sigma^{-1}\Gamma$ , called  $\beta_\lambda$  and corresponding to an eigenvalue  $\lambda$ , is used to form the reduction subspace. This will allow us to do regression in a 2-dimensional space using an inverse functional relationship. This relationship is constructed via a linear piecewise interpolation between the projection coordinates of the slices on the single eigenvector of  $\Sigma^{-1}\Gamma$  and the parameters.

The selection of  $\beta_\lambda$ , is based on a metric  $C_\lambda$  that quantifies the relationship between the spectra and the parameters.  $C_\lambda$ , defined as the “sliced inverse regression criteria” (Bernard-Michel et al. 2007, 2009), is calculated as follows:

$$C_\lambda = \frac{\beta_\lambda^t \Gamma \beta_\lambda}{\beta_\lambda^t \Sigma \beta_\lambda} \approx \frac{\text{Var}(\langle \beta_\lambda \cdot x_i \rangle)}{\text{Var}(\langle \beta_\lambda \cdot x_i \rangle) + \text{Var}(S_h)}, \quad (8)$$

where  $\beta_\lambda^t$  is the transpose of  $\beta_\lambda$  and  $\text{Var}$  is the variance function

$\beta_\lambda^t \Gamma \beta_\lambda$  is the “inter-slice” variance, whereas  $\beta_\lambda^t \Sigma \beta_\lambda$  represents the total variance. The  $\beta_\lambda$  that gives a  $C_\lambda$  value closest to 1 is considered as a the best choice for the reducing basis vector. In the present work,  $C_\lambda$  varies between 0.91 and 0.97 when using the eigenvector of  $\Sigma^{-1}\Gamma$  with the largest eigenvalue  $\lambda$ .

To invert the parameters, we apply linear piecewise interpolation on the coordinates of the projections of the  $\bar{x}_h$ -s on  $\beta_\lambda$ . Finally, the estimation of the parameters is made according to:

$$\hat{y} = \begin{cases} \bar{y}_1, & \text{if } x^p \in ]-\infty, \bar{x}_1^p], \\ \bar{y}_h + \left( x_{\text{obs}}^p - \bar{x}_h^p \right) \left( \frac{\bar{y}_{h+1} - \bar{y}_h}{\bar{x}_{h+1}^p - \bar{x}_h^p} \right) & \text{if } x^p \in ]\bar{x}_h^p, \bar{x}_{h+1}^p], \\ \bar{y}_H, & \text{if } x^p \in ]\bar{x}_H^p, +\infty[, \end{cases} \quad (9)$$

where  $\hat{y}$  is the estimated parameter;  $\bar{y}_h$  is the mean of the parameters of the spectra in slice  $h$ . The superscript “ $p$ ” represents the projected value of a selected set of data on  $\beta_\lambda$  i.e.,  $x^p = \langle \beta_\lambda \cdot x \rangle$ .

## 3 Enhancement of the computational abilities of SIR

In the present work, we are dealing with large amounts of high resolution spectra, so that  $\Sigma^{-1}\Gamma$  have typical dimension of  $\sim 10^4 \times 10^4$ . In addition, using a large LDB for SIR induces an increase in the intra-slice variance. This will lead to less accurate inverted parameters. Therefore to simultaneously address these problems, we applied two additional steps to SIR: first, using PCA, we reduce the dimension of every spectra in the LDB (Watson et al. 2017) from  $\sim 10^4$  to 12. Second, we apply a PCA-based NN-search in the re-

duced subspace to select a smaller LDB, more relevant for the spectra one wishes to analyze.

$\Sigma^{-1}\Gamma$  matrix is generally ill-conditioned. And the higher the condition number is, the more noise sensitive the system becomes (Kreyszig 2010). For the present work, values as large as  $10^{20}$  were found. In that case  $\beta_\lambda$  is very noise sensitive, leading to an unstable functional relationship. As a result, inaccurate inverted parameters may be derived. To solve this issue, we have applied Tikhonov regularization which aims to improve the conditioning of  $\Sigma^{-1}\Gamma$  and add a priori information to it based on the analysis of the noise of each observed spectrum. Several regularization methods exist, however, Tikhonov is very common and easily implemented. Other studies may address this issue, such as the truncated SVD used in Watson *et al.* (2017). Figure 1 summarizes the successive procedures we implemented, and that we discuss with more details hereafter.

### 3.1 LDB reduction via PCA

PCA is a numerical technique that allows for the reduction of dimension of each spectrum by projecting it on a set of orthogonal basis vectors called *principal components* (PC's). These components are the eigenvectors of the global covariance matrix  $\Sigma$ . Paletou *et al.* (2015 a) and Gebran *et al.* (2016) showed that for databases similar to the ones used in this study, only 12 PC's associated to the largest eigenvalues are enough to reduce the LDB, while the reconstruction error remains less than 1%. Therefore after this first pass, the new LDB has dimension of  $N_{spectra} \times 12$ .

The original LDB may reach to a dimension of  $N_{spectra} \times N_\lambda \simeq 10^6 \times 10^4$ . This is due to the fine sampling in the parameters, the high dimension of the spectra, and the large wavelength range which makes the process of SIR computationally heavy in terms of memory and time. To reduce the LDB which will be used for SIR, for each observed star, a PCA-based nearest neighbor search in the reduced subspace is applied (Paletou *et al.* 2015 a). This is done using the ‘‘PCA distance’’  $d_j^{(0)}$ , defined as:

$$d_j^{(0)} = \sqrt{\sum_{k=1}^{12} (\varrho_k - p_{jk})^2}, \quad (10)$$

where  $\varrho_k$  is the projection coordinate on the  $k^{th}$  dimension for an observed spectrum, and  $p_{jk}$  is the projection coefficient on the  $k^{th}$  dimension for the  $j^{th}$  synthetic spectrum. Finally for the SIR, a set of NN will be selected for each observed star as we will later describe in Sec. 3.3.

### 3.2 Tikhonov regularization

For the Tikhonov method (Vogel 2002), one usually inserts a regularization parameter  $\delta > 0$  into the ill-conditioned system, usually based on a priori information gathered by analyzing the noise of each observed spectrum. Considering the following matrix:

$$(\Sigma^2 + \delta I)^{-1} \Sigma \Gamma. \quad (11)$$

The eigenvector  $\beta_\lambda(\delta)$  associated to the largest eigenvalue of the matrix defined in Eq. 11 is calculated based on an optimization approach. For each parameter of each observed star, an optimum and specific  $\delta$  is calculated. This procedure is initiated by estimating the signal to noise ratio (S/N) of the observed spectrum using the procedure of Stoehr *et al.* (2008). Then a random set of synthetic spectra are selected from the LDB, and Gaussian white noise having the same S/N as the one of the observed spectrum is added to them. SIR is finally applied to this selected random set and the prediction of their parameters is done via the piecewise interpolation process described in Eq. 9. This simulated inversion leads to the selection of an optimum  $\beta_\lambda(\delta)$ .

$\delta$  is estimated by minimizing the difference between the newly inverted parameter values of the randomly selected noise added spectra ( $\hat{y}_i$ ) and their initial noiseless values ( $y_i$ ). The comparison is done using a normalized  $\chi^2$ :

$$\chi_N^2 = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2}}. \quad (12)$$

It was found that  $\log_{10}(\chi_N^2)$  as a function of  $\log(\delta)$  is a unimodal function which has a local minimum. This function is displayed in Figure 3 for a synthetic spectrum having  $T_{\text{eff}}$ ,  $\log g$ ,  $[M/H]$ ,  $v \sin i$  and S/N of 7 600 K, 2.50 dex, 0.0 dex, and 197 km s<sup>-1</sup>, and 196, respectively. The original LDB used in this example is the one of Gebran *et al.* (2016), explained in detail in Sec. 4. To find the minimum of these curves, we applied a golden-section search algorithm (Kiefer 1953). It is a classical numerical technique that minimizes unimodal functions which have a global minimum. The inversion process for each analyzed spectrum, and each parameter, has its own  $\chi_N^2 = f(\delta)$  that needs to be minimized.

### 3.3 Integrated scheme of the enhancements

Now that we have described the tools that were used to improve the SIR procedure, in what follows we discuss how these techniques are integrated to increase the accuracy of the inversion process. The flowchart in Figure 1 summarizes our adopted approach.

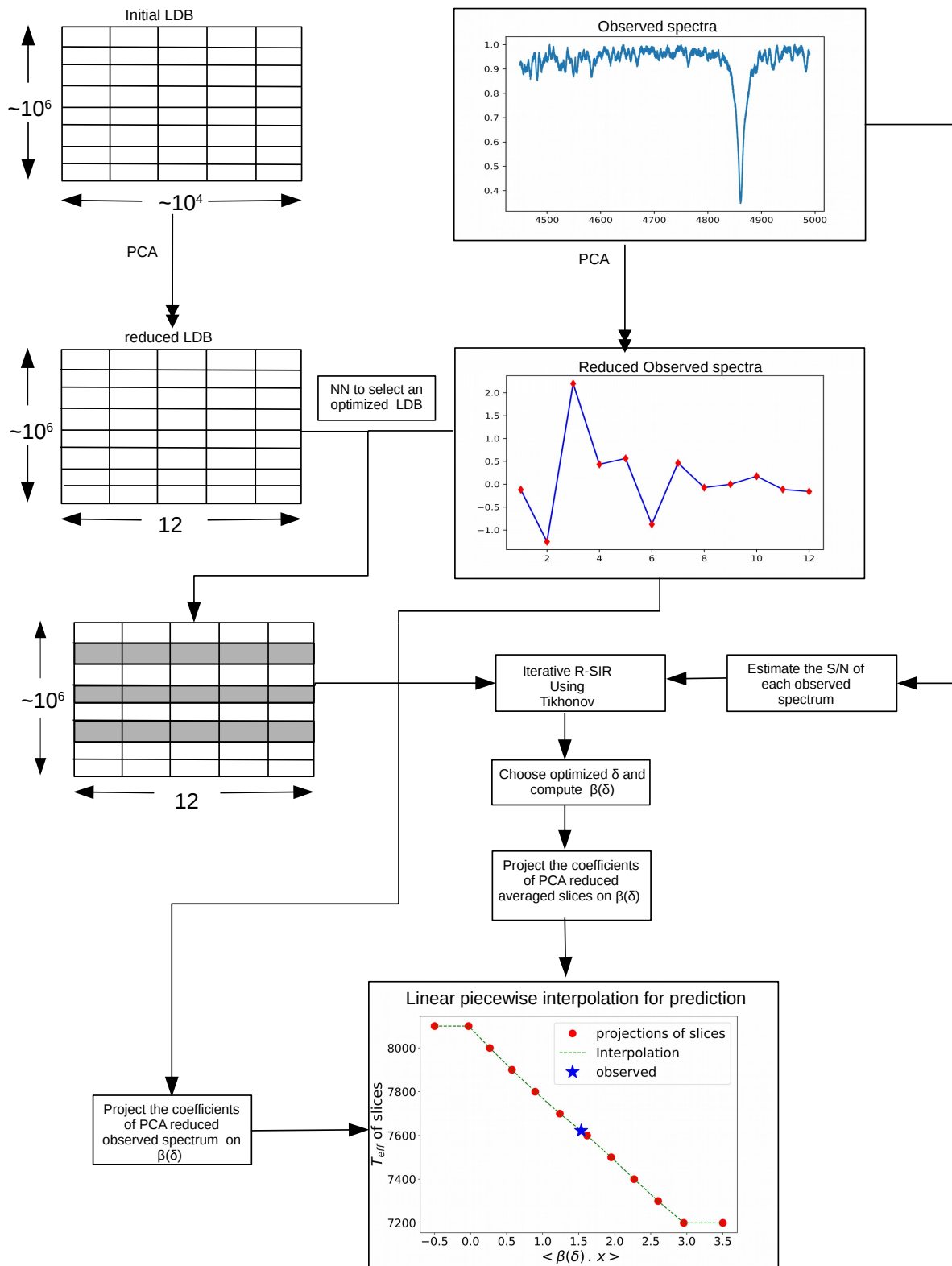


Figure 1. The flowchart of the procedure. The numerics in this figure are for the inversions of the test described in Sec. 3.

In our work,  $\Sigma$  and  $\Sigma^{-1}\Gamma$  have reached dimensions of the order of  $\sim 10^4 \times 10^4$ . For that reason and before applying the SIR process for each spectrum to be analyzed, we have reduced the dimension of these matrices by reducing the size of the original LDB using PCA as described in sec. 3.1.

During SIR, at least two distinct parameter values for each slice are required to construct the functional relationship. Therefore to select the optimum reduced LDB, a test for the construction of this relationship is required. Iteratively we tested for the number of distinct parameters by increasing the number of spectra of the nearest neighbors. Whenever the values of the distinct concerned parameters become greater or equal to 2, the iteration breaks and the inversion proceeds to the interpolation. During the tests, there were situations where  $[\Sigma^2 + \delta I]^{-1}$  was singular. This iterative approach solved this problem by adding nearest neighbors. Generally, using a smaller LDB which contain only a set of closest spectra to the observed one theoretically insures the success of SIR compared to using the entire original LDB. When selecting a set of nearest neighbors, we insure a lower minimization value of the intra-slice variance  $Var(S_h)$  in Eq. 8. Now within each slice the spectra are closer to each other and they are closer to the average spectrum of the slice. At the same time, choosing these optima reduced LDB's overcomes the issue of the degeneracies. In the PCA based NN-search (Paletou *et al.* 2015 a; Gebran *et al.* 2016), we had cases where the  $d_j^{(0)}$  were extremely close to each or even equal, with a variety of parameters. In SIR, we do not face such issue because the value are regressed for each parameter and the synthetic spectra with similar or close parameters are averaged to a single slice.

Now  $\Sigma^{-1}\Gamma$  has a dimension of  $12 \times 12$  and is constructed from the optimized reduced LDB. Its high condition number implies that it is ill-conditioned (see the example of the left panel of Figure 2). Therefore to improve the inversion process for each observed spectrum, we apply the Tikhonov regularization in SIR for our selected optima reduced LDB, iteratively. By applying this regularization, we are effectively taking advantage of the S/N ratio analysis and inserting the propagated noise information as a priori. In other words, we are applying an denoising procedure.

In Figure 2, we display the inversion results for  $T_{\text{eff}}$  of a noisy synthetic spectrum. This spectrum has a  $T_{\text{eff}}$  value of 7600 K with an added Gaussian white noise of  $S/N = 196$ . As we iterate over different sizes of optimized reduced LDB, a convergence is achieved in every case. For all of our tests, we noticed that the number of spectra in the optimized reduced LDB did not surpass 500. It is shown in this figure that the condition number of the non-regularized matrix is  $\sim 5$  orders of magnitude larger than the ones in

which the Tikhonov regularization was applied. The right panel displays the effect of the regularization on the inverted parameter ( $T_{\text{eff}}$ ) of the same spectrum. It is clearly shown that whatever the number of the nearest neighbors in the optimized reduced LDB is, inversion is achieved with higher accuracy than the one without regularization. The convergence occurs irrespectively of the value of the condition number, as long as it is smaller than the one without Tikhonov regularization.

Figure 3 represents the minimization of the  $\log_{10}(\chi_N^2)$  as a function of  $\log_{10}(\delta)$  for different sets of optimized reduced LDB. This figure shows the unimodal nature of the curves irrespectively of the size of the LDB.

## 4 Simulations and tests

In this section, we present the implementation and results of RSIR for two different sets of synthetic spectra. We also compare these results to the ones of the PCA NN-search to show the improvement in the accuracies of the derived parameters. To each of these spectra, white Gaussian noise was added with a random S/N. The spectra were calculated in the range of A to K type stars. The reason for selecting this spectral range is that in Sec. 5, we apply this procedure to a sample of the observed stars studied in Paletou *et al.* (2015 a) and Gebran *et al.* (2016).

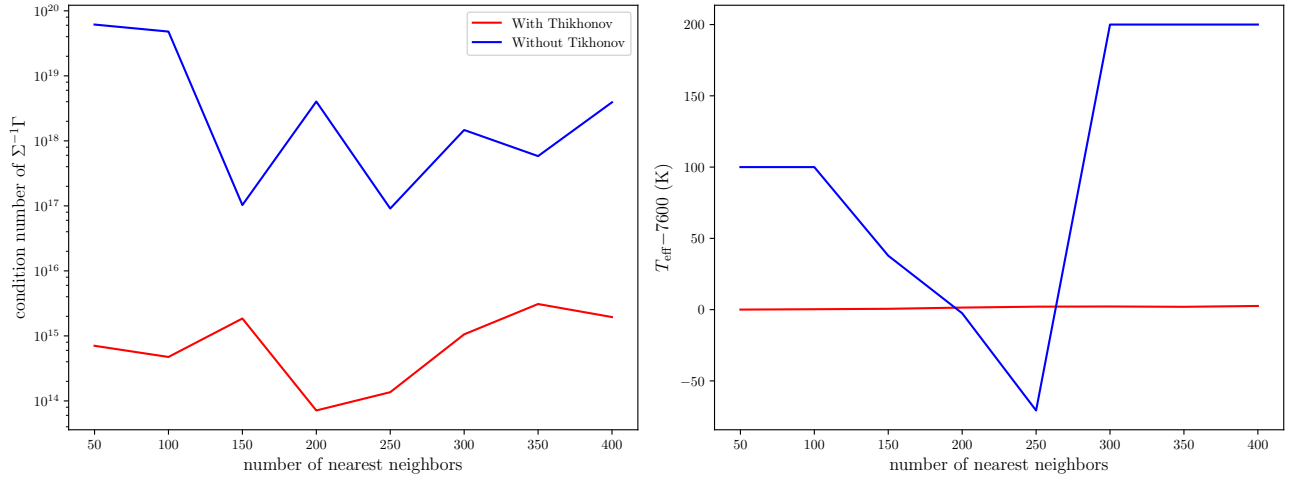
### 4.1 The learning databases

As done in Paletou *et al.* (2015 b) and Gebran *et al.* (2016), model atmospheres were calculated using ATLAS9 with the new opacity distribution function (Kurucz 1992; Castelli & Kurucz 2003). These models assume local thermodynamic equilibrium (LTE), hydrostatic equilibrium, and a 1D plane-parallel atmosphere. Convection was treated using a mixing length parameter of 0.5 for  $7000 \text{ K} \leq T_{\text{eff}} \leq 8500 \text{ K}$ , and 1.25 for  $T_{\text{eff}} \leq 7000 \text{ K}$ , following the prescriptions of Smalley (2004). Synthetic spectra were calculated using SYNSPEC48 (Hubeny & Lanz 1992). The adopted line lists were from Kurucz *ghyperall.dat*<sup>1</sup> and modified with more recent and accurate atomic data retrieved from the VALD<sup>2</sup> and the NIST<sup>3</sup> databases (for more details see Gebran *et al.* 2016). The calculation time for one spectrum depends mainly of the selected wavelength range. For instance, cal-

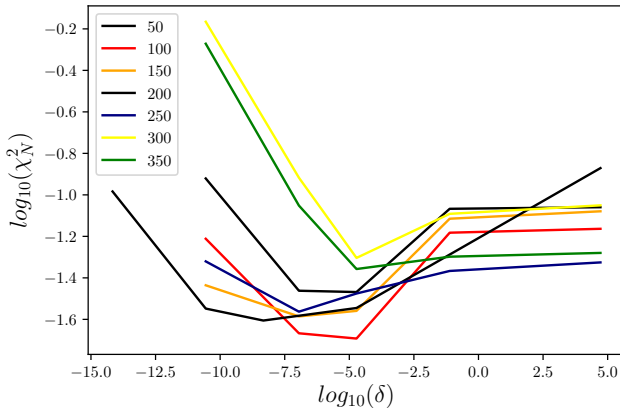
<sup>1</sup> <http://kurucz.harvard.edu>

<sup>2</sup> <http://www.astro.uu.se/~vald/php/vald.php>

<sup>3</sup> <http://physics.nist.gov>



**Figure 2.** Left: variation of the condition number as a function of the number of nearest neighbor using the optimum reduced LDB with and without the Tikhonov regularization. Right: the inverted  $T_{\text{eff}}$  as a function of the number of nearest neighbor with and without Tikhonov regularization. The inversion is done for a noise added synthetic spectrum with  $T_{\text{eff}}=7600$  K,  $\log g=2.50$  dex,  $[M/H]=0$  dex,  $v \sin i=197$  km s $^{-1}$  and  $S/N=196$ . For clarity we display on the value of  $T_{\text{eff}}-7600$  K.



**Figure 3.**  $\log_{10}(\chi^2_N)$  versus  $\log_{10}(\delta)$  for  $T_{\text{eff}}$  of a synthetic spectra having  $T_{\text{eff}}=7600$  K,  $\log g=2.50$  dex,  $v \sin i=197$  km s $^{-1}$ , and with a  $S/N$  of 196. Each curve displays the minimization using different sets of nearest neighbor generated iteratively.

culating one spectrum in the range of 4450–4990 Å at a resolution of 76 000 requires ~20 seconds on a personal computer<sup>4</sup>. Around 500 days were necessary for the calculation of the A stars LDB used in Sec. 4.2. For the wavelength range between 5000–5400 Å, one spectrum requires ~22 seconds.

## 4.2 Inversion of simulated A stars

We used the LDB of Gebran *et al.* (2016) in which the effective temperature of the data varies from 6 800 up to 11 000

**Table 1.** Ranges of the parameters used for the calculation of the A and FGK synthetic spectra LDB

Parms	A stars	F/G/K
$T_{\text{eff}}$ (K)	[6 800, 11 000]	[4 000, 8 000]
$\log g$ (dex)	[2.0, 5.0]	[3.0, 5.0]
$[M/H]$ (dex)	[-2.0, 2.0]	[-1.0, 1.0]
$v \sin i$ (km s $^{-1}$ )	[0, 300]	[0, 100]
$\lambda/\Delta\lambda$	76 000	50 000

K. The wavelength region was chosen between 4 450–4 990 Å. This wavelength region harbors lines that are sensitive to all stellar parameters, and insensitive to microturbulent velocity which was adopted to be  $\xi_t = 2$  km/s based on the work of Gebran *et al.* (2016, 2014). The adopted resolution is 76 000 as it corresponds to most of the analyzed stars in Sec. 5. The ranges of all parameters in the A-star LDB are summarized in Table 1.

Noise added synthetic spectra were calculated to be used as simulated observations. Around 1 500 spectra were calculated for A stars with parameters randomly selected within the range of the LDB but not necessarily at the grid points. To analyze the effect of the sampling on the RSIR technique, we have inverted these spectra using 3 different LDB. For the same range in all the parameter only the step was modified in each database. As an example, in the LDB 1,  $T_{\text{eff}}$  has a step of 100 K, whereas in LDB's 2 and 3, the steps are 200 K and 400 K, respectively. The same was done for all parameters and the details about the steps are found in Table 2. The sampling of the  $v \sin i$  in the LDB is not constant and depends on the value of  $v \sin i$  (Gebran *et al.* 2016).

<sup>4</sup> Intel core i7-4510U CPU at 2.00GHz × 4 with 16Gb RAM.

**Table 2.** Result of inversion for A stars using 3 different LDB with different steps on 1500 noisy synthetic spectra.

Test	Parms	step	$\Lambda_{RSIR}$	$\Lambda_{PCA}$	offset <sub>RSIR</sub>	offset <sub>PCA</sub>
1	$T_{\text{eff}}$ (K)	100	100.3	132	63.45	61.128
	$\log g$ (dex)	0.1	0.12	0.133	0.042	0.047
	$[M/H]$ (dex)	0.1	0.058	0.066	-0.0102	-0.0057
	$v \sin i$ (Km/s)	2: [0-20] 5: [20-40] 10: [40:300]	5.92	6.47	-0.279	0.144
2	$T_{\text{eff}}$ (K)	200	108	190	67.54	69.33
	$\log g$ (dex)	0.2	0.109	0.145	0.03	0.04
	$[M/H]$ (dex)	0.2	0.072	0.107	-0.0102	-0.01
	$v \sin i$ (Km/s)	10	9.63	10.25	-3.5	-1.82
3	$T_{\text{eff}}$ (K)	400	174	295	40.74	92.39
	$\log g$ (dex)	0.4	0.17	0.224	0.014	0.061
	$[M/H]$ (dex)	0.3	0.081	0.113	-0.016	0.0015
	$v \sin i$ (Km/s)	20	12.05	11.23	-4.11	-0.9

To compare the results of the inversion of PCA NN-search and RSIR for 1500 spectra, we estimate the root mean square error for both techniques,  $\Lambda$ , defined as:

$$\Lambda = \sqrt{\frac{\sum_{i=0}^N (y_i^{(\text{inv})} - y_i^{(\text{true})})^2}{N}}, \quad (13)$$

where  $y_i^{(\text{true})}$  is the known parameter of the  $i^{\text{th}}$  synthetic spectrum and  $y_i^{(\text{inv})}$  its corresponding inverted one.

Columns 4 and 5 of Table 2 display the  $\Lambda$  results using the PCA NN-search and the RSIR for the 3 LDB. The offsets, calculated as a signed mean difference, between the inverted and the true values are presented in the last two columns of Table 2. Comparing the  $\Lambda$  values of each approach, an improvement is achieved using RSIR for all parameters. One exception exists in the case of  $v \sin i$  for test 3. The large  $v \sin i$  step of the original LDB causes the PCA NN-search pre-processing stage to select inaccurate NN's. For most cases RSIR with a coarse sampling in parameters is producing more accurate inversions compared to PCA with a denser sampling. This directly infers a gain in computational time as a coarse sampling leads to smaller LDB. The time required to invert the parameters of one synthetic spectrum depends on the computational facilities. For instance, the gain in time for using the A-stars LDB of test 2 instead of the one of test 1 is  $\sim 25\%$ .

To analyze the effect of the S/N on the inversions, we display in Figure 4 the inverted  $T_{\text{eff}}$  as a function of the real  $T_{\text{eff}}$  for the 1500 A star spectra, for different S/N and different LDB (tests 1, 2 and 3). The results of the PCA NN-search is affected both by the sampling size and the S/N of the analyzed stars, whereas for RSIR, with the pre-processing of PCA and a Tikhonov regularization this effects becomes less significant on the accuracy of the inversion. In the appendix, we present the behavior of the inversion of  $\log g$ ,

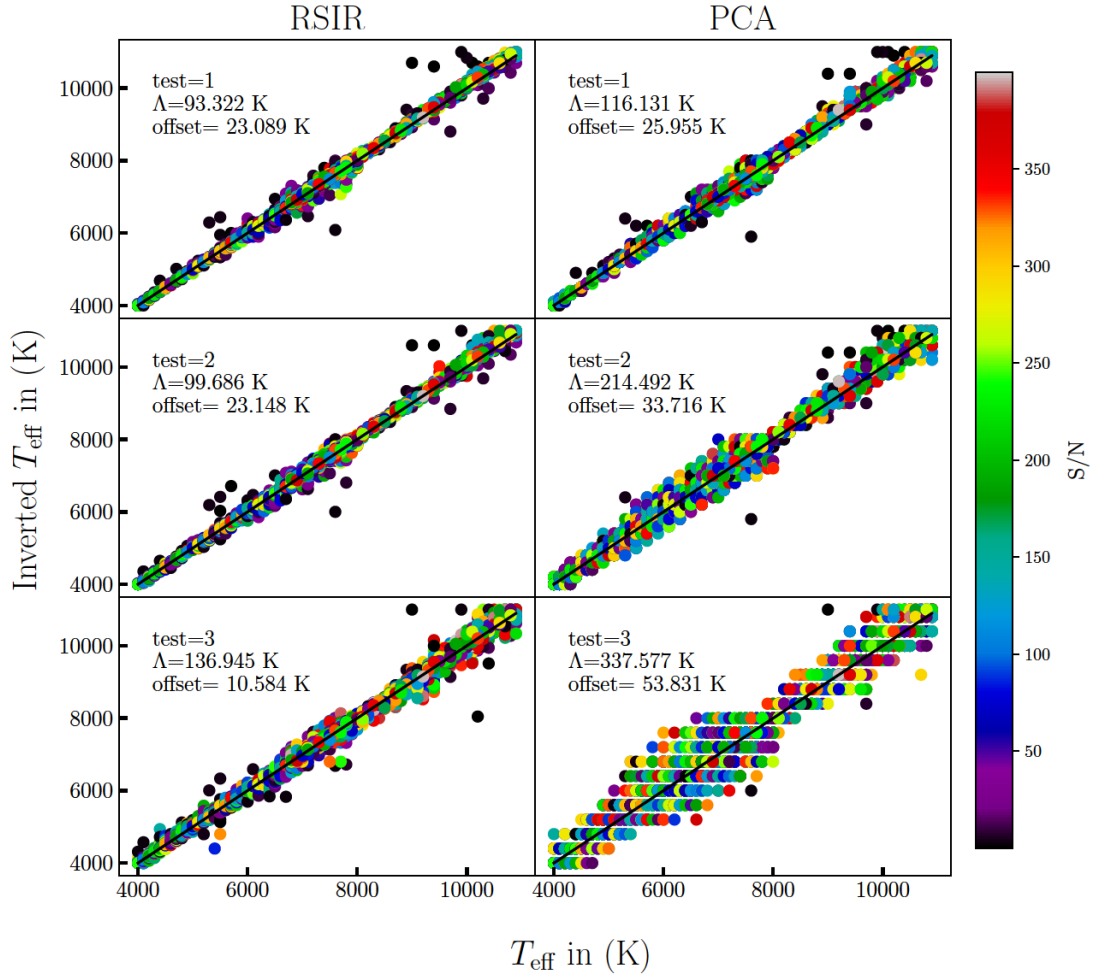
$[M/H]$ , and  $v \sin i$ . A similar behavior to the one of  $T_{\text{eff}}$  can be concluded for these parameters and this can be shown in Figures A1, A2 and A3.

### 4.3 Inversion of Simulated FGK type stars

The same procedure was applied to FGK star-like spectra. Around 2500 noisy spectra were produced in the ranges described Table 1. As done for the A stars, the parameters of the FGK synthetic spectra were also selected randomly. The chosen resolution of 50 000 is the same used in Paletou et al. (2015 a). The wavelength range was selected from 5 000 to 5 400 Å containing the Mg I b triplet, a good indicator of  $\log g$  and sensitive as well to  $T_{\text{eff}}$ . The microturbulent velocity was set to  $\xi_t \sim 1 \text{ km s}^{-1}$  (Gebran et al. 2014). The inversion results ( $\Lambda$  and offsets) as a function of the sampling steps are shown in Table 3. These results show similar behavior to that of the A stars in terms of improvement in accuracy while comparing RSIR to PCA NN-search. In Figure 4 we also overplot the  $T_{\text{eff}}$  for our F/G/K noisy synthetic spectra. The effect of inversion as a function of S/N and sampling is very similar to the one of A stars, and for all the parameters (Figures A1, A2 and A3).

## 5 Application to observed spectra

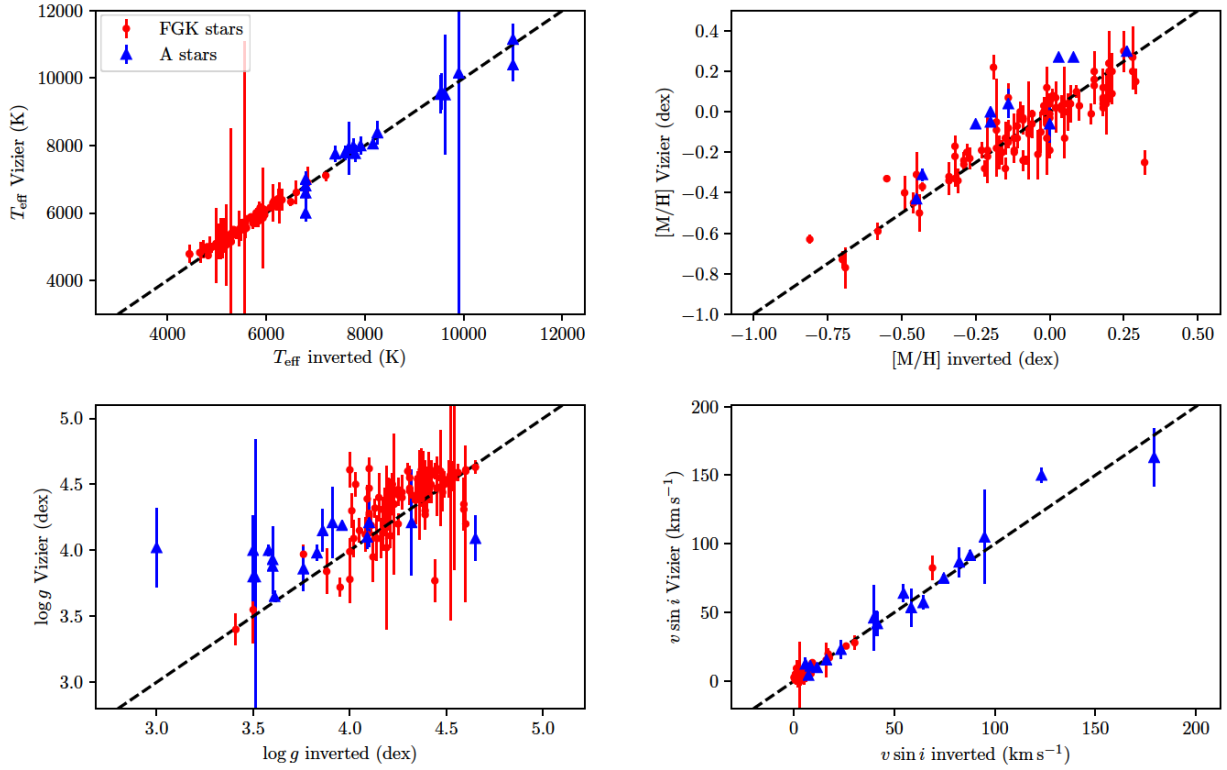
The performance of the RSIR has been tested on two samples of stellar spectra. The first sample is the one of the Spectroscopic Survey of Stars in the Solar Neighborhood ( $S^4N$ , Allende Prieto et al. 2004). These are spectra of bright FGK stars that are at distance less than 15 pc. We have estimated the S/N of these spectra in the wavelength range



**Figure 4.** Results of the effective temperature inversion for the synthetic A to K stars. The black line represents the 1-to-1 correspondence.

**Table 3.** Result of inversion using 3 different LDB with different steps on 2500 noisy synthetic FGK spectra.

Test	parms	step	$\Lambda_{RSIR}$	$\Lambda_{PCA}$	offset <sub>RSIR</sub>	offset <sub>PCA</sub>
1	$T_{\text{eff}}$ (K)	100	106	117	5.92	0.04
	$\log g$ (dex)	0.1	0.121	0.136	0.016	0.002
	$[M/H]$ (dex)	0.1	0.061	0.063	-0.45	-0.56
	$v \sin i$ (Km/s)	2: [0-20] 5: [20-40] 10: [40-100]	1.72	2.88	0.005	0.06
2	$T_{\text{eff}}$ (K)	200	109	236	8.17	-7.86
	$\log g$ (dex)	0.2	0.132	0.289	0.0027	-0.017
	$[M/H]$ (dex)	0.2	0.066	0.115	-0.47	-0.62
	$v \sin i$ (Km/s)	2: [0-20] 10: [20-40] 10: [40-100]	1.88	3.17	0.0087	0.0004
3	$T_{\text{eff}}$ (K)	400	127	368	10.43	-27
	$\log g$ (dex)	0.4	0.147	0.45	-0.01	-0.04
	$[M/H]$ (dex)	0.4	0.07	0.17	-0.84	-0.625
	$v \sin i$ (Km/s)	4: [0-20] 10: [20-40] 10: [40-100]	2.25	5.243	0.018	-0.011



**Figure 5.** Comparison between the inverted parameters for our FGK stars sample (filled circles) and A stars (filled triangle).

used for the inversion of the parameters [5 000–5 400Å]. This ratio ranges between 40 and 450. These spectra are at a resolution of  $\lambda/\Delta\lambda \sim 50\,000$ . All the details about the acquisition and the reduction procedure of the  $S^4N$  data can be found in Allende Prieto *et al.* (2004). These spectra were inverted using the database of Paletou *et al.* (2015 a), made of 905 spectra retrieved from the ELODIE stellar library (Prugniel & Soubiran 2001; Prugniel *et al.* 2007). We have used the Mg I b triplet wavelength range as explained in Sec. 4.3. We then compared the values of the inverted parameters with the ones of Allende Prieto *et al.* (2004) and to the medians found in the Vizier catalog<sup>5</sup> for all these stars. The main reason for using this catalog for our comparison is the necessity for reliable and objective catalogs which are constructed based on previous adopted values by the astronomical community.

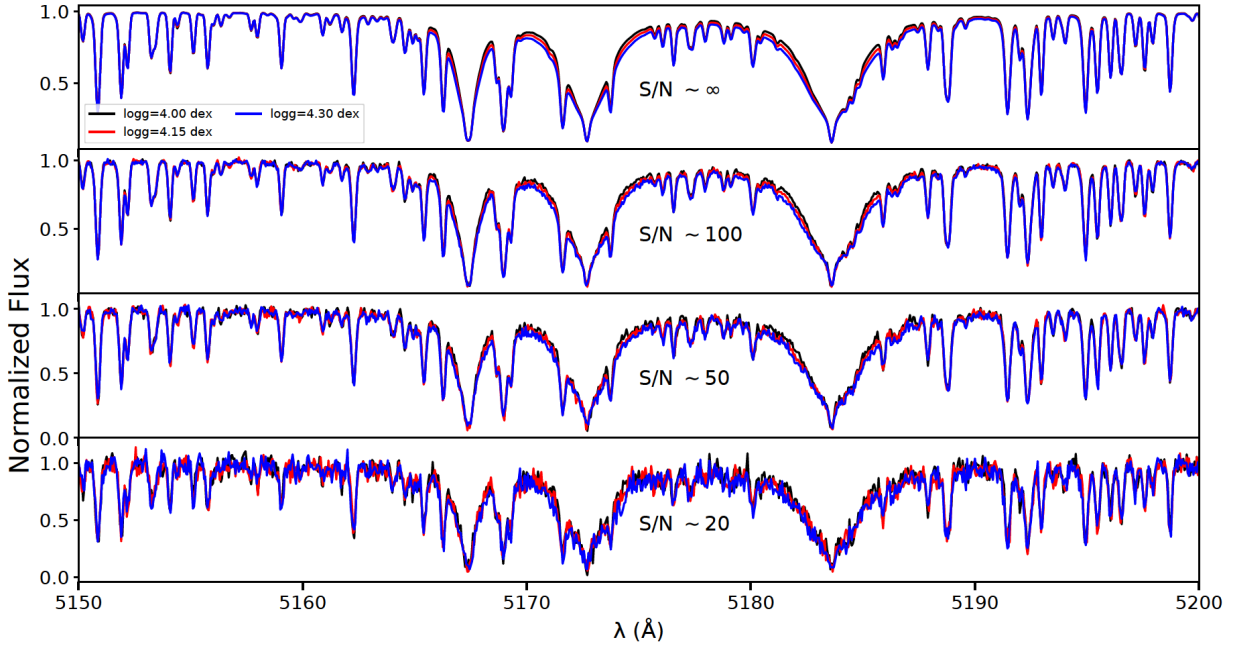
Comparing our inverted  $T_{\text{eff}}$  to the ones of Allende Prieto *et al.* (2004), we found an average signed difference of 2.09 K with standard deviation of 102 K. For  $\log g$ , the average signed difference and the standard deviation are both 0.15 dex. For  $[M/H]$  and  $v \sin i$ , we found  $-0.06 \pm 0.08$  dex and  $-0.21 \pm 1.89 \text{ km s}^{-1}$ , respectively. If we compare our

inverted values to the median of Vizier, we find  $-85 \pm 110 \text{ K}$ ,  $-0.07 \pm 0.16$  dex,  $0.01 \pm 0.10$  dex and  $-0.50 \pm 2.25 \text{ km s}^{-1}$  as a signed mean difference and a standard deviation between the catalogued values and the inverted ones for  $T_{\text{eff}}$ ,  $\log g$ ,  $[M/H]$  and  $v \sin i$ , respectively. Figure 5 displays in filled circles, for the four parameters, the comparison between our inverted values for the FGK observed spectra and the ones derived from Vizier. We have also assigned the catalogues values an error bar corresponding to the standard deviation of the dispersion in the catalogues values for each star.

The second sample of our analysis is constituted of the well studied A stars of Gebran *et al.* (2016). These are the 19 stars that have been studied extensively by different authors using different techniques (Vega, Sirius A, HD 22484, HD 15318, HD 76644, HD 49933, HD 214994, HD 214923, HD 113139, HD 114330, HD 27819, HD 5448, HD 33256, HD 29388, HD 91480, HD 30210, HD 32301, HD 28355, and HD 222603) and have more than 120 references each. The source of these high resolution spectra is explained in detail in Gebran *et al.* (2016). They were observed using ELODIE, NARVAL, ESPaDOnS and SOPHIE spectrographs. ELODIE has a resolution of 42 000 whereas NARVAL, ESPaDOnS and SOPHIE are at a resolution of  $\sim 76\,000$ .

We have applied the RSIR on these data using the database of Test 1 in Sec. 4.2 at both resolutions. The  $S/N$  of these

<sup>5</sup> The query was performed using the method described in Paletou & Zolotukhin (2014)



**Figure 6.** Synthetic spectra of G stars having  $[T_{\text{eff}}, \log g, [M/H], v \sin i]$  of 5200 K, [4.00, 4.15, 4.30 dex], 0.0 dex, 6 km s<sup>-1</sup>. Each plot displays the spectra with different S/N.

spectra is between 180 and 360. The inverted parameters of each star were compared to the ones retrieved from VizieR and added to the plots of Figure 5 as filled triangles. We found an average signed difference and a standard deviation of  $-0.14 \pm 245$  K,  $-0.20 \pm 0.30$  dex,  $-0.11 \pm 0.09$  dex and  $-2.07 \pm 8.5$  km s<sup>-1</sup> for  $T_{\text{eff}}$ ,  $\log g$ ,  $[M/H]$  and  $v \sin i$ , respectively between the inverted and the VizieR parameters.

### 5.1 The case of $\log g$

These results show that most of our inverted parameters are in agreement with previous studies. Considering that the most accurate parameters of these A and FGK stars are the VizieR median, our values are less spread with respect to the median than the ones of Paletou *et al.* (2015 a) and Gebran *et al.* (2016). The standard deviations that we found could be assigned as an estimation of the errors on the derived parameters. We can therefore assign precision of 110 K, 0.16 dex, 0.10 dex and 2.25 km s<sup>-1</sup>, on  $T_{\text{eff}}$ ,  $\log g$ ,  $[M/H]$ , and  $v \sin i$ , respectively for FGK stars. For A stars, we found precision of 245 K, 0.30 dex, 0.09 dex, and 8.50 km s<sup>-1</sup>, on  $T_{\text{eff}}$ ,  $\log g$ ,  $[M/H]$ , and  $v \sin i$ , respectively. These errors are summarized in Table 4.

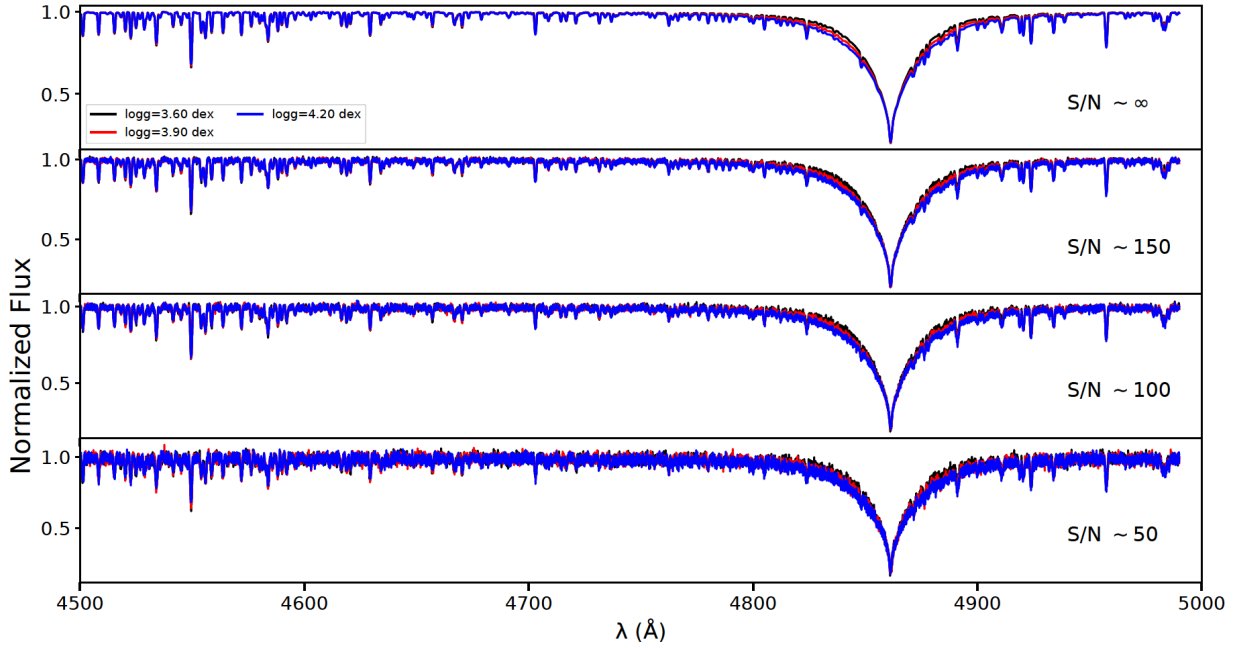
Surface gravity is systematically the most difficult parameter to determine, with typical errors of the order of 0.15 to 0.3 dex. This parameter is very important for chemical analysis as some line profiles could be very sensitive to

**Table 4.** Estimation of the offset (signed mean difference) and the errors on the derived parameters for FGK and A stars.

Parameter	Offset (FGK)	$\sigma_{\text{FGK}}$	Offset (A)	$\sigma_{\text{A}}$
$T_{\text{eff}}$ (K)	-85	110	-0.14	245
$\log g$ (dex)	-0.07	0.16	-0.20	0.30
$[M/H]$ (dex)	0.01	0.10	-0.11	0.09
$v \sin i$ (km s <sup>-1</sup> )	-0.5	2.25	-2.07	8.50

$\log g$  values. Spectroscopic determinations of surface gravity have always been assigned moderately large error bars, especially for A stars (Smalley 2005). The same applies to FGK stars but with smaller error bars. Asteroseismic  $\log g$  determinations remain the best tools for achieving accuracies less than 0.05 dex (Hekker *et al.* 2013; Creevey *et al.* 2013; Chaplin *et al.* 2014). RSIR is mainly based on finding the best set of spectra in the database that correspond to the observed one. As it is a spectroscopic method, we should not expect an accurate recovery for  $\log g$ . Using our values for  $\Delta \log g$  we can, a posteriori figure out what it means in terms of discernibility between two spectra whose respective  $\log g$  differ from this quantity. This also gives us relevant information about (i) which specific bandwidth(s) are the most sensitive to such differences, and (ii) how significant they are for various S/N.

Figures 6 and 7 display the variation in the spectrum profile as a function of  $\log g$ , fixing all the remaining parameters, for G and A stars, respectively. In Figure 6, we



**Figure 7.** Same as for Figure 6 but for A stars having having  $[T_{\text{eff}}, \log g, [M/H], v \sin i]$  of  $[8500 \text{ K}, [3.60, 3.90, 4.20 \text{ dex}], 0.0 \text{ dex}, 40 \text{ km s}^{-1}]$ .

calculated synthetic spectra for a typical G star with a  $T_{\text{eff}}$  of  $5200 \text{ K}$ ,  $[M/H]$  of  $0.0 \text{ dex}$ ,  $v \sin i$  of  $6 \text{ km s}^{-1}$  at a resolution of  $50\,000$  and in the wavelength range of  $5\,000\text{--}5\,400 \text{ \AA}$ . The only parameter that differs between the 3 spectra is  $\log g$ , ranging between  $4.00 \text{ dex}$  and  $4.30 \text{ dex}$  with a step of  $0.15 \text{ dex}$ . The upper panel of Figure 6 displays the normalized flux of the synthetic spectra in the Mg I b triplet region. The flux level in this region is very sensitive to variation in  $\log g$ . The following panels displays the same spectra for different values of  $S/N$ . When no noise is added (panel with  $S/N \sim \infty$ ), the distinction between the 3 spectra is clear but when  $S/N$  starts to decrease, the distinction between the noisy spectra becomes harder to detect. This shows that for a  $S/N$  in the order of  $100$ , the noisy spectra with  $\log g$  of  $4.00$  and  $4.15 \text{ dex}$  are very similar and therefore the best corresponding synthetic spectrum in our LDB could have a  $\log g$  varying at least  $0.15 \text{ dex}$  from the correct value. We are not trying to quantify the minimum  $S/N$  required for an accurate inversion of  $\log g$  as the RSIR is not based on a pixel-to-pixel comparison, but we are showing the effect of our derived standard deviations in  $\log g$  on the flux for noisy spectra. Figure 7 displays a similar behaviour for A stars having similar  $T_{\text{eff}}$  of  $8\,500 \text{ K}$ ,  $[M/H]$  of  $0.0 \text{ dex}$ ,  $v \sin i$  of  $40 \text{ km s}^{-1}$ , at a resolution of  $76\,000$  in the wavelength range of  $4\,500\text{--}5\,000 \text{ \AA}$ . Surface gravity of these spectra ranges between  $3.60$  and  $4.20 \text{ dex}$  with a step of  $0.30 \text{ dex}$ . This figure shows a similar behavior to that of Figure 6. At a  $S/N$  of  $\sim 150$ , the distinction between spectra having a difference of  $0.30 \text{ dex}$  in  $\log g$ , becomes hardly noticeable.

Figures 6 and 7 also show that the effect of weak metallic lines, on the derivation of  $\log g$ , becomes negligible as the  $S/N$  decreases. The  $\log g$  information that these lines contain is mainly lost in the noise.

## 6 Discussion and conclusion

RSIR tests for nearly  $4\,000$  synthetic stars of different spectral type and different noise levels showed an improvement over the PCA-based method of Paletou *et al.* (2015 a,b) and Gebran *et al.* (2016) for the inversion of stellar parameters. Results of Tables 2 and 3 and Figure 4, for FGK and A stars, show that for most of the tests, the  $\Lambda$  values of RSIR are lower than nearest neighbor PCA approach. Having a prior information about the star using PCA as a pre-process allows us to narrow down the selection of the optimized reduced databases. This decreases drastically the size of the LDB. Achieving lower  $\Lambda$  with bigger steps helps in decreasing the prohibitive computation time for the construction of databases and the calculations of the PC's. Simulated tests revealed that computation time of RSIR is nearly 1% of that of the process of PCA nearest neighbor approach.

One should be very careful while increasing the size of steps of the parameters because the PCA pre-processing step could deviate drastically from the true inverted values therefore excluding the spectra that actually best describes the observed ones.

Application to observed FGK and A stars reveal a good agreement between the inverted parameters and the ones derived in previous studies. The comparison with Vizier catalog values show an improvement in the derived parameters as compared to the results of Paletou *et al.* (2015 a) and Gebran *et al.* (2016) for the same stars and LDB. Surface gravity remains the parameter with the least accuracy. Our derived errors on  $\log g$  are in the order of 0.15-0.30 dex. Smarter LDB should be therefore considered, say, “adaptive sampling” (in the parameters under study), taking care with caution of the flux typical variations at the most sensitive wavelength (sub-)domains, *together with* the S/N of the observations, instead of the a priori sampling in the parameters. Also, a commonly reported issue with the inversion of stellar parameters using a LDB of synthetic spectra are the so-called “ambiguities”. This means that two sets of distinct parameters may generate spectra which are beyond “discernibility”. Given a set of observed spectra to characterize, we could naturally relate that discernibility to their level of S/N. Using a nearest neighbor search PCA-based method, for instance, such a level of S/N can easily be translated into a threshold of distance  $\delta_{\text{PCA}}$ . Then, we can anticipate that, instead of relying on LDB usually made using a priori sampling *in the parameters*, a smarter DB should rely on  $\delta_{\text{PCA}}$  instead. This would imply to set up LDB for fundamental stellar parameters in a radically different fashion vs. common practices. In the frame of PCA, it would be more relevant to sample properly the full range of parameters with a “constrained-random” process ensuring that there are no nearest neighbors closer than  $\delta_{\text{PCA}}$ . Such a “sieve algorithm” was first proposed by López Ariste & Casini (2002) in the context of the characterization of magnetic fields from spectropolarimetric data (see also Casini *et al.* 2013). Another line of development relates to the “structure” of our LDB. Smarter, or optimal LDB, using different methods of samplings, should be considered. Such a general issue was already evoked by Bijaoui *et al.* (2012) for instance.

Available online databases are usually calculated with large steps in  $T_{\text{eff}}$  and  $\log g$ . Our RSIR technique, as it does not require small steps in the LDB, is a good tool to be used with online available synthetic spectra such as the POLLUX<sup>6</sup> database (Palacios *et al.* 2010) that contains models with temperature ranging between 3 000 and 50 000 K or TLUSTY Non-LTE Line-blanketed Model Atmospheres of O-Type Stars (Lanz & Hubeny 2003) with  $T_{\text{eff}}$  ranging between 27 500 and 55 000 K with 2 500 K steps, and  $\log g$  between 3.0 and 4.75 with steps of 0.25 dex. We can also mention the PHOENIX (Husser *et al.* 2013) models database

for stars having  $T_{\text{eff}} < 12\,000$  K and the AMBRE (de Laverny *et al.* 2012) project that contains high-resolution FGKM stellar synthetic spectra.

As an output of the new Gaia Data Release 2, Cropper *et al.* (2018) describe the Gaia RVS specification as well as the predicted performance at the end of the mission. Gaia RVS will provide us with a large number of spectra in the calcium triplet regime (845 – 872 nm). This triplet is very sensitive to  $T_{\text{eff}}$  and  $\log g$ . The medium resolution (11 500) of the RVS and the small range in wavelength would require LDB smaller than the ones used in our work, leading to a fast application of the RSIR. As we did for the inversion of the  $S^4N$  data in Sec. 5, LDB could be constructed with real observed stars having well known fundamental parameters and with the same resolution. Finally, since RSIR is based on single parameter inversion process, one can also incorporate other parameters at the cost of computing and handling more numerous individual spectra, for example, microturbulence velocity and individual chemical abundances.

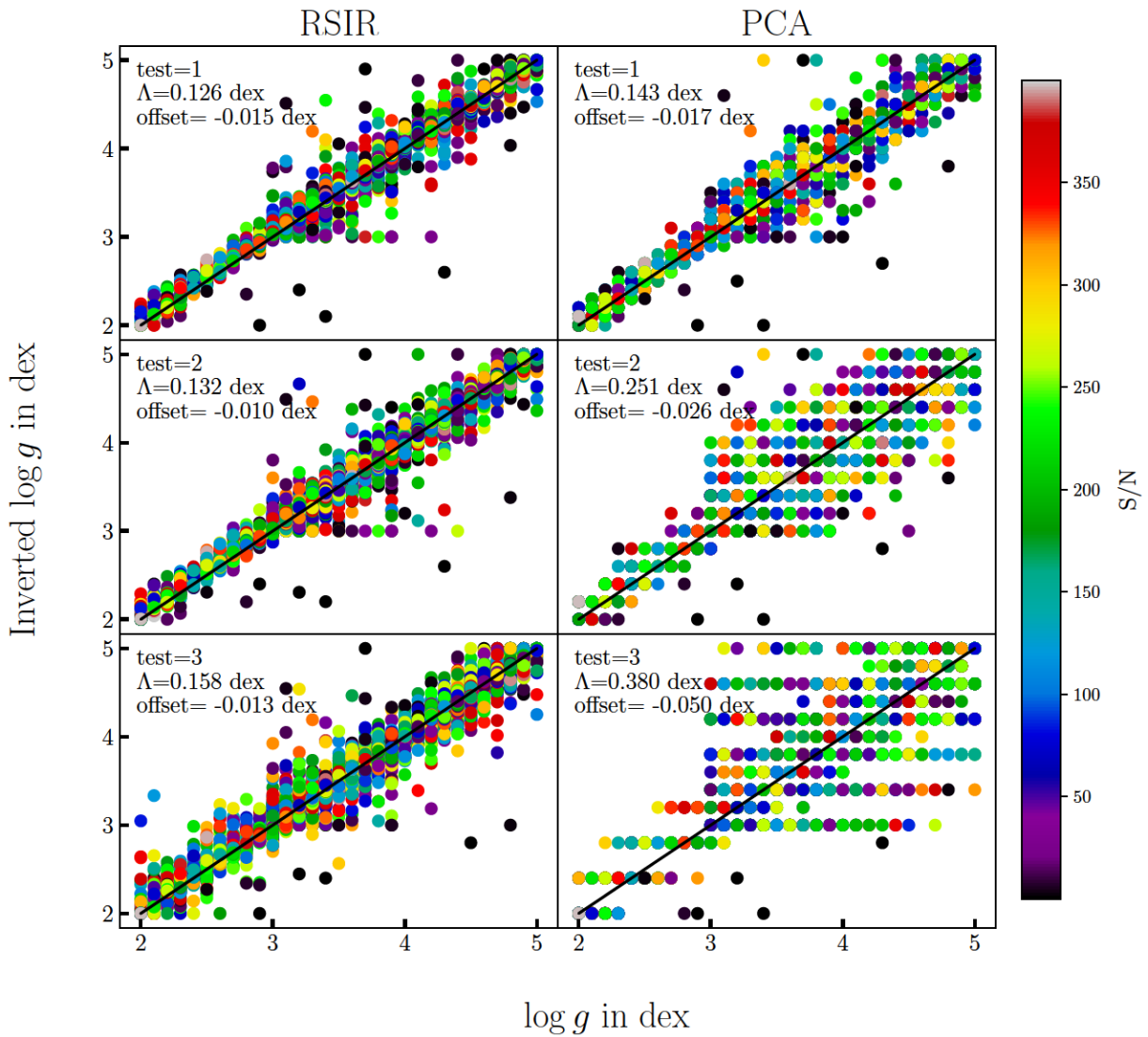
## References

- Abolfathi, B., Aguado, D. S., Aguilar, G., Allende Prieto, C., Almeida, A., Ananna, T. T. *et al.* 2018, *ApJS*, 235, 42.
- Allende Prieto, C., Barklem, P. S., Lambert, D. L., & Cunha, K. 2004, *A & A*, 420, 183–205.
- Alves, S., Benamati, L., Santos, N. C., Adibekyan, V. Zh., Sousa, S. G., Israelian, G. *et al.* 2015, *MNRAS*, 448, 2749–2765.
- Bailer-Jones, C. A. L., Irwin, M., & von Hippel, T. 1998, *MNRAS*, 298, 361–377.
- Bernard-Michel, C., Douté, S., Fauvel, M., Gardes, L., & Girard, S. 2007, Research Report, INRIA(inria-00187444v2), 91.
- Bernard-Michel, C., Douté, S., Fauvel, M., Gardes, L., & Girard, S. 2009, *Journal of Geophysical Research (Planets)*, 114, E06005.
- Bijaoui, A., Recio-Blanco, A., de Laverny, P., & Ordenovic, C. 2012, *Statistical Methodology*, 9, 55–62.
- Boeche, C., Smith, M. C., Grebel, E. K., Zhong, J., Hou, J. L., Chen, L. *et al.* 2018, *AJ*, 155, 181
- Buchhave, L. A., Latham, D. W., Johansen, A., Bizzarro, M., Torres, G., Rowe, J. F. *et al.* 2012, *Nature*, 486, 375–377.
- Casini, R., Asensio Ramos, A., Lites, B. W., & López Ariste, A. 2013, *APJ*, 773, 180.
- Chaplin, W. J., Basu, S., Huber, D., Serenelli, A., Casagrande, L., Silva Aguirre, V. *et al.* 2014, *ApJS*, 210, 1.
- Casey, A. R., Hogg, D. W., Ness, M., Rix, H.-W., Q Y Ho, A., & Gilmore, G. 2016, arXiv:1603.03040
- Castelli, F. & Kurucz, R. L. 2003, *Modelling of Stellar Atmospheres*, 210, A20
- Cayrel, G. & Cayrel, R. 1963, *APJ*, 137, 431–469.
- Cayrel de Strobel, G. 1969, In: Gingerich O. (Ed.), *Proc. of the 3rd Harvard-Smithsonian Conf. on Stellar Atmospheres*, 35–39.
- Creevey, O. L., Thévenin, F., Basu, S., Chaplin, W. J.; Bigot, L.; Elsworth, Y. *et al.* 2013, *MNRAS*, 431, 2419–2432.

- Cropper, M., Katz, D., Sartoretti, P., Prusti, T., de Bruijne, J.H.J., Chassat, F. et al. 2018, arXiv:astro-ph/1804.09369
- Cui, X.-Q., Zhao, Y.-H., Chu, Y.-Q., et al. 2012, *Research in Astronomy and Astrophysics*, 12, 1197.  
Data: <http://dr5.lamost.org/>
- Dieterich, S., Henry, T. J., Benedict, G. F., Jao, W.-Ch., White, R. & RECONS Team 2017, *American Astronomical Society Meeting Abstracts #229*, 229, 240.30.
- Fabbro, S., Venn, K. A., O'Briain, T., Bialek, S., Kielty, C. L., Jahandar, F. et al. 2018, *MNRAS*, 475, 2978–2993.
- Gebran, M., Monier, R., Royer, F., Lobel, A., & Blomme, R. 2014, *Putting A Stars into Context: Evolution, Environment, and Related Stars*, 193–198.
- Gebran, M., Farah, W., Paletou, F., Monier, R., & Watson, V. 2016, *A & A*, 589, A83.
- Gill, S., Maxted, P. F. L., & Smalley, B. 2018, arXiv:1801.06106.
- Hekker, S., Elsworth, Y., Mosser, B., Kallinger, T., Basu, S., Chaplin, W. J. et al. 2013, *A & A*, 556, A59.
- Hubeny, I. & Lanz, T. 1992, *A & A*, 262, 501–514.
- Husser, T.-O., Wende-von Berg, S., Dreizler, S., Homeier, D., Reiners, A., Barman, T. et al. 2013, *A & A*, 553, A6.
- Jolliffe, I. T. 1986, *Springer Series in Statistics*, Berlin: Springer, 1986.
- Katz, D. & Brown, A. G. A. 2017, *SF2A-2017: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics*, 259–263.
- Kiefer, J. 1953, *Proceedings of the American Mathematical Society*, 4, 502–506.
- Kreyszig, E. 2010, *Advanced Engineering Mathematics*, John Wiley & Sons, 864-871.
- Kurucz, R. L. 1992, *RMXAA*, 23.
- Lanz, T. & Hubeny, I. 2003, *ApJS*, 146, 417–441.
- Latham, D. W., Stefanik, R. P., Torres, G., Davis, R. J., Mazeh, T., Carney, B. W. et al. 2002, *AJ*, 124, 1144–1161.
- de Laverny, P., Recio-Blanco, A., Worley, C. C., & Plez, B. 2012, *A & A*, 544, A126.
- Li, K. C. 1991, *Journal of the American Statistical Association*, 86, 414, 316–342.
- López Ariste, A., & Casini, R. 2002, *ApJ*, 575, 529–541.
- McWilliam, A. 1990, *ApJS*, 74, 1075–1128.
- Morris, M., Kaiser, M. E., Bohlin, R., Kurucz, R., & ACCESS Team 2018, *American Astronomical Society Meeting Abstracts #231*, 231, #355.27
- Ness, M., Hogg, D. W., Rix, H.-W., Ho, A. Y. Q., & Zasowski, G. 2015, *APJ*, 808, 16.
- O'Mullane, W., & LSST Data Management Team 2018, *American Astronomical Society Meeting Abstracts #231*, 231, #362.10
- Palacios, A., Gebran, M., Josselin, E., Martins, F., Plez, B., Belmas, M. et al. 2010, *A & A*, 516, A13.
- Paletou, F., & Zolotukhin, I. 2014, arXiv:1408.7026.
- Paletou, F., Böhm, T., Watson, V., & Trouilhet, J.-F. 2015, *A & A*, 573, A67.
- Paletou, F., Gebran, M., Houdebine, E. R., & Watson, V. 2015, *A & A*, 580, A78.
- Perryman, M. A. C., de Boer, K. S., Gilmore, G., Høg, E., Lattanzi, M. G., Lindegren, L. et al. 2001, *A & A*, 369, 339–363.  
Data: <https://www.cosmos.esa.int/web/gaia/dr2>
- Prugniel, P. & Soubiran, C. 2001, *A & A*, 369, 1048–1057.
- Prugniel, P., Soubiran, C., Koleva, M., & Le Borgne, D. 2007, arXiv:astro-ph/0703658.
- Re Fiorentin, P., Bailer-Jones, C. A. L., Lee, Y. S., Beers, T. C., Sivarani, T., Wilhelm, R. et al. 2007, *A & A*, 467, 1373–1387.
- Schönrich, R. & Bergemann, M. 2014, *MNRAS*, 443, 698–717.
- Shevlyakova, M. & Morgenthaler, S. 2014, *Statistical Papers*, 55, 1.
- Smalley, B., Zverko, J., Žižňovský, J., Adelman, S. J. & Weiss, W. W. 2004, *The A-Star Puzzle*, Proc. IAU Symp. 224, Cambridge Univ. Press Cambridge.
- Smalley, B. 2005, *Memorie della Societa Astronomica Italiana Supplementi*, 8, 130.
- Stoehr, F., White, R., Smith, M., Kamp, I., Thompson, R., Durand, D. et al. 2008, *Astronomical Data Analysis Software and Systems XVII*, 394, 505–508.
- Ting, Y.-S., Conroy, C., Rix, H.-W., & Cargile, P. 2018, arXiv:1804.01530.
- Torres, G., Neuhäuser, R., & Guenther, E. W. 2002, *AJ*, 123, 1701–1722.
- Valenti, J. A. & Piskunov, N. 1996, *A & AS*, 118, 595–603.
- Vogel, C. R. 2002, *Computational methods for inverse problems. Society for Industrial and Applied Mathematics.*
- Watson, V., Trouilhet, J., Paletou, F., & Girard, S. 2017, *IEEE International Workshop of Electronics, Control, Measurement, Signals and their Application to Mechatronics (ECMSM).*
- Wilkinson, D. M., Maraston, C., Goddard, D., Thomas, D., & Parikh, T. 2017, *MNRAS*, 472, 4297-4326.
- Xiang, M.-S., Liu, X.-W., Shi, J.-R., Yuan, H.-B., Huang, Y., Luo, A.-L. et al. 2017, *MNRAS*, 464, 3657-3678.

## Appendix

In this appendix, we display the results of the inversion of  $\log g$ ,  $[M/H]$ , and  $v \sin i$ . The black line corresponds to the 1-to-1 associated values. The test number, the root mean square error  $\Lambda$ , and the offset (see Tables 2 and 3 for details) are presented.



**Figure A1.** Results of the surface gravity inversion for the synthetic A to K stars.

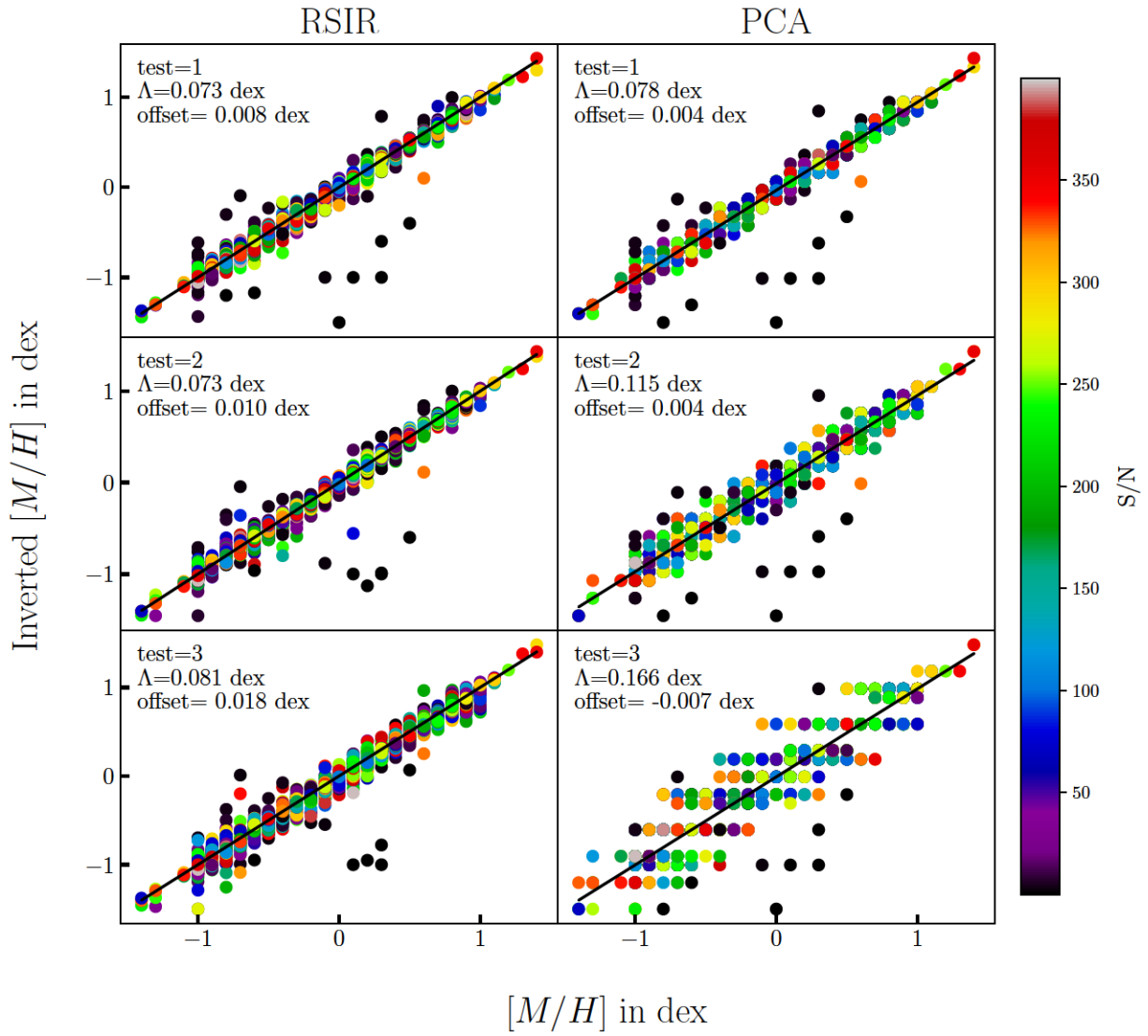
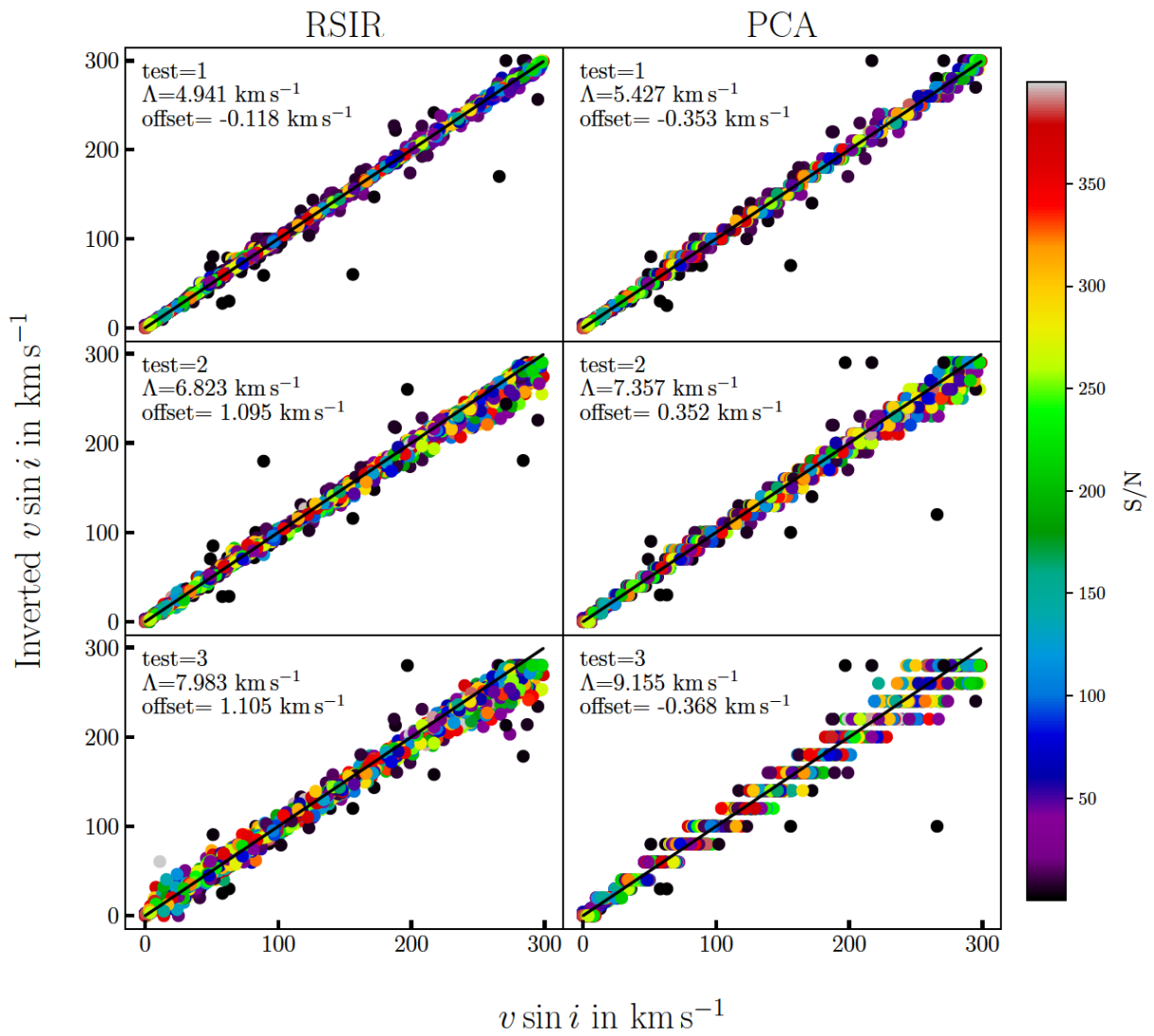


Figure A2. Results of the metallicity inversion for the synthetic A to K stars.



**Figure A3.** Results of projected equatorial velocity inversion for the synthetic A to K stars.