Stellar Populations as Building Blocks of Galaxies
Proceedings IAU Symposium No. 241, 2006
A. Vazdekis & R. F. Peletier, eds.

doi:10.1017/S1743921307007557

Filling Gaps in Indo-US Stellar Spectral Library using Principal Component Analysis

Harinder P. Singh[1], S. Jotin Singh[1], Ranjan Gupta[2] and M. Yuasa[3]

[1]Department of Physics & Astrophysics, University of Delhi, Delhi – 110 007, India
[2]IUCAA, Post Bag 4, Ganeshkhind, Pune 411 007, India
[3]Research Institute of Science & Technology, Kinki University, Osaka, Japan

Abstract. The Indo-US coudé feed stellar spectral library (CFLIB) published recently by Valdes *et al.* (2004) contains spectra of 1273 stars in the spectral region 3460 to 9464 Å at a resolution of 1 Å. About 500 stars in this database have gaps ranging from a few Å to several tens of Å in this wavelength range. We use a variation of Principal Component Analysis (PCA) technique to fill gaps of up to 5Å in a subset of spectra from the CFLIB. We hope to exploit the full potential of the scheme and attempt to fill larger gaps in stellar spectra in a subsequent study.

Keywords. Stars: Catalogs, Principal Component Analysis, Methods: Data Analysis

1. Principal Component Analysis for Data Reconstruction

We describe here the PCA based method for restoration of missing data for a set of spectra of 300 stars in the wavelength region 4000-4300 Å selected from the Indo-US CFLIB (Valdes *et al.* 2004). To begin with, we have 301 flux values at 1 Å interval in the range for 4000-4300 Å for 300 stars. For the i-th star, let F_j^i and w_j^i be the j-th observed flux value and its weight respectively, where j = 1,....., n (n = 301) and i = 1,....., N (N = 300). If a particular flux value F_j^i for a particular star is missing, its weight (w_j^i) is equal to zero.

For applying the PCA, the normalized data f_j^i is defined as

$$f_j^i = \frac{[F_j^i - \langle F_j \rangle]}{\sigma_j}, \tag{1.1}$$

where $< F_j >$ and σ_j are the mean and the standard deviation of F_j respectively. Following Unno and Yuasa (1992) and Singh *et al.* (2006), we define virtual data x_j^i and their corresponding weight v_j^i for each observed flux value f_j^i for each star as

$$v_j^i = 1 - w_j^i, \tag{1.2}$$

$$\sum_{i=1}^{N} v_j^i x_j^i = 0, \quad \sum_{i=1}^{N} v_j^i (x_j^i)^2 = \sum_{i=1}^{N} v_j^i. \tag{1.3}$$

Eqn. (1.3) represents the statistical constraint that the mean value of the virtual data is zero and the standard deviation is unity. The most probable value of x_j^i are thus given by the following set of n simultaneous linear algebraic equations:

$$\sum_{l=1}^{n} \frac{1}{\lambda_l} \left[\mu_{lj}^2 x_j^i + \sum_{k \neq j} \mu_{lj} \mu_{lk} (w_k^i f_k^i + v_k^i x_k^i) \right] = 0, \quad (j = 1, ..., n), \tag{1.4}$$

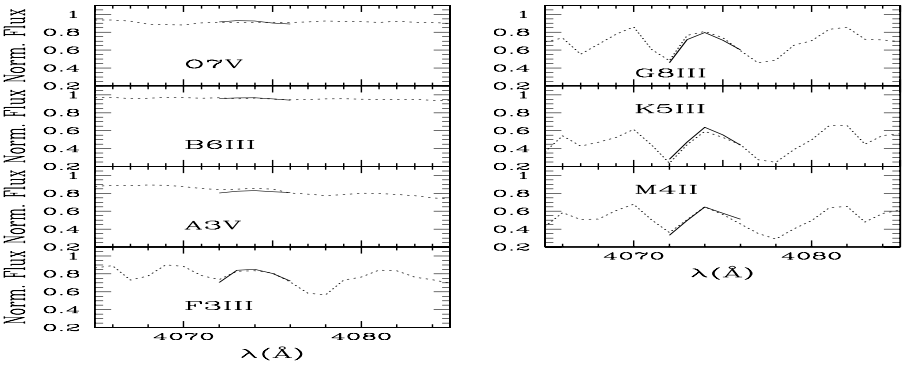

Figure 1. Reconstructed fluxes (solid lines) plotted together with the original normalized fluxes (dashed lines) for seven representative spectral types. Five fluxes were reconstructed in the range 4072 Å to 4076 Å.

where λ_l is the l-th eigenvalue and μ_l^j represent the j-th component of the l-th eigenvector in the PCA. To check the veracity of this procedure let us assume that from the data set only one flux value F_1^s is missing. This means that $w_1^s = 0$ and all the other weights w_j^i except w_1^s are equal to unity. In this simplified situation, Eqns. (1.4) can be easily solved to get x_1^s for the missing flux f_1^s. By exchanging the columns, one can compute x_2^s, x_3^s, and so on for the missing flux for any wavelength and for any star.

2. Results and Discussion

We tested the validity of this flux reconstruction method by attempting to reproduce fluxes around a strong absorption feature, namely the 4077 Å SrII feature, visible in the representative spectra in Fig. 1. We use a flux region of 20 Å starting from 4076 Å and thus 20 principal components to reconstruct the fluxes at 4076 Å for all the 300 stars. The procedure assumes a gap (zero flux) at 4076 Å and goes on to compute this missing flux. We find that 256 stars have restoration error $(f_1^i - x_1^i)$ within ± 0.1. For most of the stars the reconstruction is good with a standard deviation of 0.0278.

Next we repeat the procedure with 20 principal components with starting wavelength of 4075 Å in the region 4075Å - 4094 Å. The flux used at 4076 Å is the earlier restored flux. The method, starting with the assumption that the first flux (at 4075 Å) is missing, proceeds to calculate this missing flux. In this case, 269 stars have restoration error $(f_1^i - x_1^i)$ within ± 0.1. The standard deviation is 0.0411 for the flux at 4075 Å. We carry forward the procedure to reconstruct fluxes at successive wavelengths of 4074 Å, 4073 Å, and 4072 Å. Fig. 1 shows the original and reconstructed fluxes for seven representative spectral types picked out of the 300 stars. The reconstruction is good for all the spectral types. The procedure was also extended to restoring fluxes at other wavelengths in the interval 4000-4300 Å with similar accuracies. Another interesting offshoot of this analysis was in picking outliers, stars which have either no known MK spectral types or have noisy spectra.

References

Singh, H.P., Yuasa, M., Yamamoto, N. & Gupta, R. 2006, PASJ 58, 177
Unno, W. & Yuasa, M. 1992, Ap&SS 189, 271
Valdes, F., Gupta, R., Rose, J. A., Singh, H. P. & Bell, D. J. 2004,ApJS 152, 251 (CFLIB)

https://doi.org/10.1017/S1743921307007557 Published online by Cambridge University Press