

PAPER • OPEN ACCESS

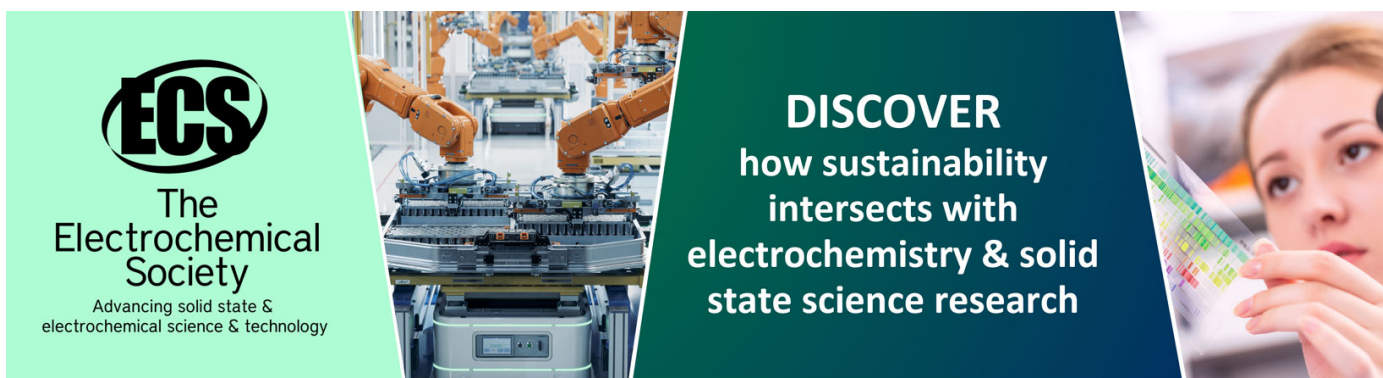
Pattern recognition in spectra

To cite this article: M Gebran and F Paletou 2017 *J. Phys.: Conf. Ser.* **869** 012075

View the [article online](#) for updates and enhancements.

You may also like

- [The Puzzle of the Metallic Line Stars](#)
Erika BöhmVitense
- [Elemental Abundance Analyses with DAO Spectrograms. XXXIV. A Three-Dimensional Graphical Examination of the Elemental Abundances of the Mercury-Manganese and Metallic-Line Stars](#)
Kutluay Yüce and Saul J. Adelman
- [The Chemical Compositions of the Two New HgMn Stars HD 30085 and HD 30963: Comparison to Lupi A, Cap. and HD 174567](#)
R. Monier, E. Griffin, M. Gebran et al.



ECS
The
Electrochemical
Society
Advancing solid state &
electrochemical science & technology

DISCOVER
how sustainability
intersects with
electrochemistry & solid
state science research

Pattern recognition in spectra

M Gebran¹ and F Paletou^{2,3}

¹Department of Physics and Astronomy, Notre Dame University-Louaize, PO Box 72, Zouk Mikael, Lebanon

²Université de Toulouse, UPS-Observatoire Midi-Pyrénées, IRAP, 31000 Toulouse, France

³CNRS, Institut de Recherche en Astrophysique et Planétologie, 14 av. E. Belin, 31400 Toulouse, France

E-mail: mgebran@ndu.edu.lb

Abstract. We present a new automated procedure that simultaneously derives the effective temperature T_{eff} , surface gravity $\log g$, metallicity $[Fe/H]$, and equatorial projected rotational velocity $v_e \sin i$ for stars. The procedure is inspired by the well-known PCA-based inversion of spectropolarimetric full-Stokes solar data, which was used both for Zeeman and Hanle effects. The efficiency and accuracy of this procedure have been proven for FGK, A, and late type dwarf stars of K and M spectral types. Learning databases are generated from the Elodie stellar spectra library using observed spectra for which fundamental parameters were already evaluated or with synthetic data. The synthetic spectra are calculated using ATLAS9 model atmospheres. This technique helped us to detect many peculiar stars such as Am, Ap, HgMn, SiEuCr and binaries. This fast and efficient technique could be used every time a pattern recognition is needed. One important application is the understanding of the physical properties of planetary surfaces by comparing aboard instrument data to synthetic ones.

1. Introduction

Pattern recognition techniques are being intensively studied in Astronomy due to the large amount of data that we are receiving from space telescope/mission and on-ground instruments such as the RAdial Velocity Experiment (RAVE; [1, 2]), the Sloan Digital Sky Survey (SDSS; [3]), the Gaia ESO Survey (GES; [4]), the LAMOST Experiment for Galactic Understanding and Exploration (LEGUE; [5]), the Apache Point Observatory Galactic Evolution Experiment (APOGEE; [6]), the High Efficiency and Resolution Multi-Element Spectrograph survey (HERMES; [7]), and the Gaia Radial Velocity Spectrograph (RVS) survey (e.g. [8]). Photometric and spectroscopic surveys are becoming numerous. To give an example, the Large Synoptic Survey Telescope (LSST) project will conduct a ten-year survey of the sky that will deliver about 15 terabytes (TB) of raw data per night [9]. The total amount of data collected over the ten years of operation will be 60 PB, and processing these data will produce a 15 PB catalogue database. The need of automated techniques that can handle and analyze this amount of information becomes evident.

In this context, different semi- and fully automated techniques have been developed in order to determine fundamental parameters and/or element chemical abundances in stars. Techniques such as pixel to pixel comparison based on χ^2 algorithms are the most common [10–12]. Other



techniques, based on multivariable analysis, have shown efficiency in reducing the dimension of physical problems [13, 14].

This work compiles the results of previous studies [14–16] in which we have applied an automated techniques, based on the Principal Component Analysis (PCA), to derive simultaneously effective temperatures (T_{eff}), surface gravities ($\log g$), equatorial projected rotation velocities ($v \sin i$), Metallicities ($[F/H]$), and microturbulence velocities (ξ_t) of A, F, G, K, and M stars. This technique has proven to be fast and accurate in inverting observed high-resolution spectra. Unlike classical techniques such as χ^2 fitting, PCA is based on a dimensionality reduction for the fast search of the nearest neighbour of an observed spectrum in a learning database.

2. Observations

For this study, we have acquired spectra of stars of different spectral types. High-resolution HARPS spectra of late-type dwarf K stars and early-type dwarf M stars were taken from the ESO Science Archive Facility. High-resolution spectra of A, F, G and K stars observed with Narval and ESPaDOnS spectropolarimeters were taken from the PolarBase (<http://polarbase.irap.omp.eu>). Narval is a spectropolarimeter, mounted on the 2-m TBL telescope, operating in the optical domain (380–1000 nm), with a spectral resolution of $R \sim 65\,000$ in its polarimetric mode. ESPaDOnS spectropolarimeter has mainly the same characteristics as Narval and is operating on the 3.6-m CFHT telescope. We have also retrieved A-type stars spectra from the ELODIE (<http://atlas.obs-hp.fr/elodie/>) and SOPHIE (<http://atlas.obs-hp.fr/sophie/>) archives. The ELODIE is a fiber-fed, cross-dispersed Echelle spectrograph that was attached to the 1.93-m telescope at Observatoire de Haute-Provence and had a resolution of $R \sim 42\,000$. SOPHIE replaced ELODIE in September 2006. SOPHIE spectra stretch from 3870 to 6940 Å at a resolution of $R \sim 75\,000$ in HR mode.

3. The learning database

Different types of learning databases were constructed for this project. One based on observed spectra and others based on calculated synthetic spectra. The first one was created after the Elodie stellar spectral library [17, 18]. The stellar parameters associated with each ELODIE spectrum were extracted from the CDS. The Synthetic databases were constructed using different sets of resolution ($R \sim 65\,000$, 76 000, and 42 000 for A stars, and $R \sim 115\,000$ for late-type dwarf K and early-type dwarf M stars). These different grids are constructed using a combination of different parameters such as T_{eff} , $\log g$, $[\text{Fe}/\text{H}]$, $v \sin i$, and ξ_t . Line-blanketed ATLAS9 model atmospheres [19, 20] were calculated for the purpose of this work. ATLAS9 models are LTE plane parallel and assume radiative and hydrostatic equilibrium. ATLAS9 uses the opacity distribution function (ODF) of [21]. We included convection in the atmospheres of stars cooler than 8500 K. Convection was treated according to Smalley’s prescriptions [22]. The grid of synthetic spectra was computed using SYNSPEC48 [23]. The effective temperatures of the model atmospheres vary depending of the spectral types with a typical step of 100 K. Logarithm of surface gravities vary between 2.0 and 5.0 with a step of 0.1 dex, respectively. The projected equatorial rotational velocities $v \sin i$ were varied from 0 up to 300 km s^{-1} with a non-constant step, metallicity was scaled from -2.0 dex up to +2.0 dex with respect to the solar value [24] with a step of 0.1 dex. The wavelength range that we used depend on the spectral type of the observations as explained in [14–16].

4. Principal component analysis and parameters derivation procedure

We represent all the spectra of the learning database in a matrix $\mathbf{S}$ of size N_{spectra} by N_{λ} . $\bar{S}$ will be the average of $\mathbf{S}$ along the N_{spectra} -axis. Then, we derive the eigenvectors $\mathbf{e}_k(\lambda)$ of the variance-covariance matrix $\mathbf{C}$ ($\mathbf{C} = (\mathbf{S} - \bar{S})^T \cdot (\mathbf{S} - \bar{S})$) using Singular Value Decomposition

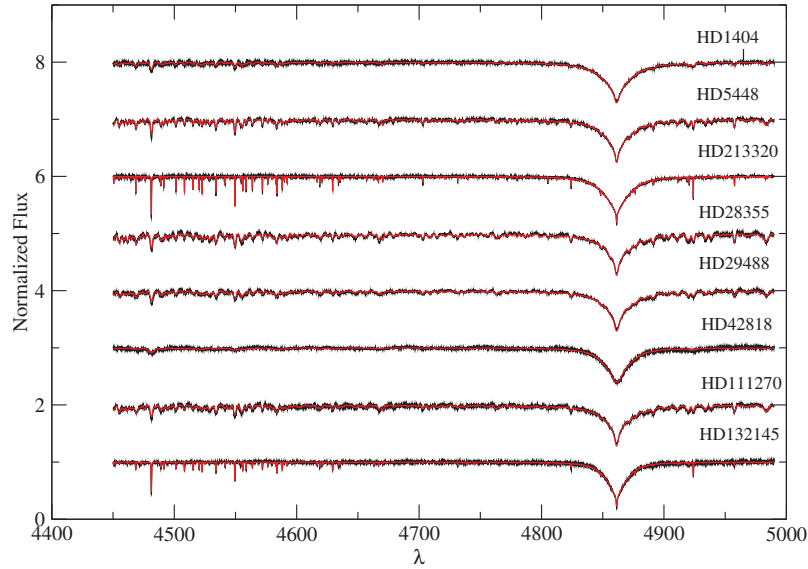


Figure 1. First neighbour fit of synthetic spectra (in red) to the observed spectra (in black). The wavelengths are in Å. Spectra were shifted up for clarity.

method or the power iteration method (see [14] for more details). The variance-covariance matrix has a dimension of $N_\lambda \times N_\lambda$. The eigenvectors of the symmetric $\mathbf{C}$ matrix are then sorted in decreasing eigenvalues magnitude; they are more usually referred to as “principal components”. Each spectrum of the dataset $\mathbf{S}$ is then represented by a small number of coefficients p_{jk} defined as

$$p_{jk} = (S_j - \bar{S}) \cdot e_k \quad (1)$$

The main point of reducing the dimensionality is the selection of $k_{max} \ll N_\lambda$. For all databases and in case of all spectral types and wavelength spectral ranges, we showed that $k_{max} = 12$ is a justified choice [14–16].

The observed spectra are denoted by $O(\lambda)$ having the same wavelength range and sampling as the learning dataset spectra. After correcting the observed spectra for radial velocities, we compute the set of projection coefficients ϱ_k defined as

$$\varrho_k = (O - \bar{S}) \cdot e_k \quad (2)$$

The nearest neighbour is then found by minimizing a χ^2 in the low dimensional space of the coefficients

$$d_j^{(O)} = \sum_{k=1}^{k_{max}} (\varrho_k - p_{jk})^2, \quad (3)$$

where j covers the number of synthetic spectra. The parameters of the synthetic spectrum having the minimum d will be considered for the inversion. For each observed spectrum, an iteration is performed, according to [25], in order to correct bad normalization as explained in [14]. An example of an overall spectral fitting between the first neighbour synthetic spectrum and several observation from our sample stars is shown in figure 1.

5. Results and discussion

In order to characterize the accuracy of the PCA technique, we compared our results to previous published values using Vizier catalogue (see [26] for more details). In overall, we found very good

agreement between our inverted values and the already published ones. It is worth mentioning that depending on the effective temperature and surface gravity of the star, learning databases are constructed in different wavelength regions. As spectral lines behave differently with T_{eff} and $\log g$, the affected errors do also vary with spectral types. On average mean signed differences between inverted and catalog values are smaller than the adopted spectroscopic errors on T_{eff} , $\log g$, and $[\text{Fe}/\text{H}]$. Typical errors of ± 150 K are found for T_{eff} , ± 0.15 dex for $\log g$, ± 0.1 dex for $[\text{Fe}/\text{H}]$, $\pm 5 \text{ km s}^{-1}$ for $v \sin i$. Large differences between catalog and inverted values yielded to detection of peculiar and/or binary stars as shown in [14]. In some cases, large inverted $v \sin i$ is an indicator of spectroscopic binary system and large or low metallic abundances is an indicator, in case of A and late B stars, of chemically peculiar stars such as Am, Ap, HgMn, SiEuCr... The PCA techniques is being also used in order to derive fundamental parameters of stars for the GES survey [27]. We are now investigating other multivariable techniques such as the Slice Inverse Regression (SIR) that was once used to retrieve Mars surface physical properties from spectroscopic data [28, 29].

References

- [1] Steinmetz M 2003, *GAIA Spectroscopy: Science and Technology* **298** 381
- [2] Munari U, Zwitter T and Siebert A 2005 *The Three-Dimensional Universe with Gaia* **576** 529
- [3] York D G *et al* 2000 *AJ* **120** 1579
- [4] Gilmore G *et al* 2012 *The Messenger* **147** 25
- [5] Deng L C *et al* 2012 *Research in Astronomy and Astrophysics* **12** 735
- [6] Majewski S R, Wilson J C, Hearty F, Schiavon R R and Skrutskie M F 2010 *Chemical Abundances in the Universe: Connecting First Stars to Planets* **265** 480
- [7] Zucker D B, de Silva G, Freeman K, Bland-Hawthorn J and Hermes Team 2012 *Galactic Archaeology: Near-Field Cosmology and the Formation of the Milky Way* **458** 421
- [8] Bailer-Jones C A L *et al* 2013 *A&A* **559** A74
- [9] Kantor J *et al* 2007 *Astronomical Data Analysis Software and Systems XVI* **376** 3
- [10] Allende P C *et al* 2006 *ApJ* **636** 804
- [11] Zwitter T *et al* 2008 *AJ* **136** 421
- [12] García Pérez A E *et al* 2016 *AJ* **151** 144
- [13] Traven G *et al* 2016 *arXiv:1612.02242*
- [14] Gebran M, Farah W, Paletou F, Monier R and Watson V 2016 *A&A* **589** A83
- [15] Paletou F, Böhm T, Watson V and Trouilhet J F 2015 *A&A* **573** A67
- [16] Paletou F, Gebran M, Houdebine E R and Watson V 2015 *A&A* **580** A78
- [17] Prugniel P, Soubiran C, Koleva M and Le Borgne D 2007 *arXiv:astro-ph/0703658*
- [18] Prugniel P and Soubiran C 2001 *A&A* **369** 1048
- [19] Kurucz R L 1992 *Rev. Mex. Astron. Astrofis.* **23** 45
- [20] Kurucz R L 2005 *Mem. Soc. Astron. It. Suppl.* **8** 14
- [21] Castelli F and Kurucz R L 2003 *Modelling of Stellar Atmospheres* **210** 20P
- [22] Smalley B 2005 *Mem. Soc. Astron. It. Suppl.* **8** 130
- [23] Hubeny I and Lanz T 1992 *A&A* **262** 501
- [24] Grevesse N and Sauval A J 1998 *Solar Composition and Its Evolution – From Core to Corona* 161
- [25] Gazzano J C, de Laverny P, Deleuil M *et al.* 2010 *A&A* **523** A91
- [26] Paletou F and Zolotukhin I 2014 *arXiv:1408.7026*
- [27] Farah W, Gebran M, Paletou F and Blomme R 2015 *SF2A-2015: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics* 3
- [28] Bernard-Michel C, Douté S, Fauvel M, Gardes L and Girard S 2009 *Journal of Geophysical Research (Planets)* **114** E06005
- [29] Watson V, Trouilhet J F, Paletou F and Girard S 2017 *IEEE (Proceeding of the International Workshop of Electronics, Control, Measurement, Signals and their application to Mechatronics)* (in preparation)