

Removing IFU signatures with Wavelet and PCA combined

Fabricio Ferrari¹, João E. Steiner², Tiago V. Ricci², Roberto B. Menezes²

1. Universidade Federal do Pampa, Bagé, RS
2. Instituto de Astronomia, Geofísica e Ciências Atmosféricas, Universidade de São Paulo, São Paulo, SP

Introduction

We present a new data cube analysis method based on discrete wavelet and principal component transforms. The method consist in decomposing the data cube in wavelet and principal components spaces, remove unwanted structures, and then reconstruct the data cube. The method is used to remove instrumental signatures in integral field units (IFU) data. IFU signatures are structures present both in spatial and spectral dimensions; usually they cannot be removed with standard techniques without affecting the science data. Many IFUs present these signatures and no general technique has been developed so far to attack the problem. Using discrete wavelets and principal components we can isolate the signature and remove it. As an example of the technique, we apply it to Gemini/GMOS IFU data of the galaxy NGC 1399.

Principal Component Analysis – PCA

Besides the relevant information, any set of data has some amount of redundant information and noise. The goal of PCA is to find a new basis on which the data in written in a more meaningful way. The basis vectors – **the principal components** – are chosen so that they contain the maximum variance. There is as many vectors as the original variables, but usually the first few vectors contain most of the information of the data set: they provide a more concise description of the data. Often the physical content are restricted to few vectors and thus interpretation can be easier.

Mathematically we proceed as follows. In a basis where the variables are all uncorrelated (orthogonal) their covariance or correlation matrix is diagonal. We then find the principal components by diagonalizing the covariance or correlation matrix of the original data. In the case of covariance matrix, the **eigenvectors** are the principal components and the **eigenvalues** are the variance associated with each of the eigenvectors. (See Figure 2)

In the case of a hyperspectral cube, it is informative to measure the correlation between each principal component and the spectra in each spatial pixel: the **tomograms**. For example, tomogram 1 is the projection (scalar product) of eigenvector 1 and the spectra relative to each spatial pixel. Tomograms are shown in Figure 2. For a full description on the method applied to astronomical data, see Steiner et al. 2009.

Discrete Wavelet Transform

The wavelet transform consists of describing a signal in term of a basis of functions which are square integrable and have compact support (Mallat 1999). In the case of time series, the transform is a time frequency representation of the signal with **good time and frequency localisation**. The continuous wavelet transform W_x of a function $x(t)$ is the scalar product between the function and scaled and shifted versions of the mother wavelet $\psi_{a,b}$:

$$W_x(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \psi^* \left(\frac{t-b}{a} \right) dt,$$

where the instant (or position) b and the period (or scale) a vary continuously.

In the case of a discrete signal $x[n]$, then a and b assume discrete values, the integral becomes a sum, and we have the **discrete wavelet transform**, which is a sparse representation of the signal. Dyadic translations and dilations are commonly used, where $a = 2^{-j}$ and $b = k2^{-j}$ for $j, k \in \mathbb{Z}$. The discrete wavelet transform is a **multiresolution or multiscale** analysis because it splits the signal into different components each with a characteristic scale.

The WPCA Algorithm

The idea behind the use of the wavelet and the PCA transforms together is to decompose the original signal both in the wavelet and in the PCA space, then select unwanted features and finally reconstruct the signal (Steiner et al. 2010). The process is shown in Figure 1. The original data cube (orange) is first decomposed in wavelets scales W_0, W_1, \dots, W_k . Then, the PCA analysis is performed in each of the wavelet components, resulting in several eigenvectors for each scale. After removing those eigenvectors which contain unwanted structures (black dots), each of the wavelets components is restored with the PCA reconstruction, PCA^{-1} , and then the resulting corrected wavelets W'_0, W'_1, \dots, W'_k are combined to form the resulting corrected cube (green).

Acknowledgements



The WPCA Algorithm

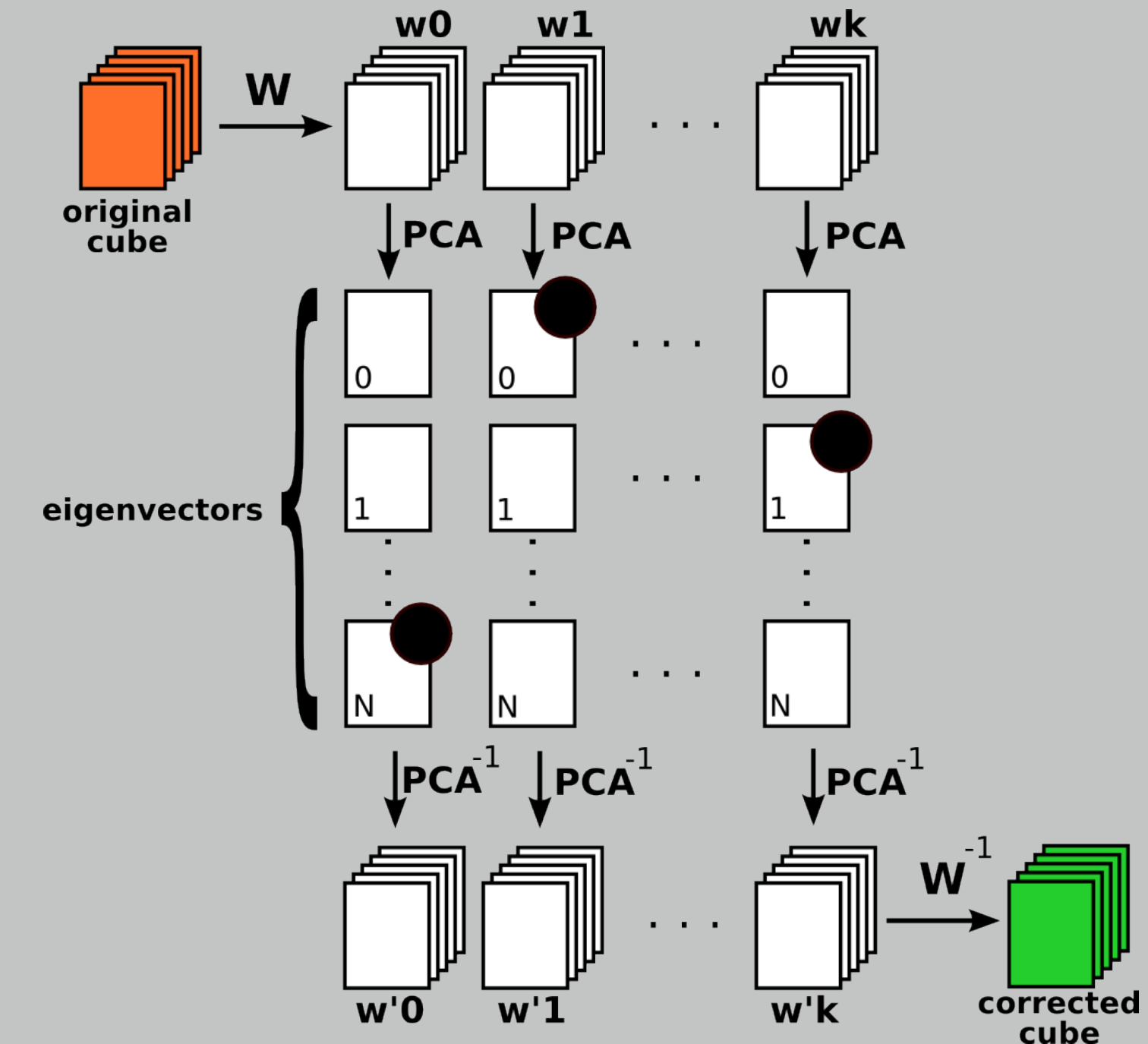


Figure 1: Wavelet and PCA combined: the original data cube is first decomposed W in wavelet and then the PCA analysis is performed. After correcting for unwanted effects (instrumental signatures, fingerprints or noise, for example) the cube is reconstructed.

NGC 1399

The method was applied to Gemini/GMOS observations of the elliptical galaxy NGC 1399. In this case, there is a noticeable **instrumental signature** which present itself as **vertical stripes** in the tomograms (plots marked with large black dots in Figure 2). This signature is also found in many other IFUs with optical fibers. The problem in these case are **the vertical stripes which cannot be removed with standard techniques**. High frequency noise can be removed with Butterworth filters (see Ricci et al. 2010.) Both the eigenvectors and tomograms of the original data (column **O** in Figure 2) show this characteristic structure, which can be identified in the wavelet-PCA decomposition as a single component, thus permitting its removal.

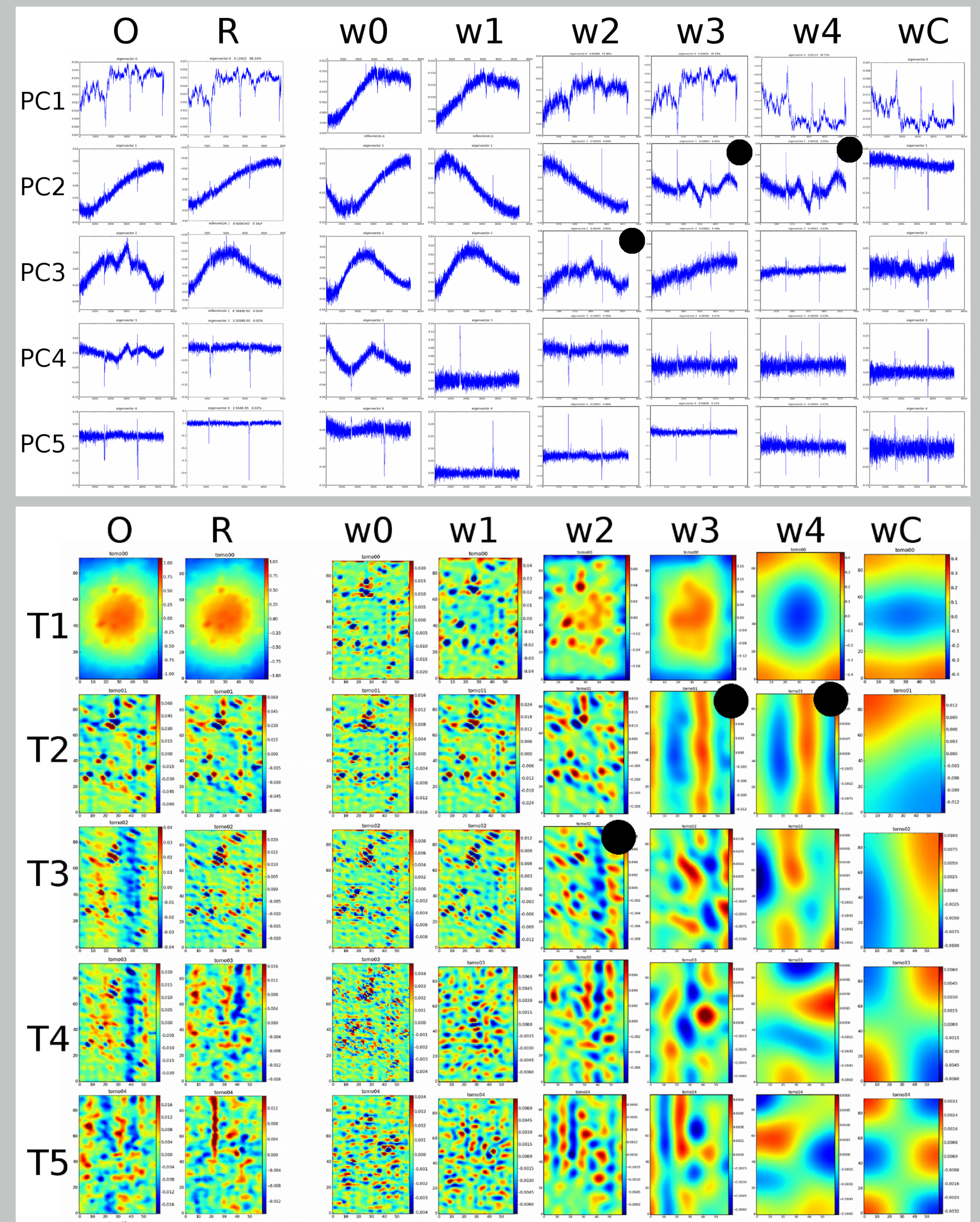


Figure 2: **NGC 1399**: eigenvectors (**PC**, top plots) and tomograms (**T**, bottom images) of the original (**O**), reconstructed (**R**) and each wavelet component (**W_k**) of the data. Black dot ● indicates components that were removed from **O** to build **R**

Conclusions

It is presented a new method for data analysis which combines wavelet and principal component analysis techniques. The technique is able to remove instrumental IFU signatures. The data is represented in wavelet and PCA spaces where features can be selected and removed. As an example, it is shown the case of Gemini/GMOS signature in the observations of the galaxy NGC 1399, where a vertical structure is identified and then removed.

References

- J. E. Steiner, R. B. Menezes, T. V. Ricci, A. S. Oliveira, *PCA Tomography: how to extract information from data cubes*, MNRAS, **395**, 64, 2009.
- S.G. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, 1999.
- J. E. Steiner, F. Ferrari, T. Ricci, R. Menezes, 2010, in preparation.
- T. V. Ricci, J. E. Steiner, R. B. Menezes, *Remoção de Ruídos de Cubos de Dados do GMOS-IFU*, XXXV Reunião da SAB (poster 244), 2010.