

PCA in Astronomy

or

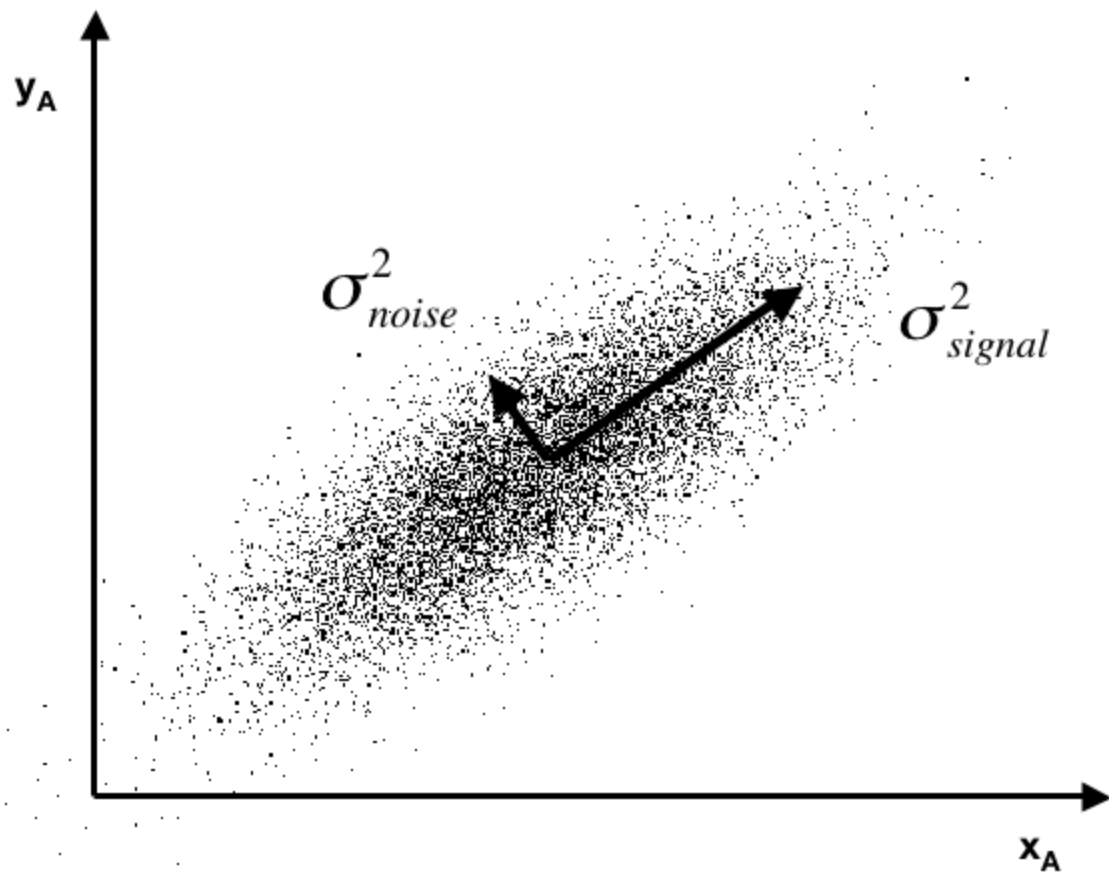
PCA Tomography: how to extract information from data cubes

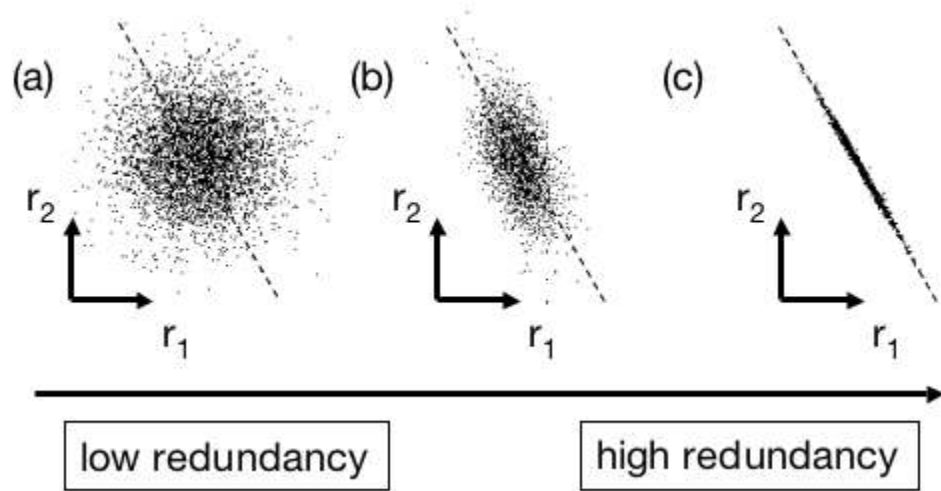
João Steiner

IAG - Universidade de São Paulo

IWCCA-UNICSUL

3 de junho de 2011





Data cubes

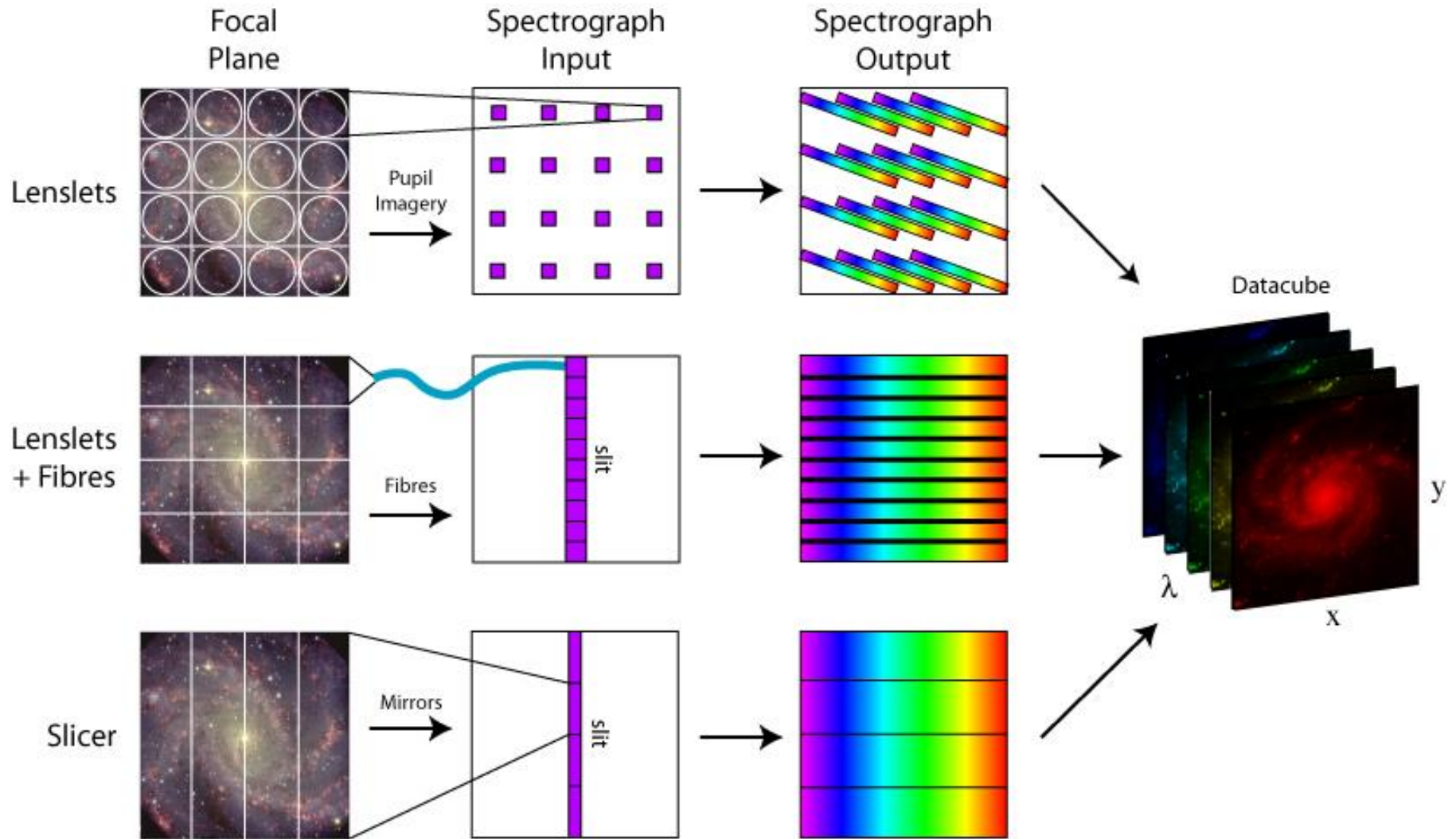
Traditional astronomical data:

- *Spectrum: λ (pixel)*
- *Image: $x.y$ (spaxel)*

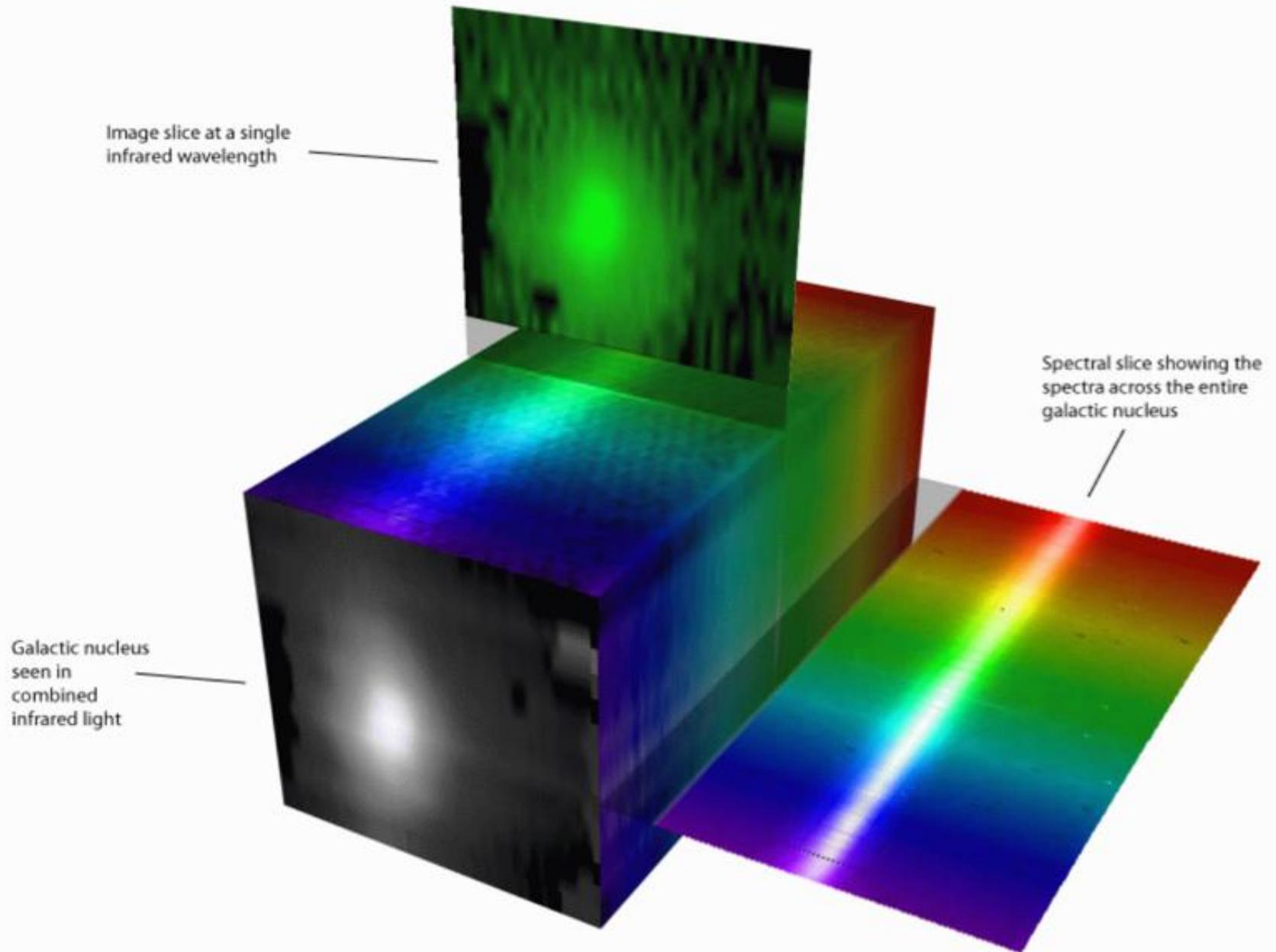
Integral Field Unit (IFU) spectroscopy:

- *3D spectroscopy*
- *Data cube: $x.y.\lambda$*
- *Gemini – SOAR – JWST (5 IFUs)*

IFU spectrographs



Data Cubes



Principal Component Analysis - PCA

- Data cubes have huge number of pixels (6×10^6). How can we extract information?
- PCA tomography:
- Not a set of objects; a set of spatial pixels of the same data cube. The wavelength pixels are the properties.
- Linear transformation to a new system of coordinates
- The coordinates are orthogonal
- Dimensional reduction
- A lot of redundancy!
- Noise reduction – background subtraction
- Data organization and analysis

From a datacube to a data matrix

- The datacube has $n = v \times \mu$ spatial pixels and m spectral pixels.

- The mean intensity
$$Q_\lambda = \frac{1}{n} \cdot \sum_{i=1}^{\mu} \sum_{j=1}^v (I_{ij\lambda})_O$$

- The intensity adjusted
$$I_{ij\lambda} = (I_{ij\lambda})_O - Q_\lambda$$

- The data cube $I_{ij\lambda}$ has to be transformed into a matrix, $I_{\beta\lambda}$

where $\beta = \mu(i-1) + j$

PCA — Principal Component Analysis

- Covariance matrix
$$C_{\text{cov}} = \frac{[I_{\beta\lambda}]^T \cdot I_{\beta\lambda}}{n-1}$$

- Properties
$$C_{\text{cov}} = [C_{\text{cov}}]^T$$

- PCA transformation:
$$T_{\beta k} = I_{\beta\lambda} \cdot E_{\lambda k}$$

- D must be diagonal:
The diagonal elements of D are the eigen-values Λ_k
$$D_{\text{cov}} = \frac{[T_{\beta k}]^T \cdot T_{\beta k}}{n-1}$$

Eigen-spectra and Tomograms

- $E_{\lambda k}$ are the eigen-spectra

and, (transforming β into ij)

- T_{ijk} are the tomograms

Analysing both ***together*** may reveal a wealth of information!

- Λ_k are the eigenvalues

Reconstruction, Compression Cosmetics

- Reconstructing the datacube with relevant eigenvectors to $k=r$.

$$I'_{\beta\lambda}(\leq r) = T_{\beta k}(\leq r) \cdot [E_{\lambda k}(\leq r)]^T$$

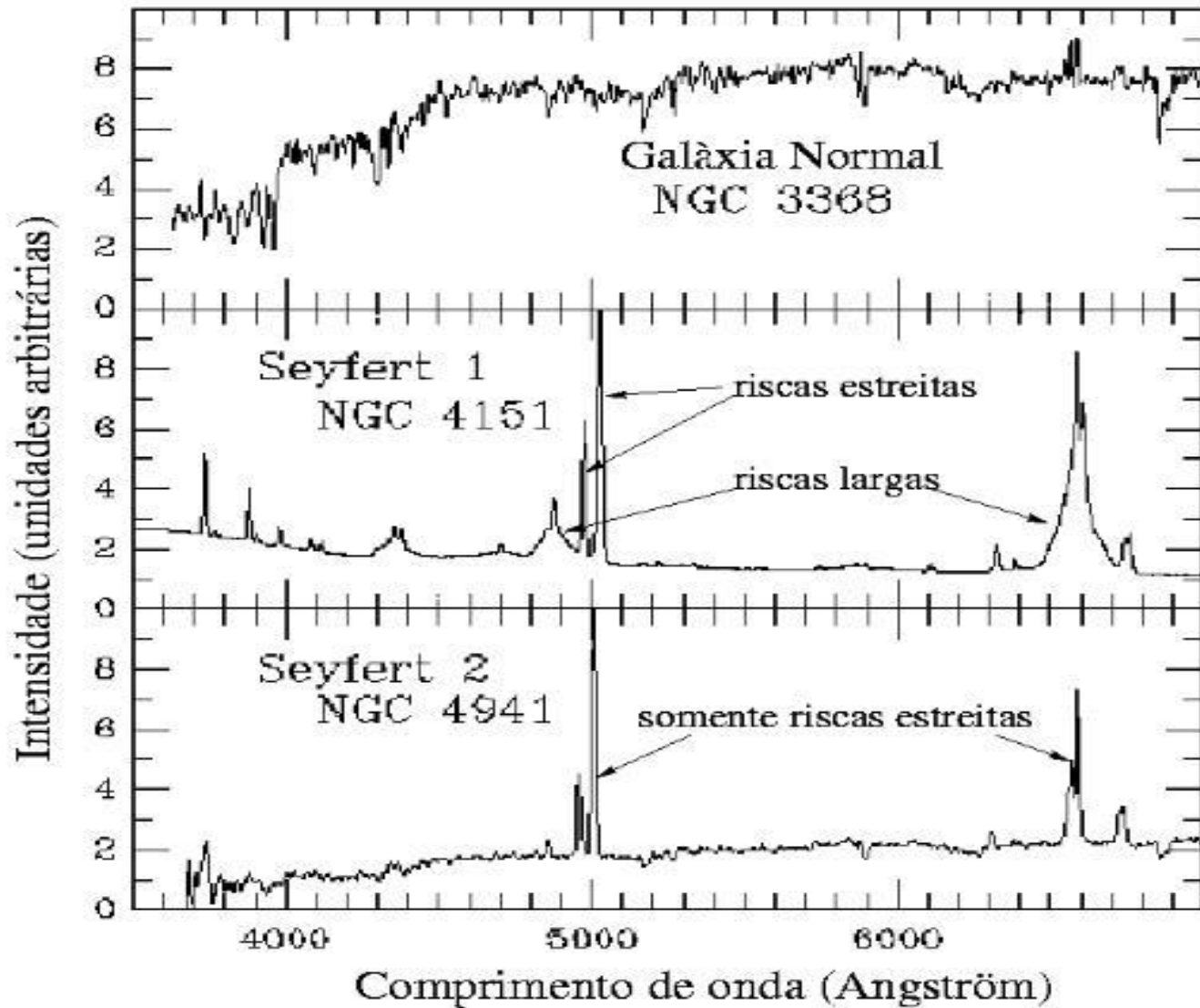
- Calibrating flux back

$$(I'_{ij\lambda}(\leq r))_O = I'_{ij\lambda}(\leq r) + Q_\lambda$$

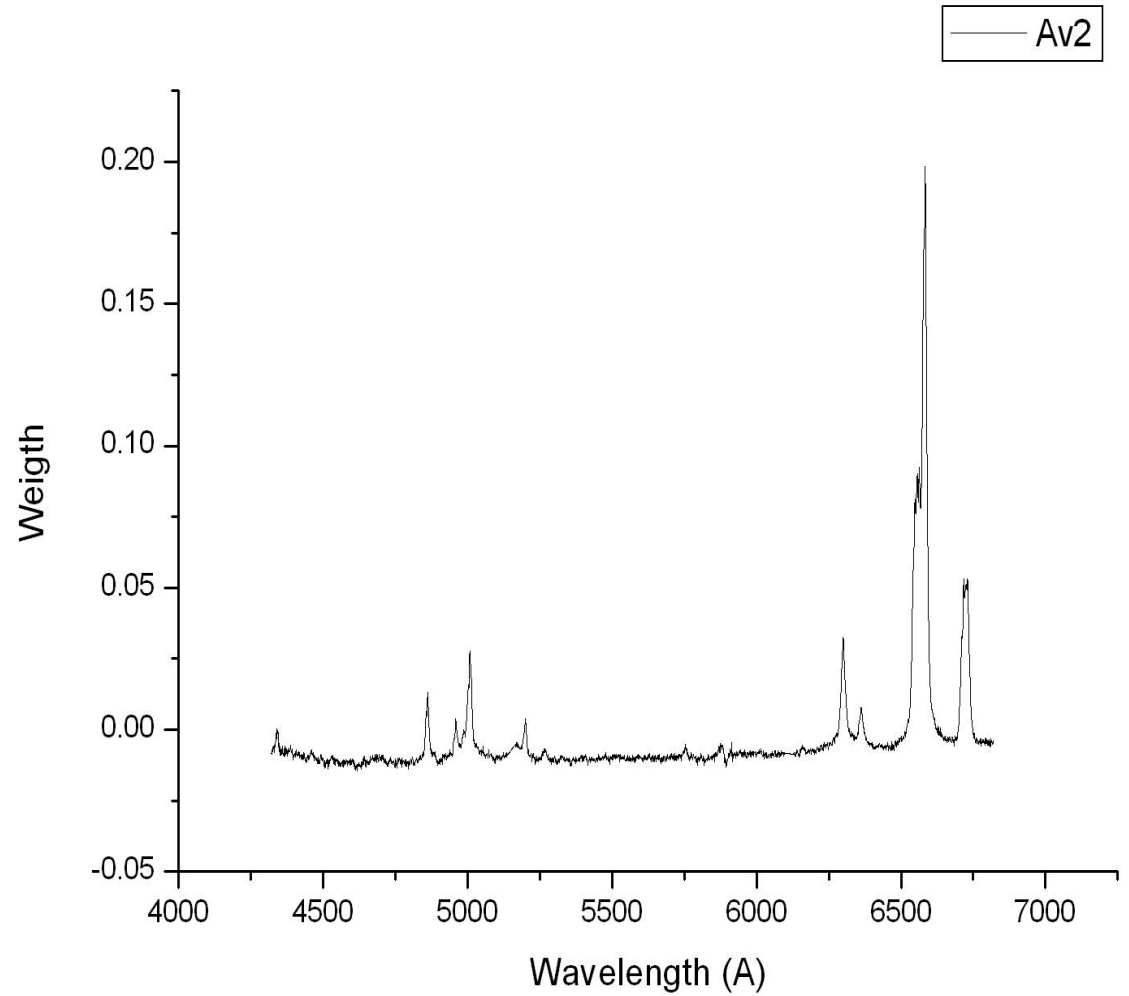
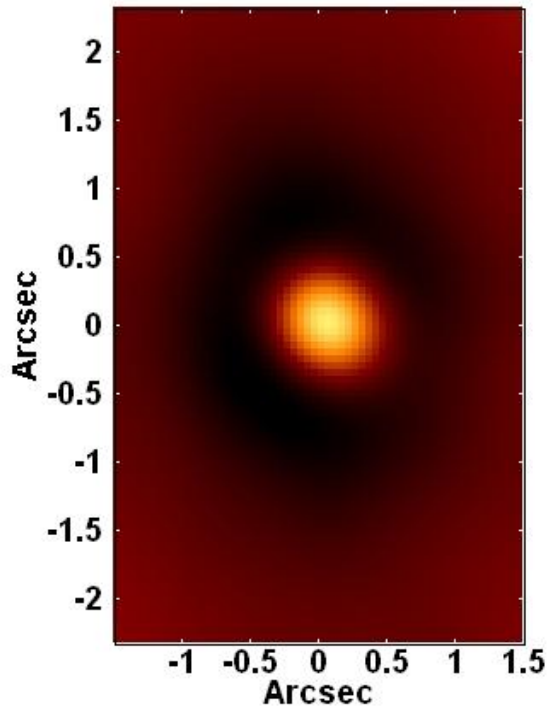
- Noise may be evaluated as

$$\sigma^2 = \sum_{k=r+1}^{k=m} \Lambda_k$$

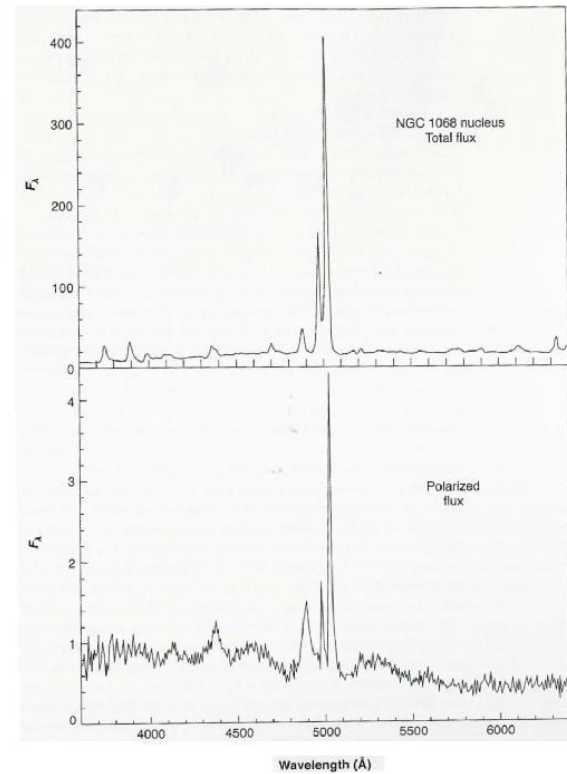
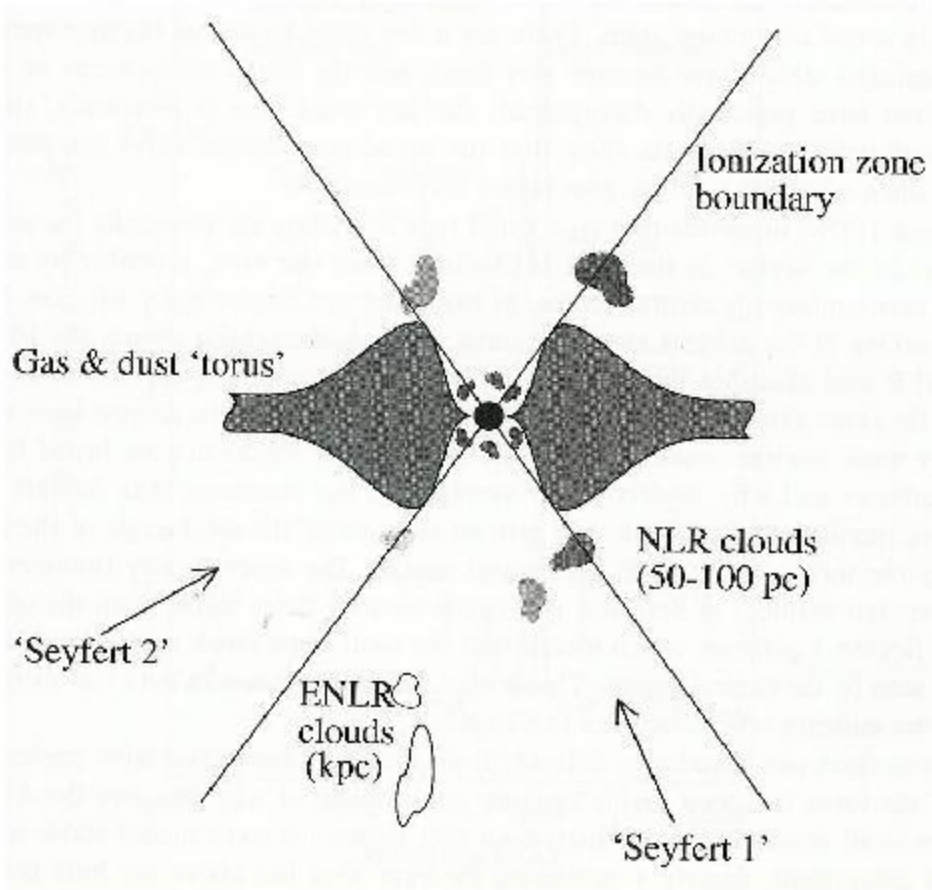
Seyfert 1 vs 2



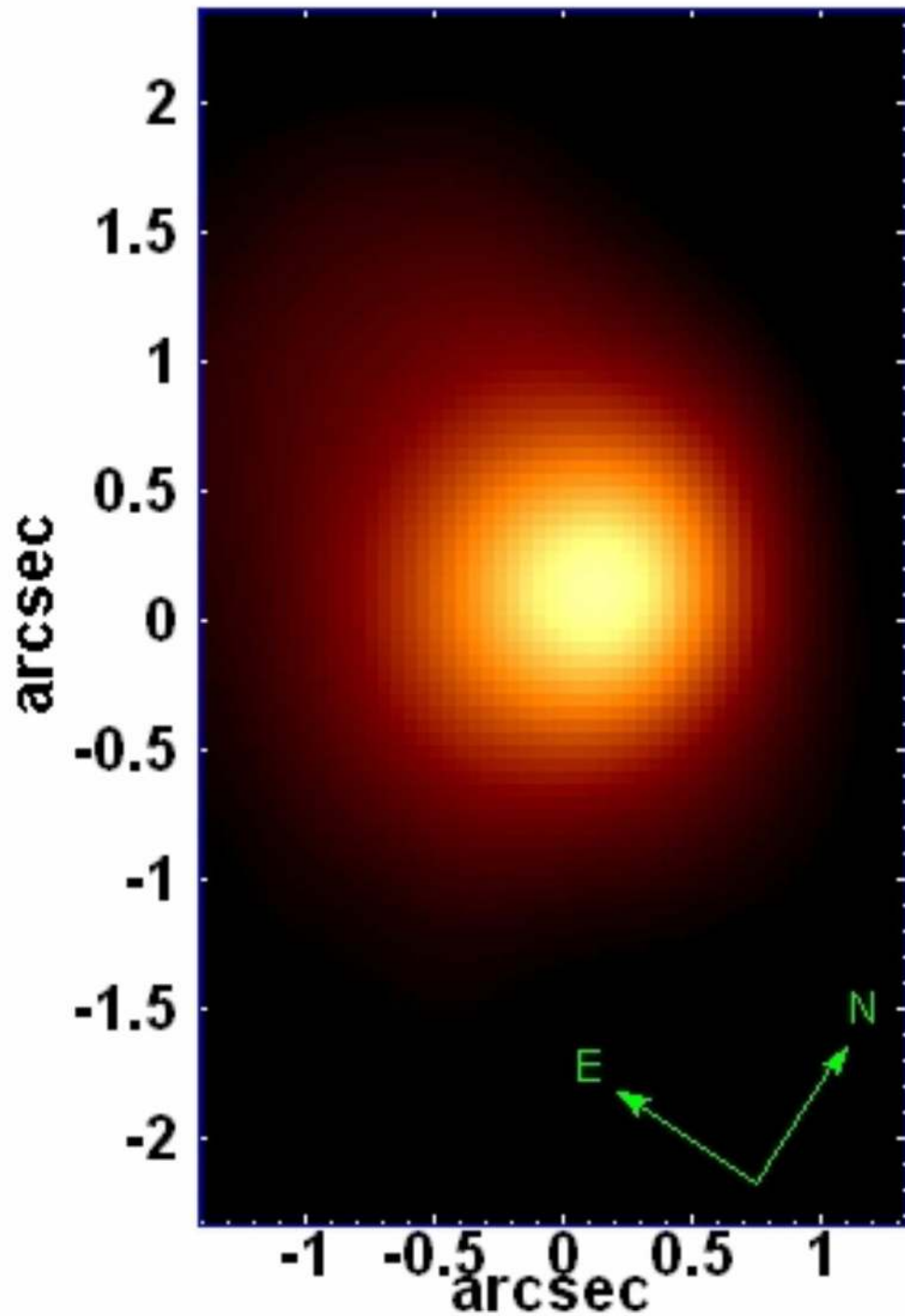
IC1459 – Av 2



AGN – the unified model

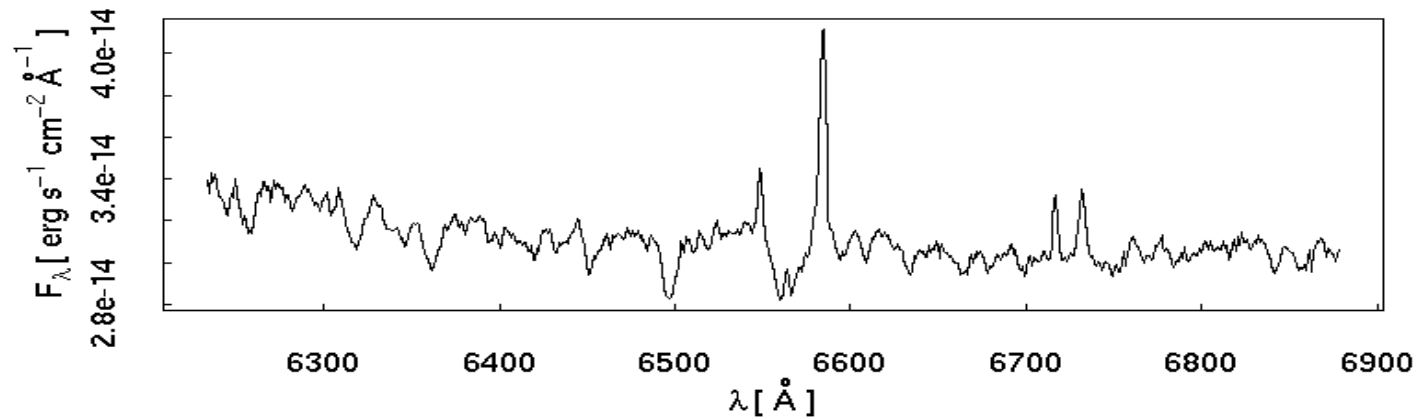
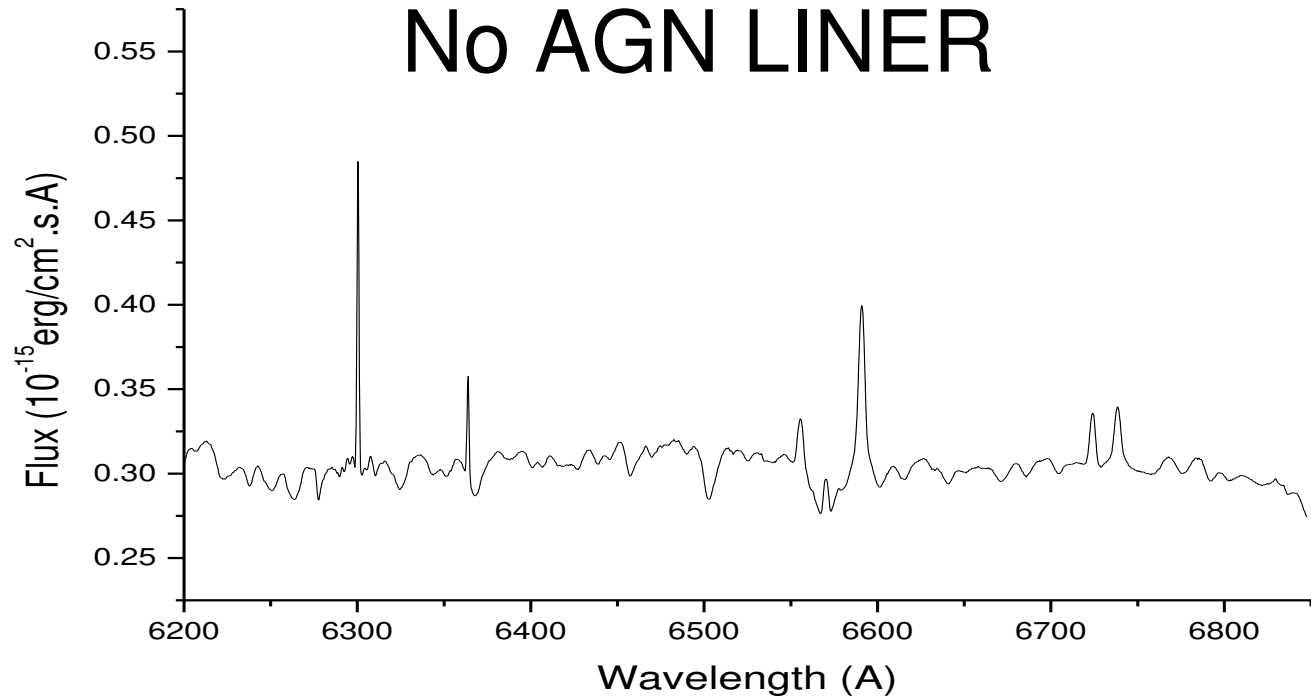


NGC 4736 (5 Mpc)
GMOS – IFU
Steiner et al 2009, MNRAS



NGC 4736 – Gemini 8m x Palomar 5m

No AGN LINER

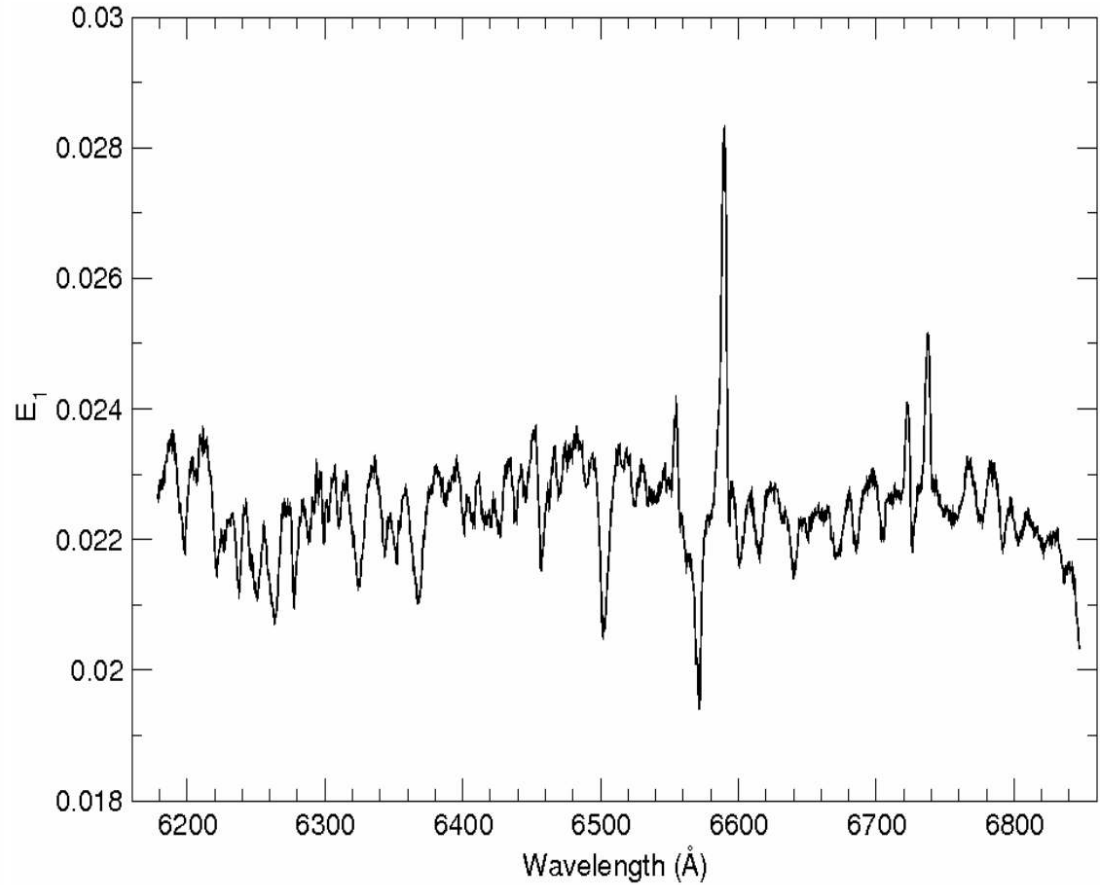
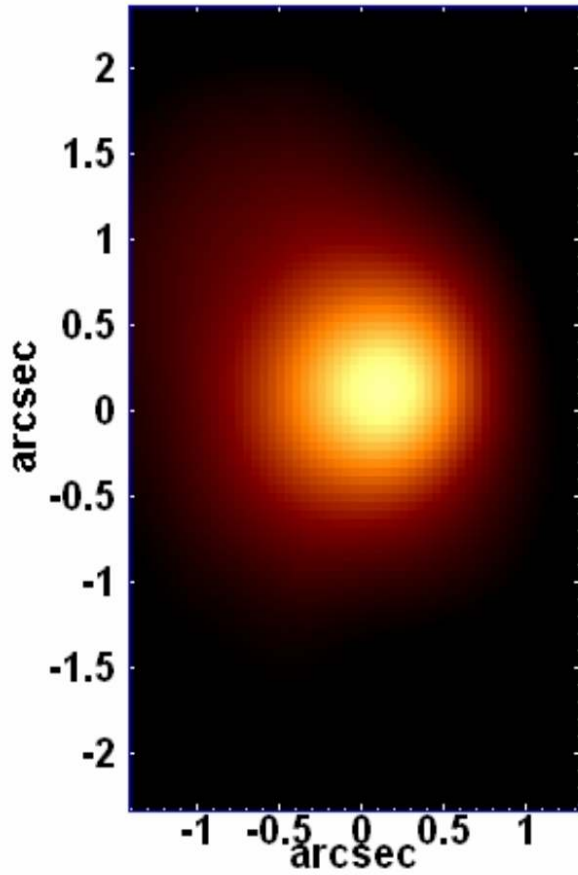


Eigenvalues and variance explained

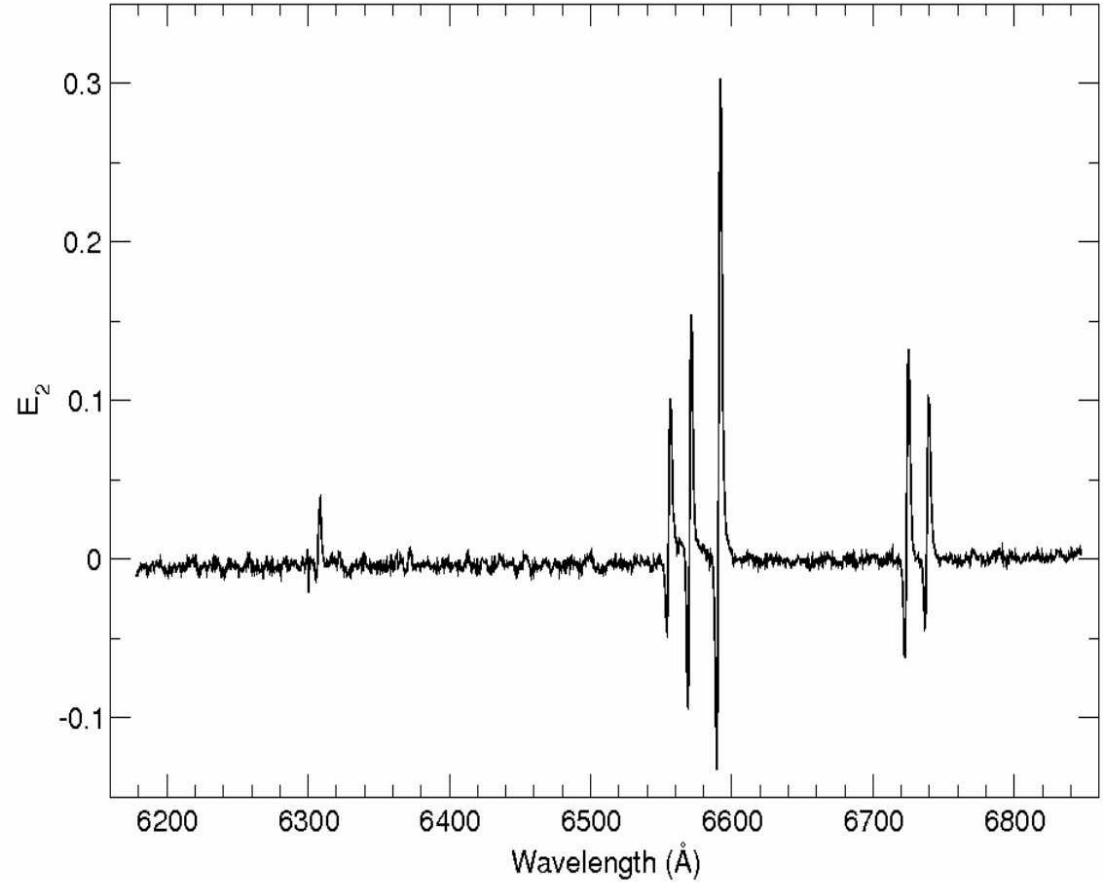
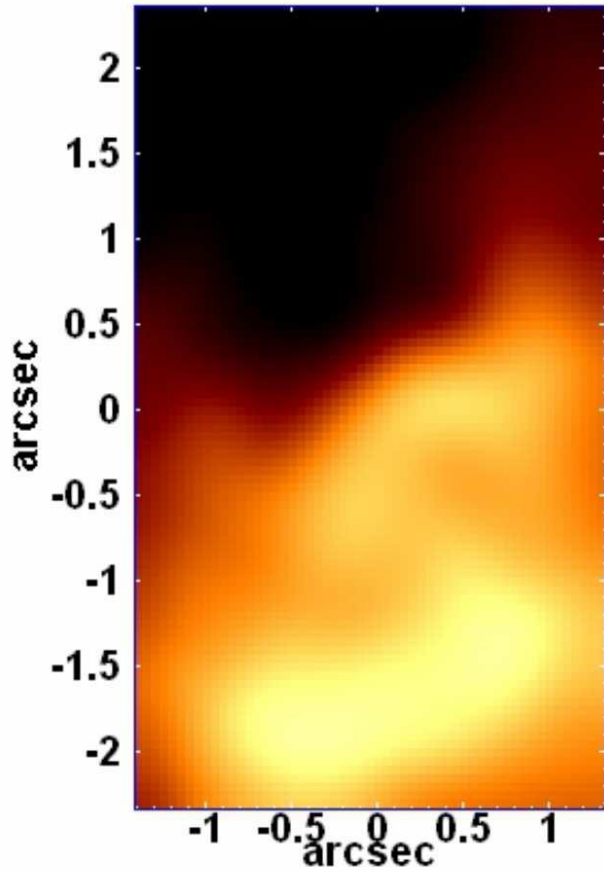
(Huge redundancy)

Eigenvector E_k	Eigenvalue (% of the variance)	Accumulated fraction (% of the variance)
E_1	99.7443	99.7443
E_2	0.0883	99.8326
E_3	0.0325	99.8651
E_4	0.0129	99.8781
E_5	0.0084	99.8864
E_6	0.0048	99.8912
E_7	0.0039	99.8952
E_8	0.0027	99.8979

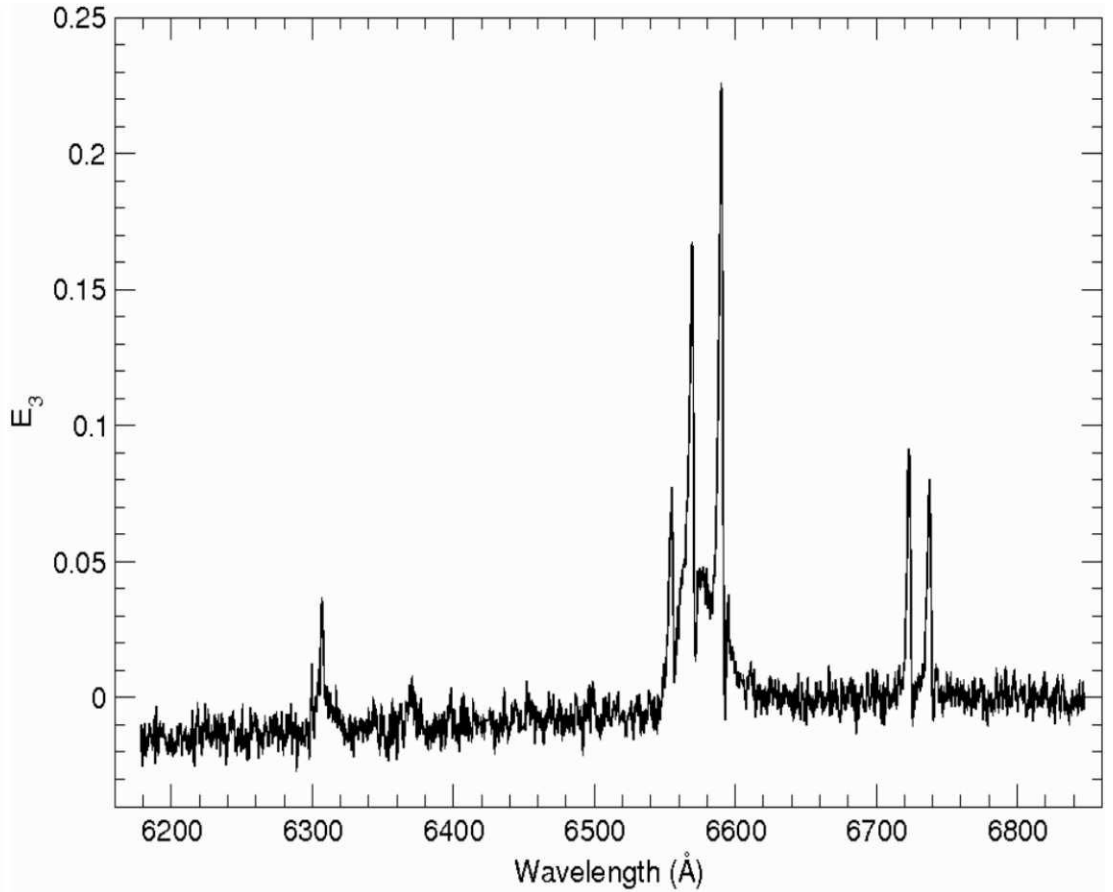
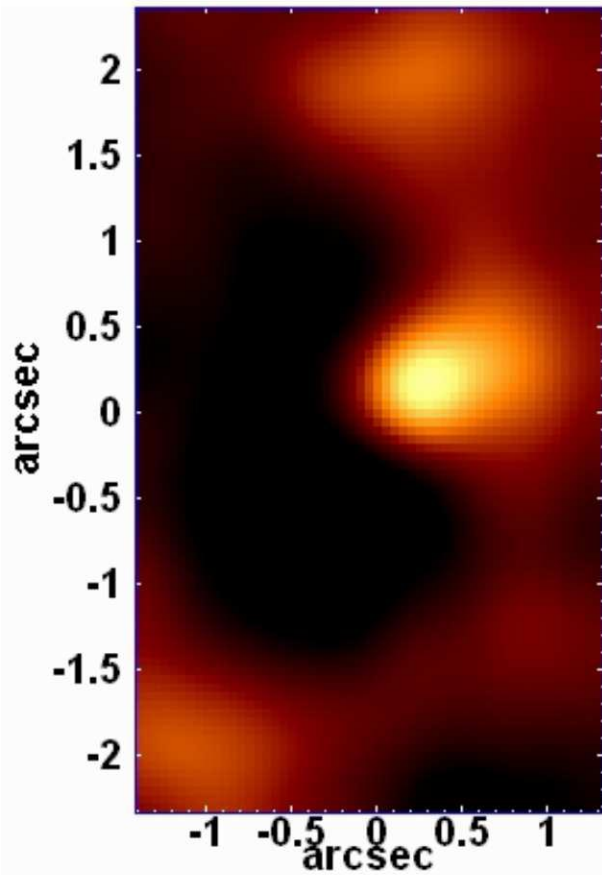
Tomogram 1 and eigenspectrum 1

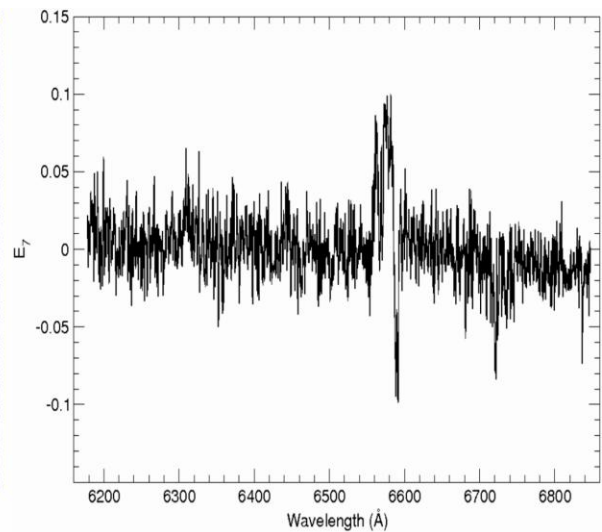
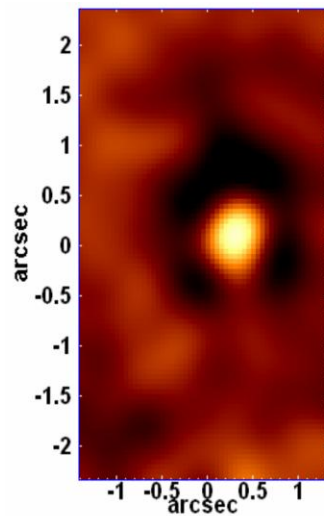
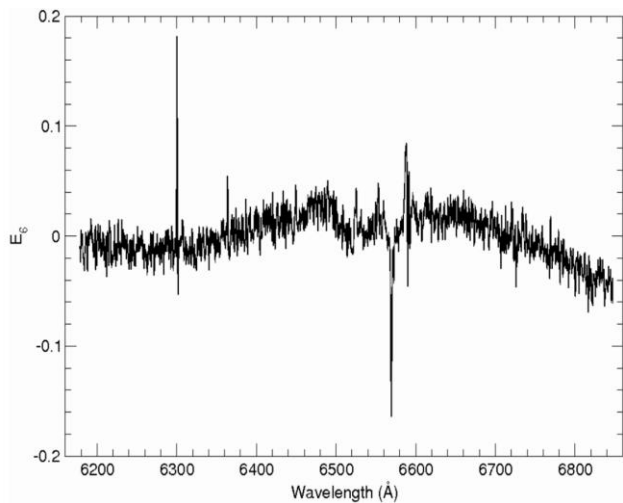
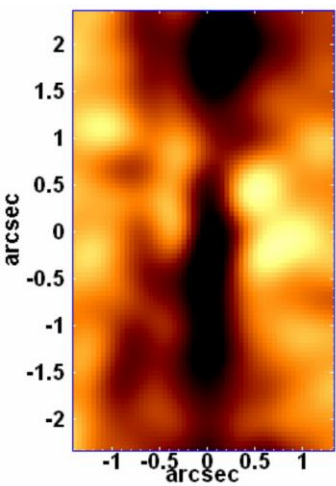
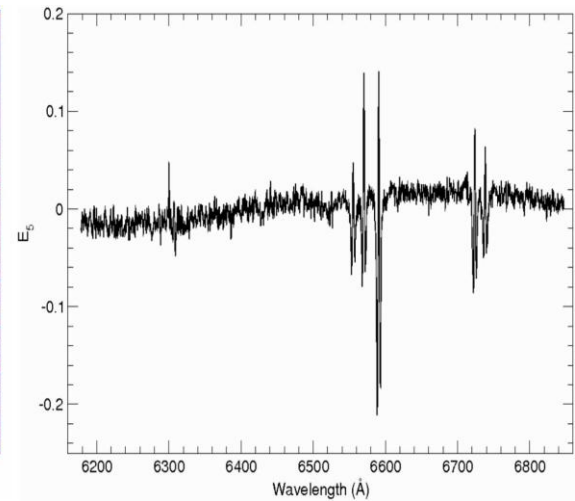
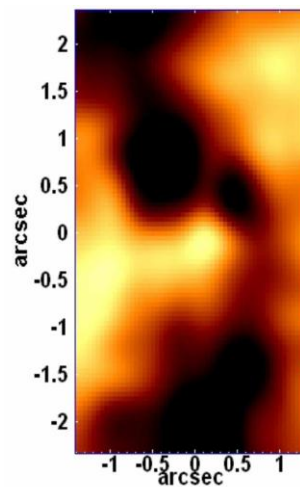
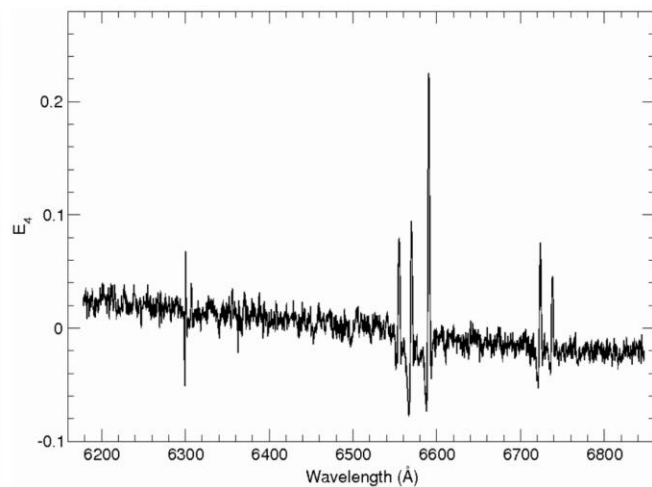
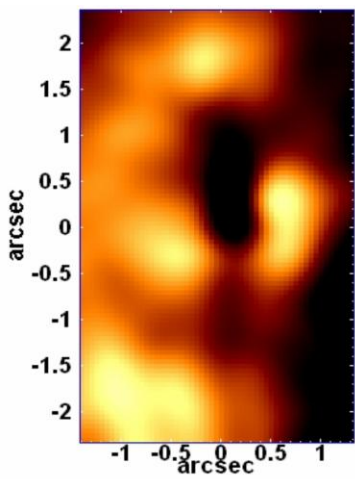


Tomogram 2 and eigenspectrum 2

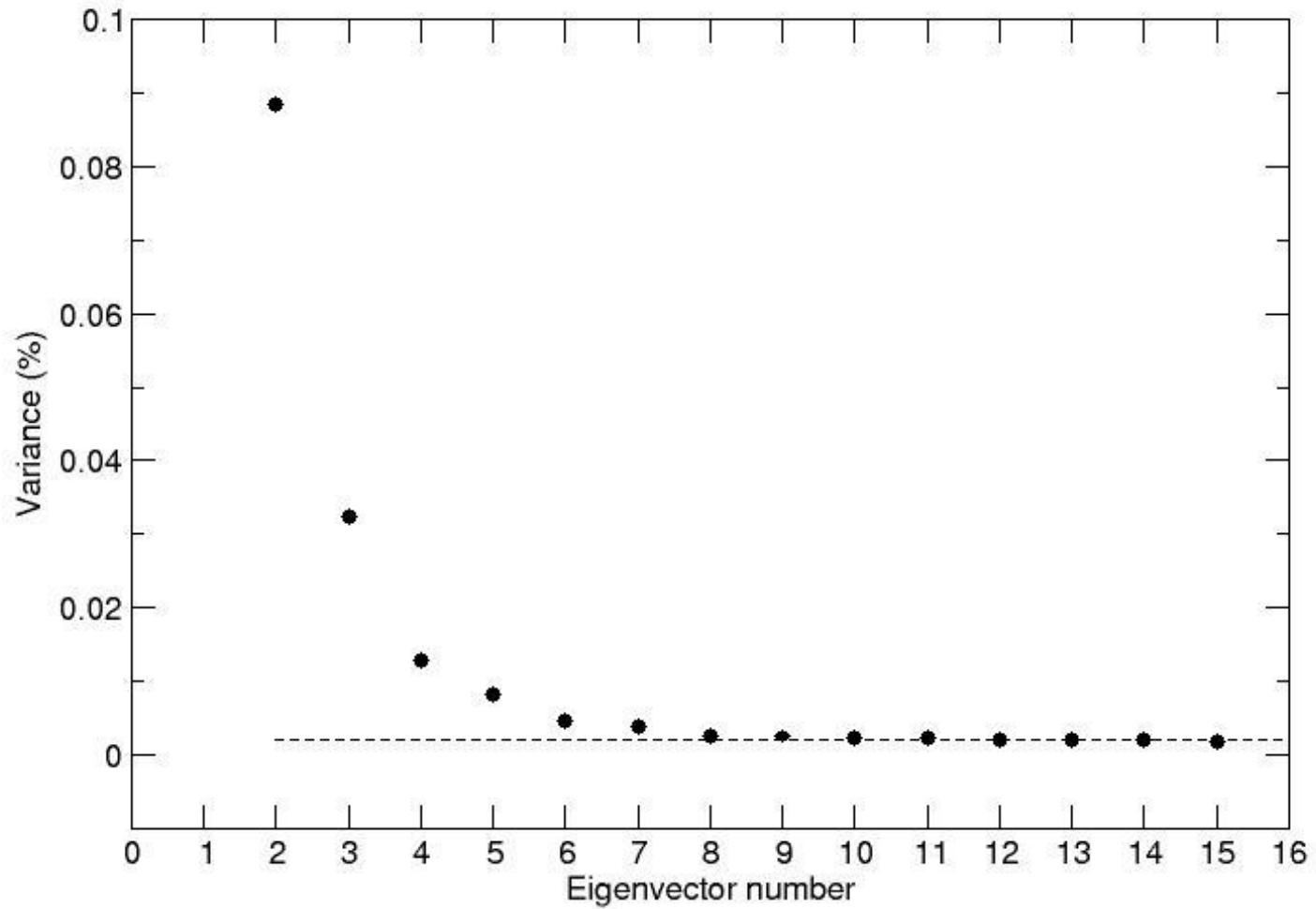


Tomogram 3 and eigenspectrum 3



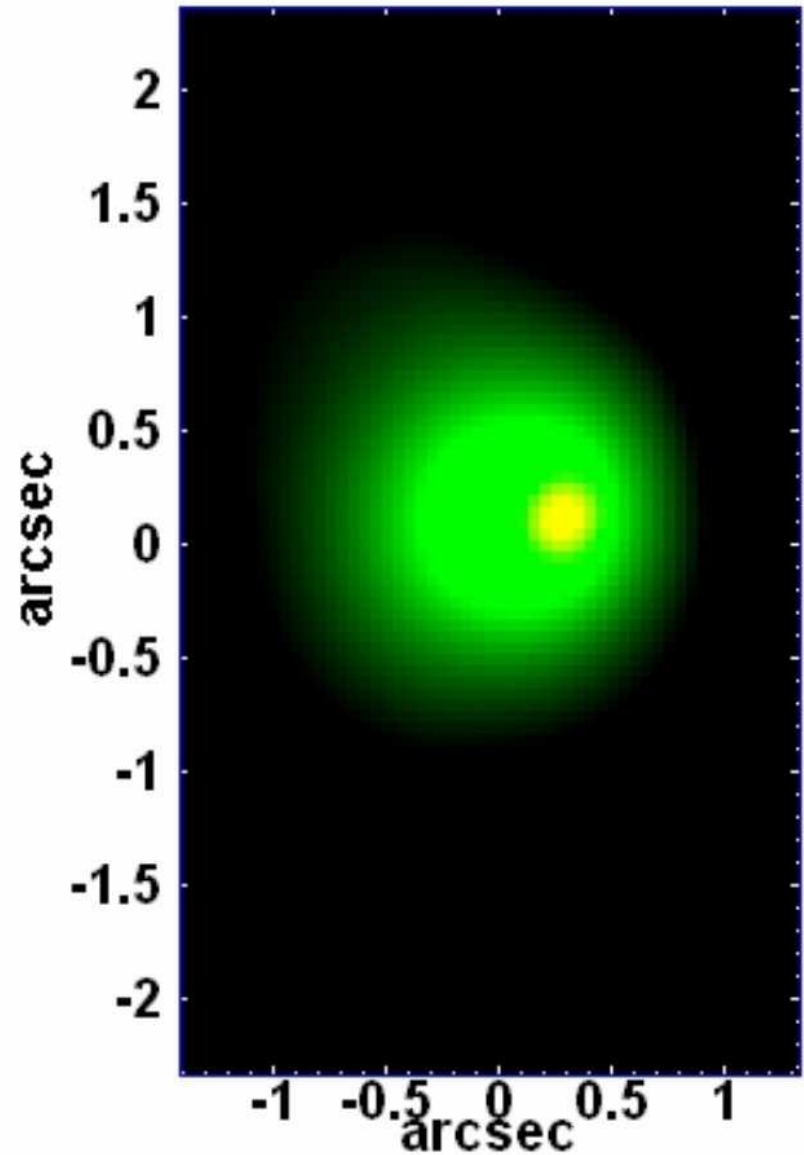


The “scree test”



Tomogram 1 (green) + 2 (yellow)

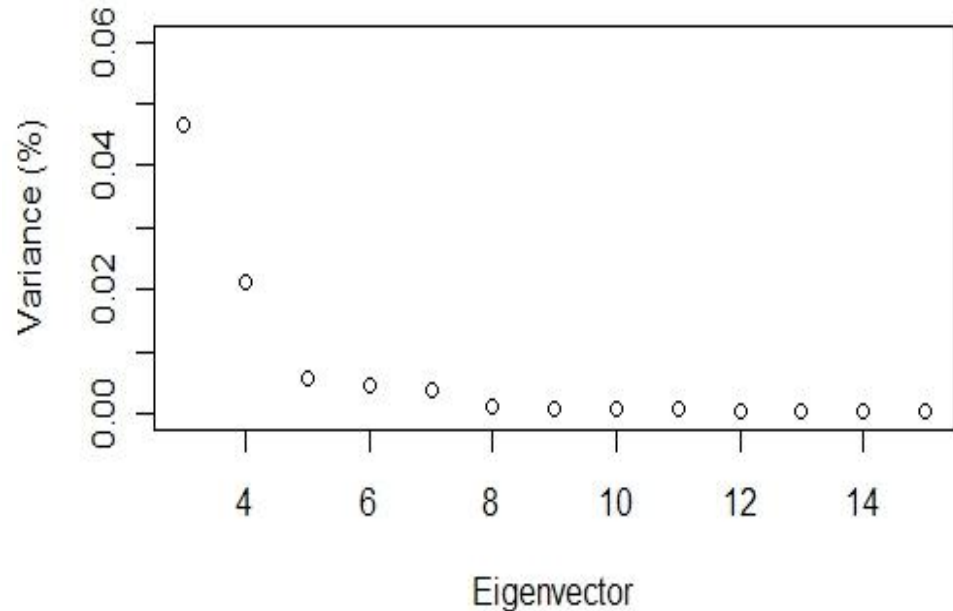
The stellar bulge and the BLR



NGC 7097: Eigenvalues

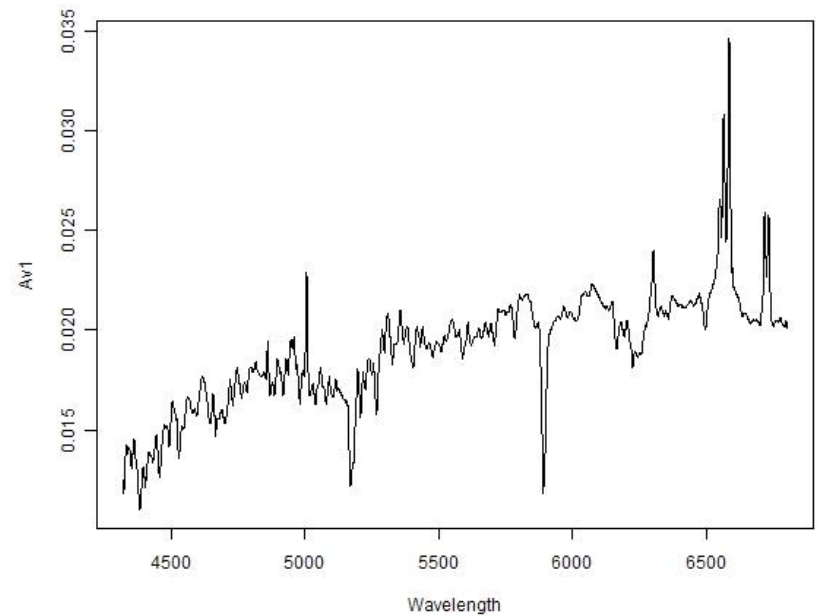
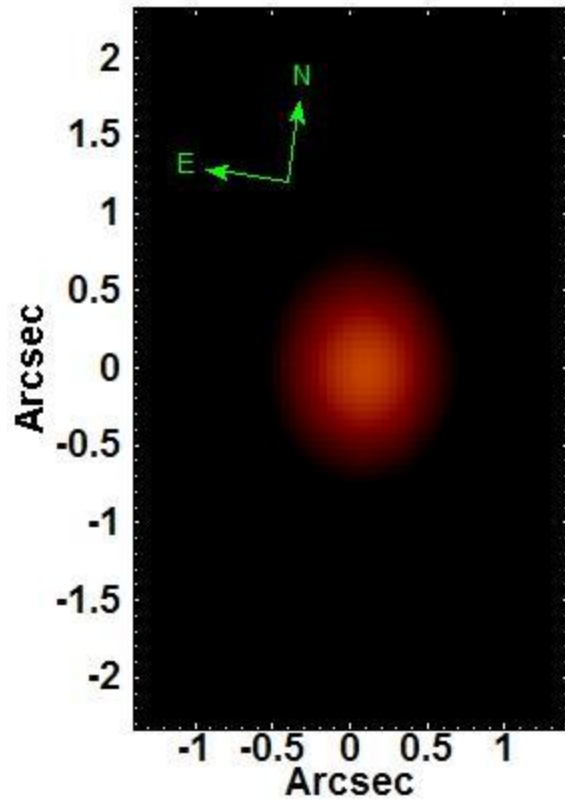
Eigenvector	Variance (%)
1	99.5256
2	0.3849
3	0.0464
4	0.0212
5	0.0059
6	0.0046
7	0.0038
8	0.0012
9	0.0009
10	0.0008

Scree test – Indicates that until Eigenvector 8 we have useful results



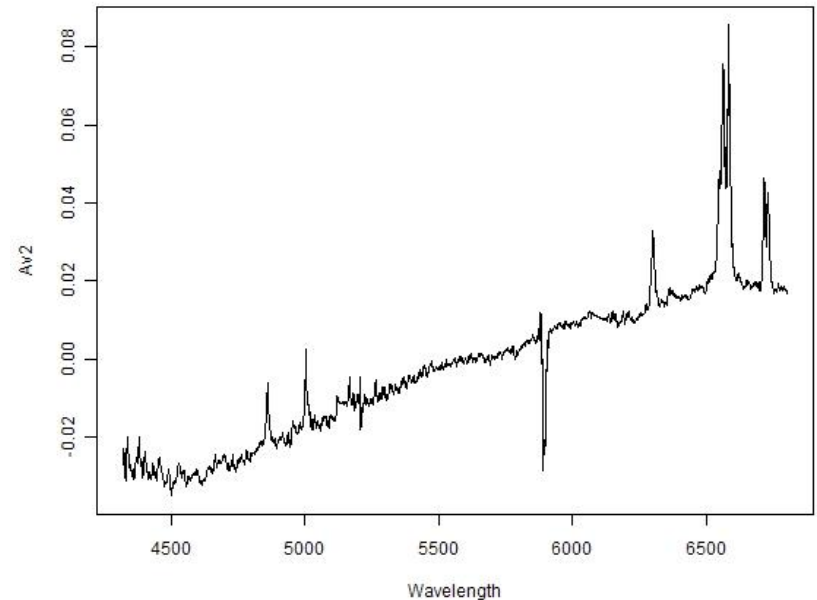
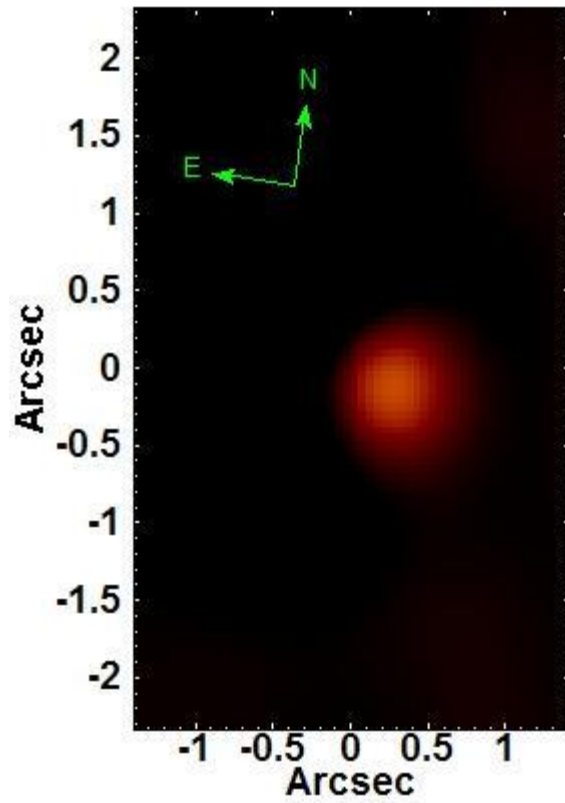
NGC 7097: Eigenvector 1 (99.52%):

Mainly the galaxy Bulge



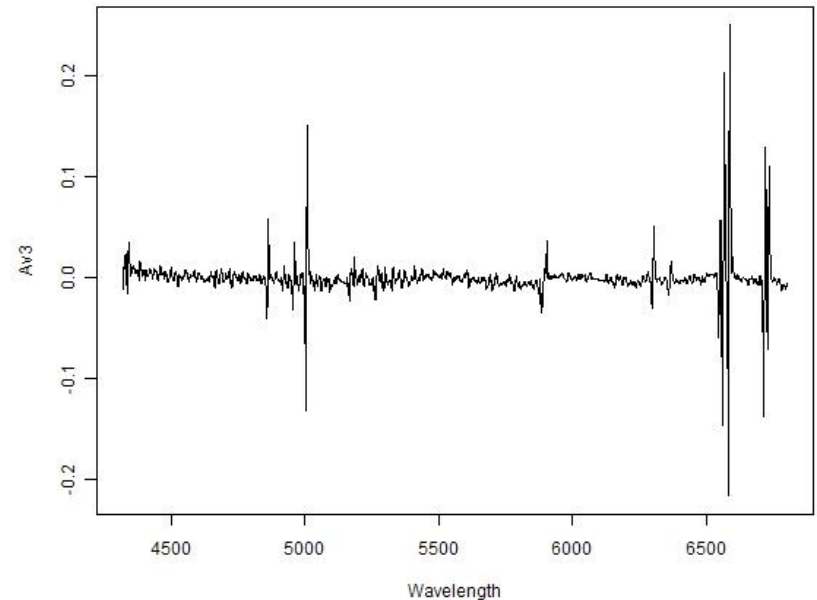
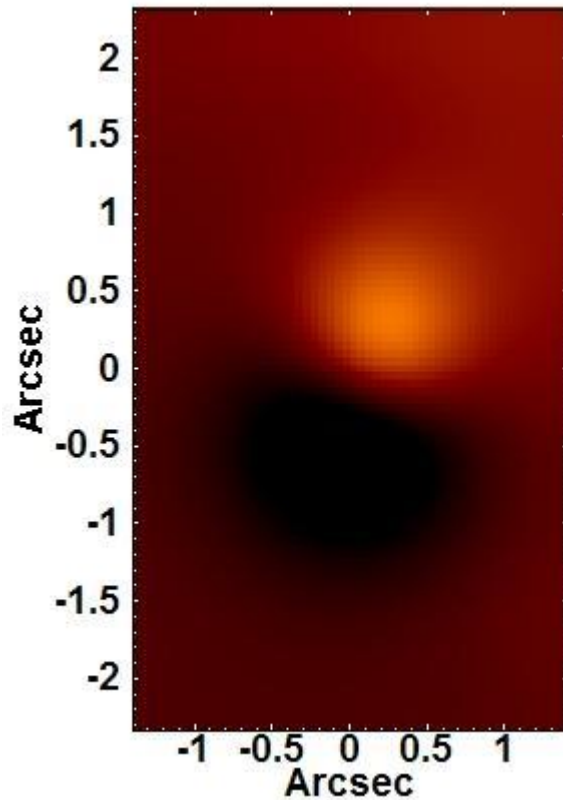
NGC 7097: Eigenvector 2 (0.38%):

LINER signature. Red component correlated with the emission line features. Interstellar gas (Na I) in the line of sight to the AGN.

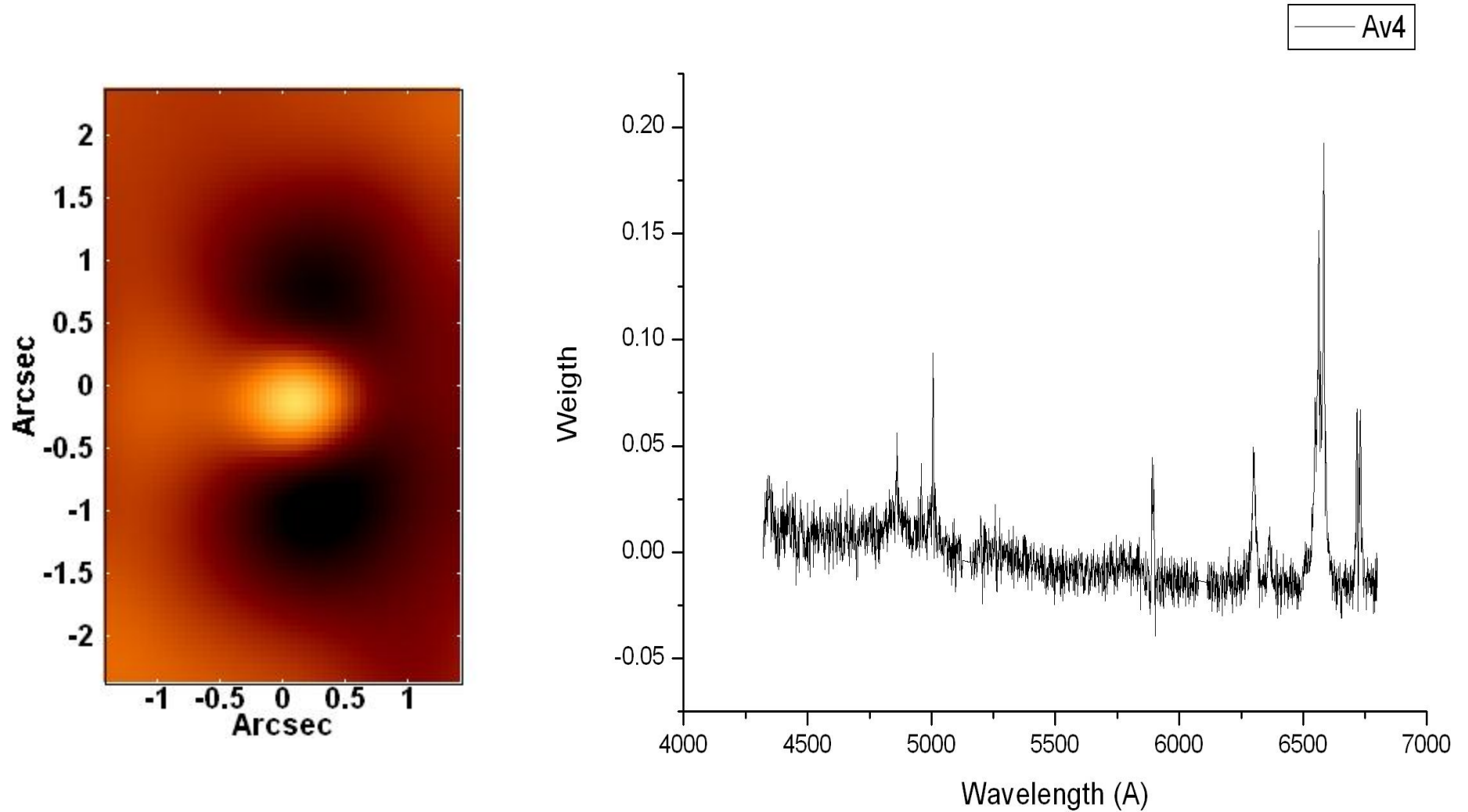


NGC 7097: Eigenvector 3 (0.046%):

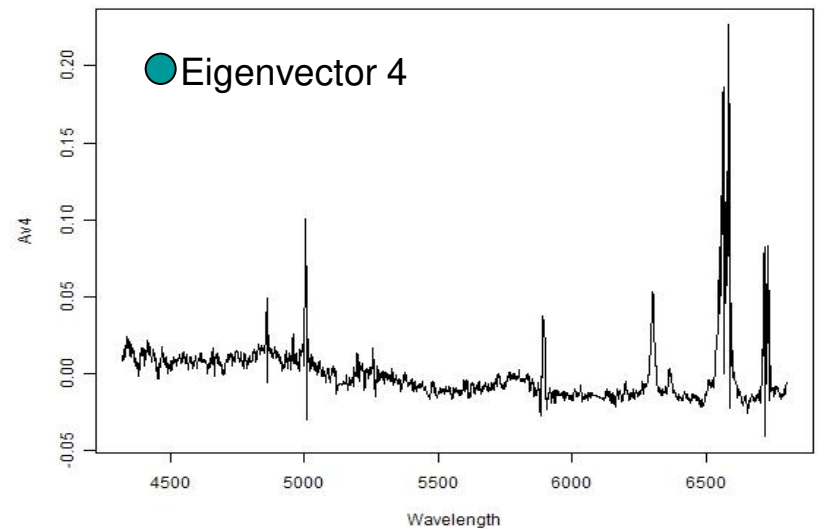
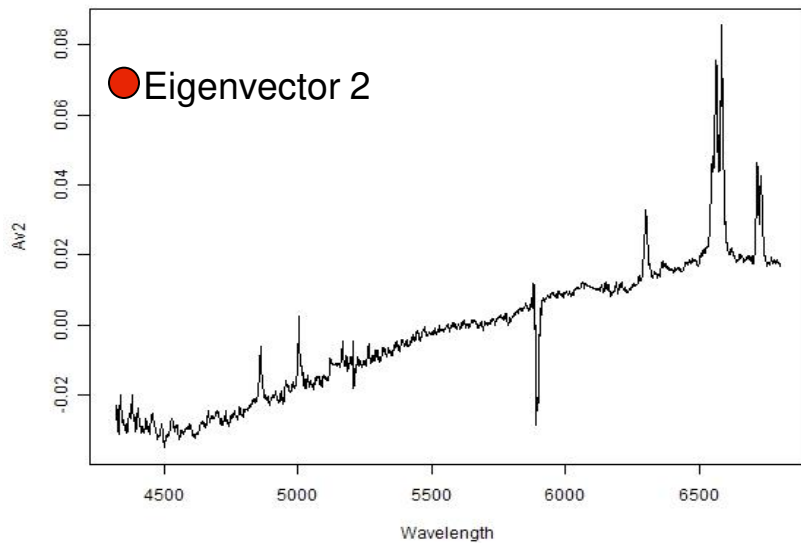
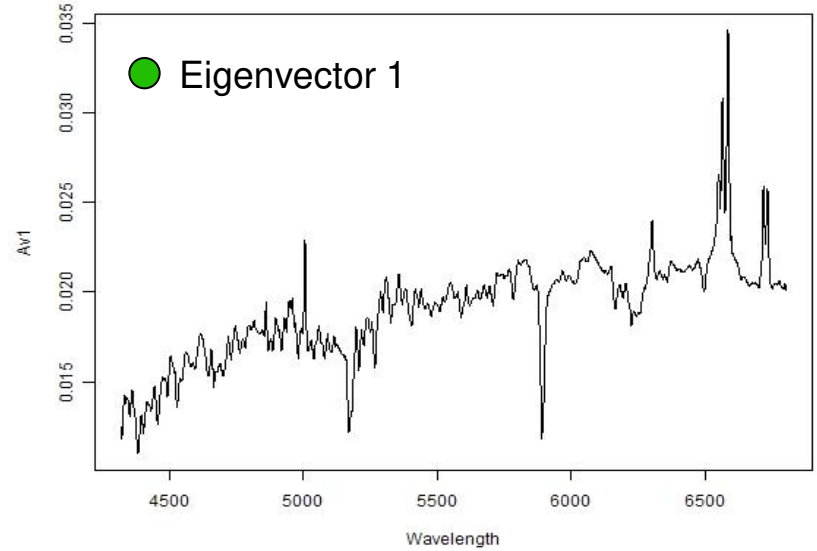
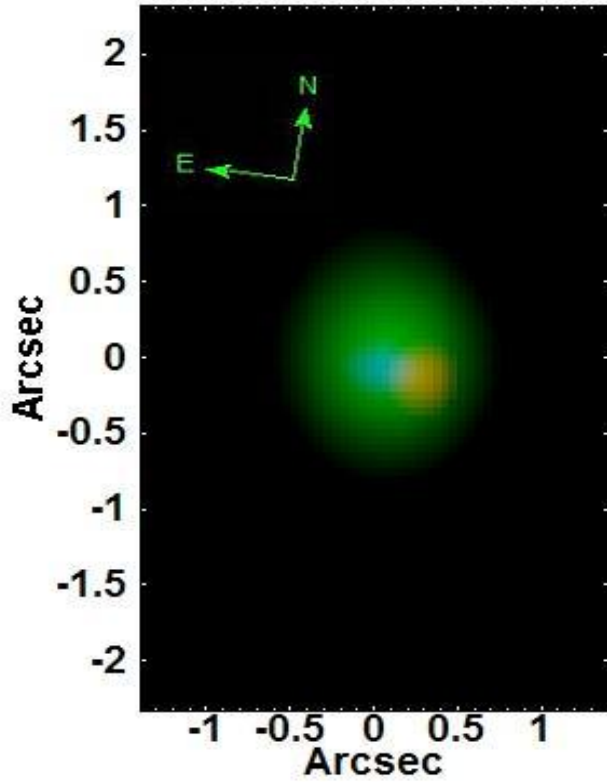
Positive correlation is the red component of the line features.
Negative correlation is the blue component of the line features.
Rotating disc



NGC 7097 – Av 4

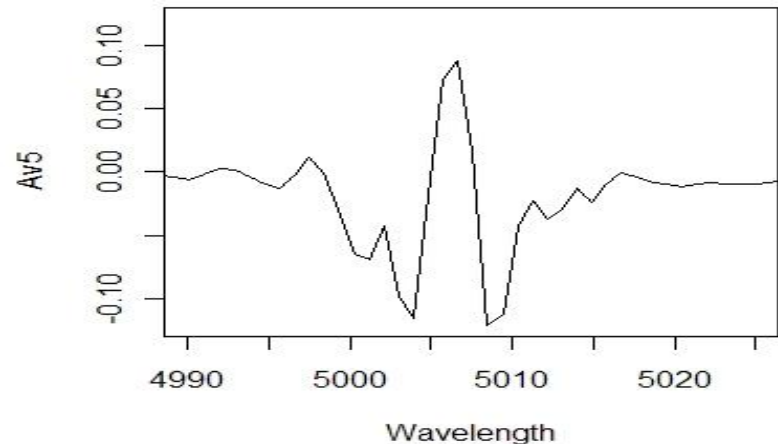
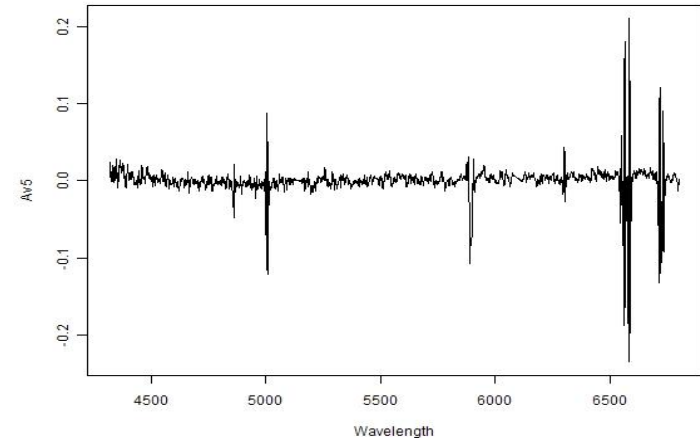
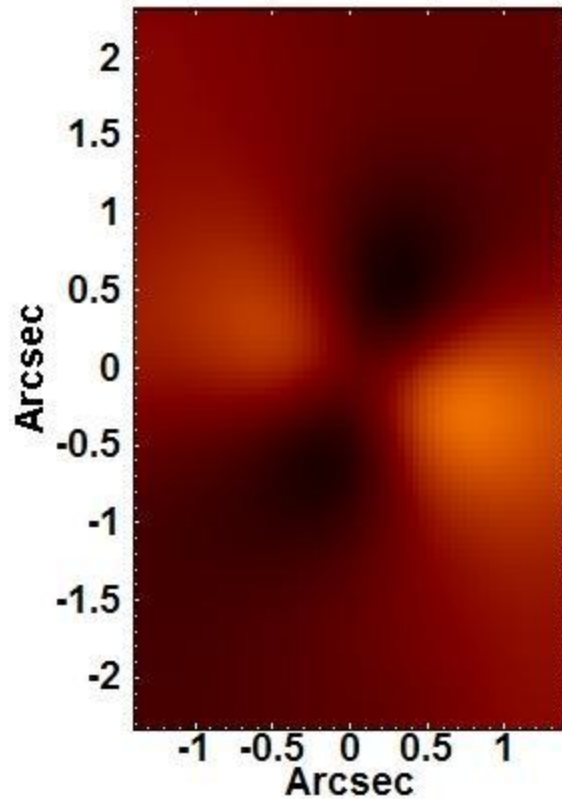


Combining Tomograms 1, 2 and 4

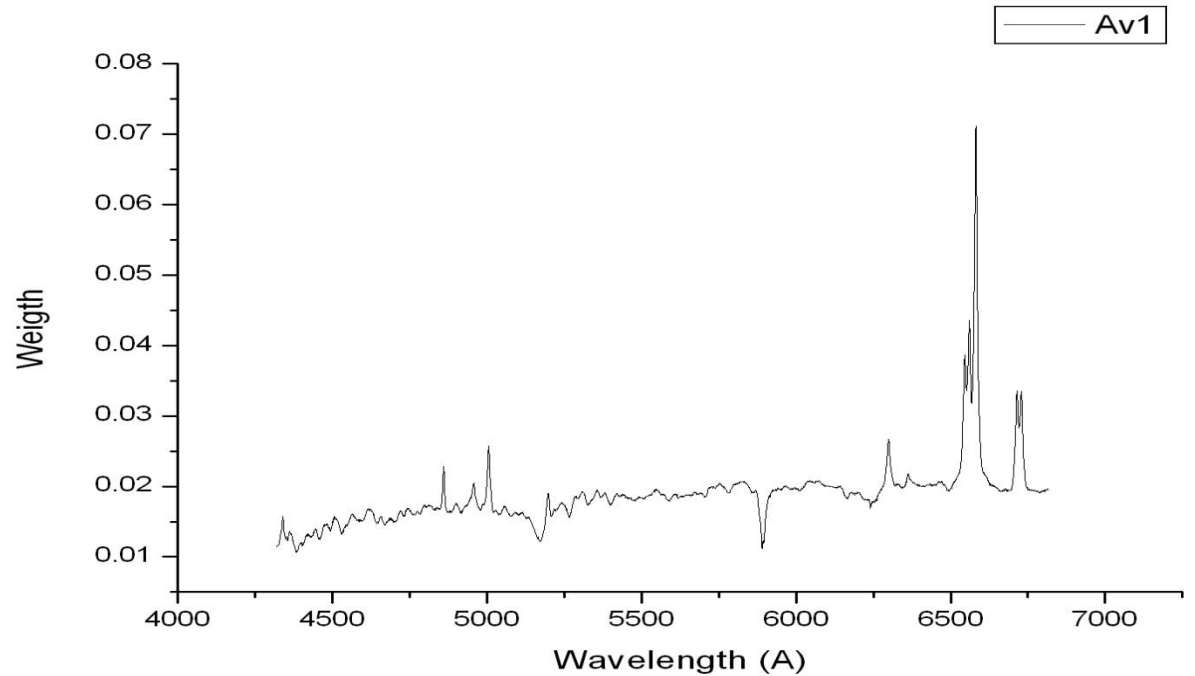
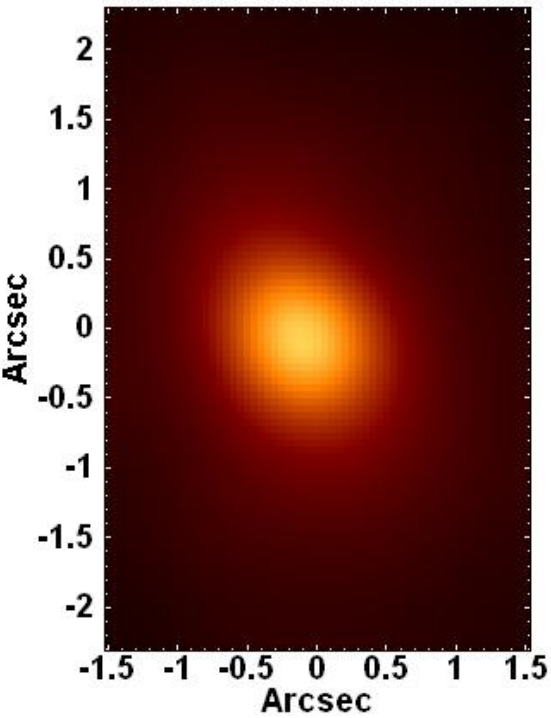


NGC 7097: Eigenvector 5 (0.0059%):

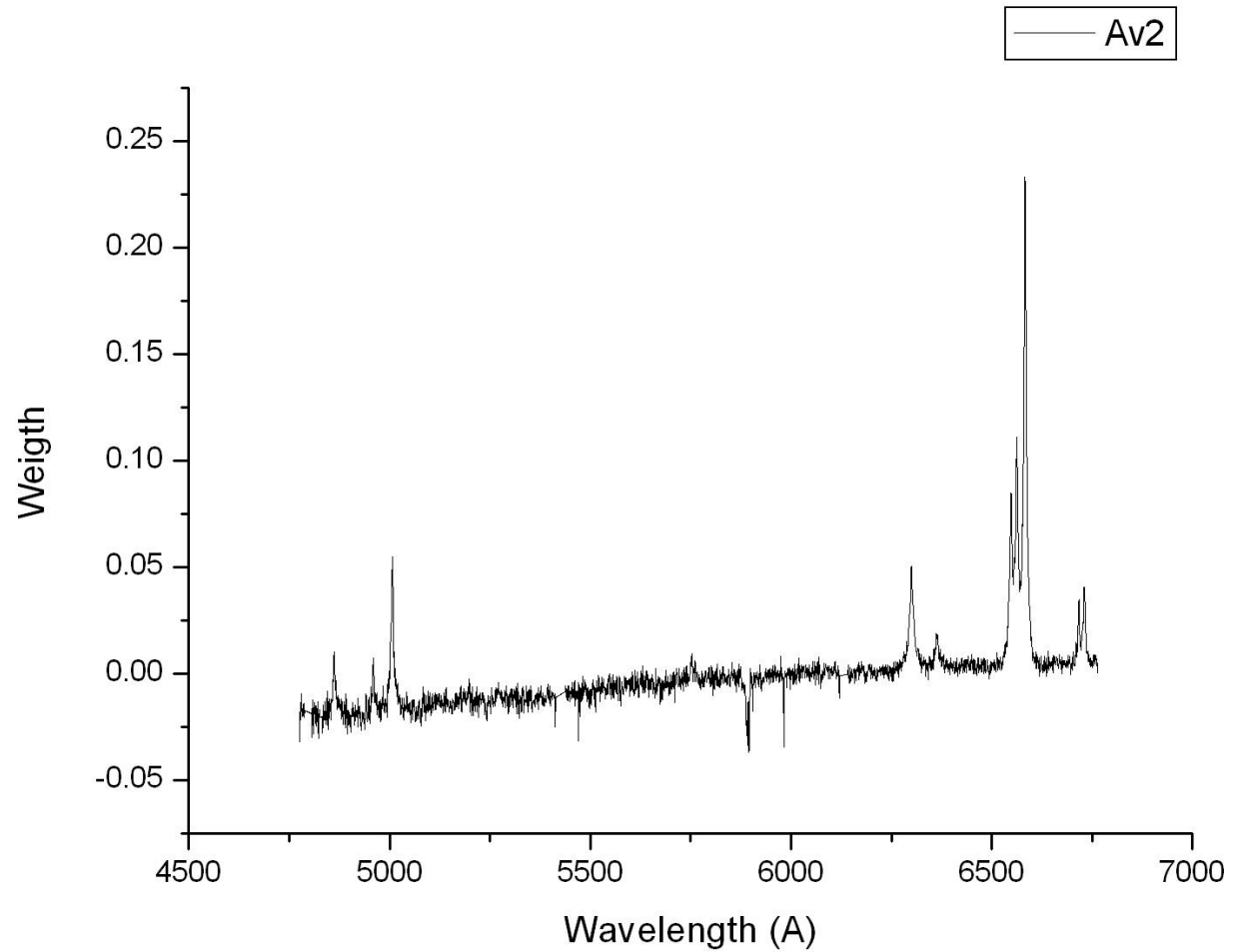
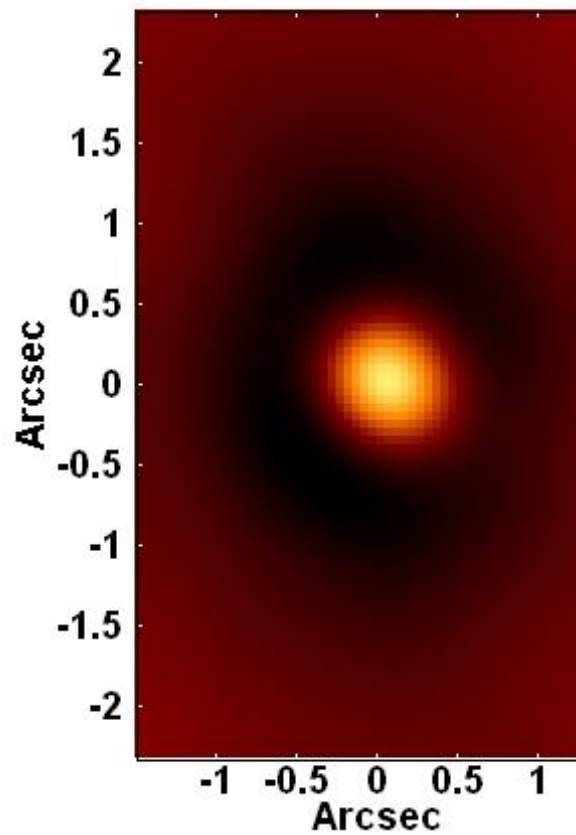
- The central component of the emission line features is anti-correlated with their blue and red components.
- Ionization cone (correlated) and rotating disc (anti-correlated).



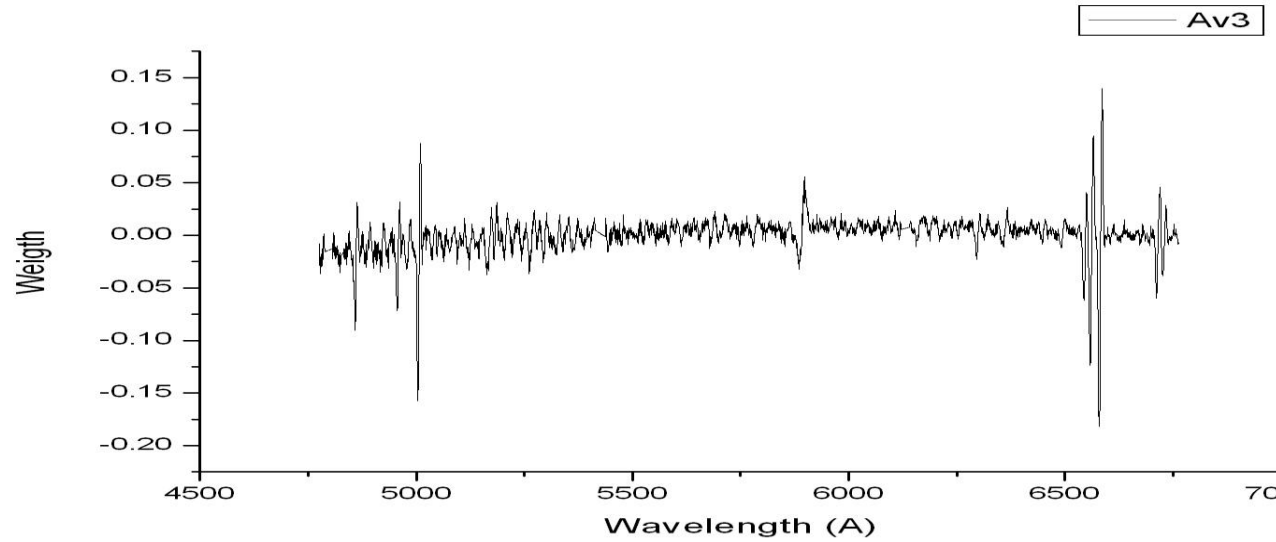
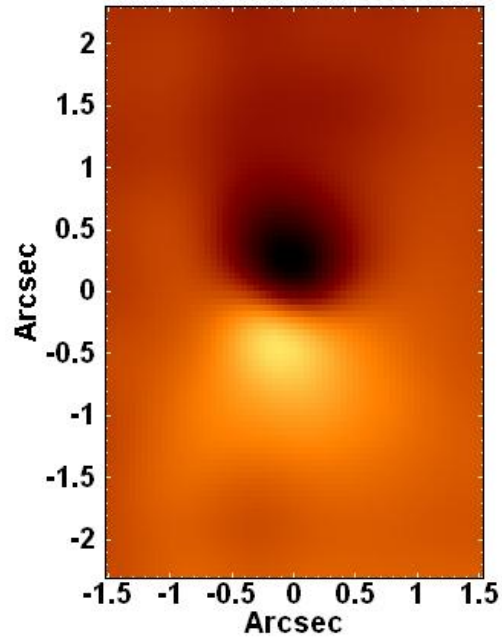
NGC 4546 (SB(s)0-::; 13 Mpc) – Av 1



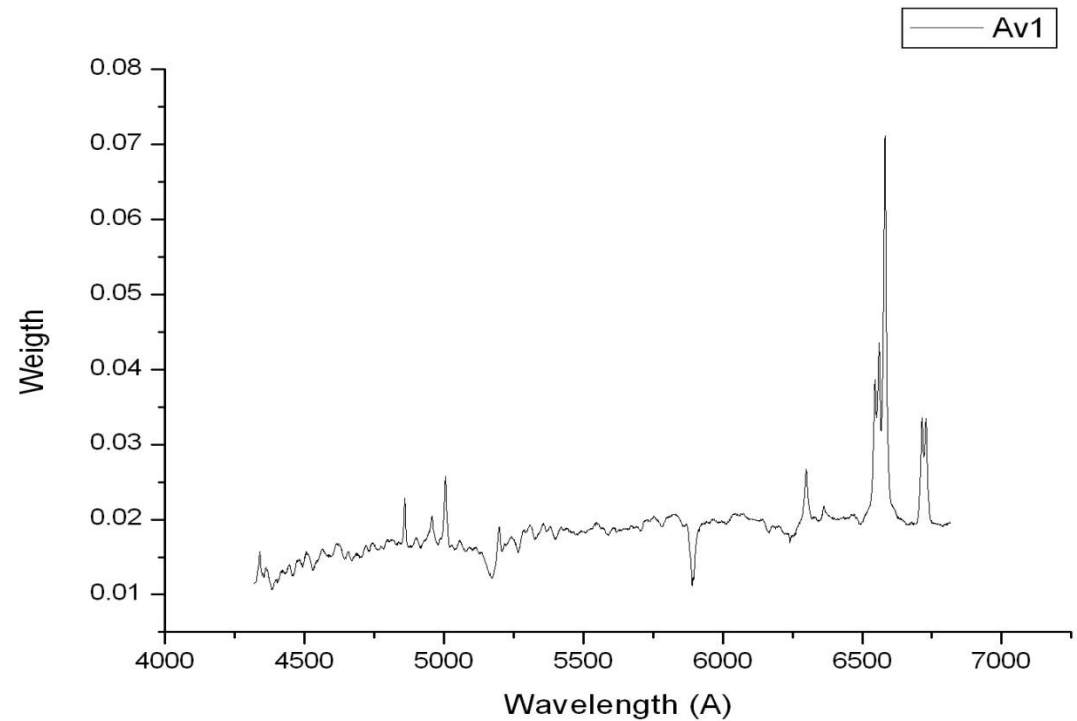
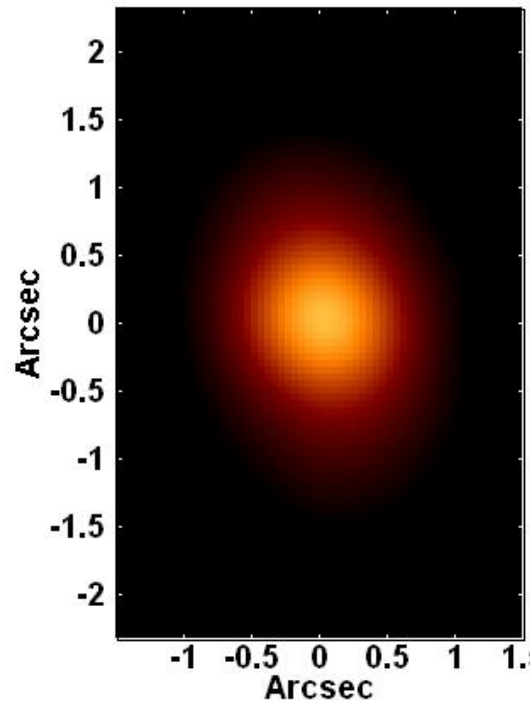
NGC 4546 – Av 2



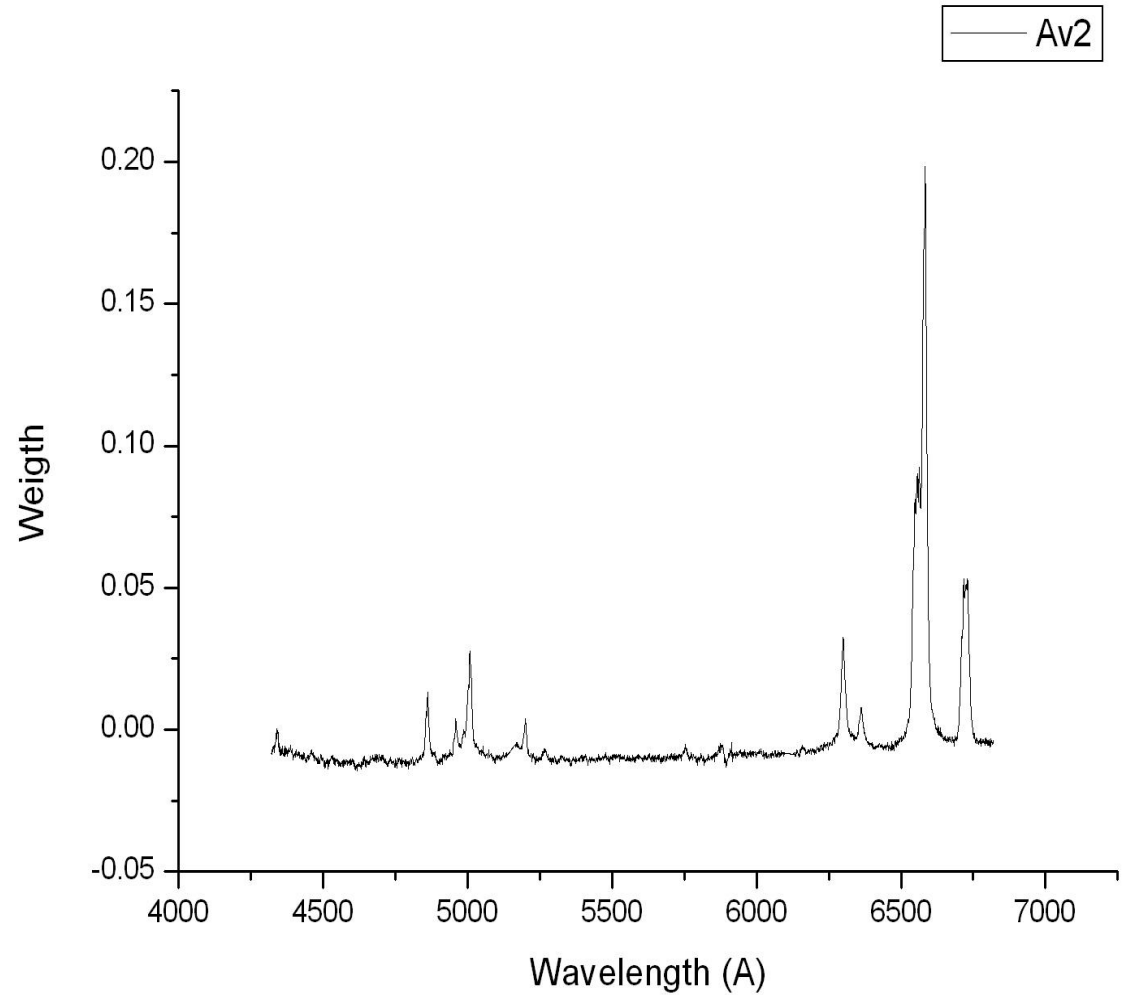
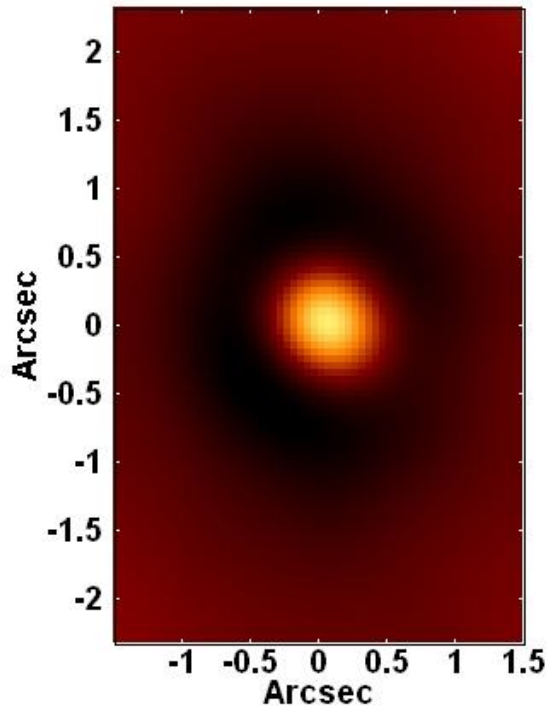
NGC 4546 – A_v 3



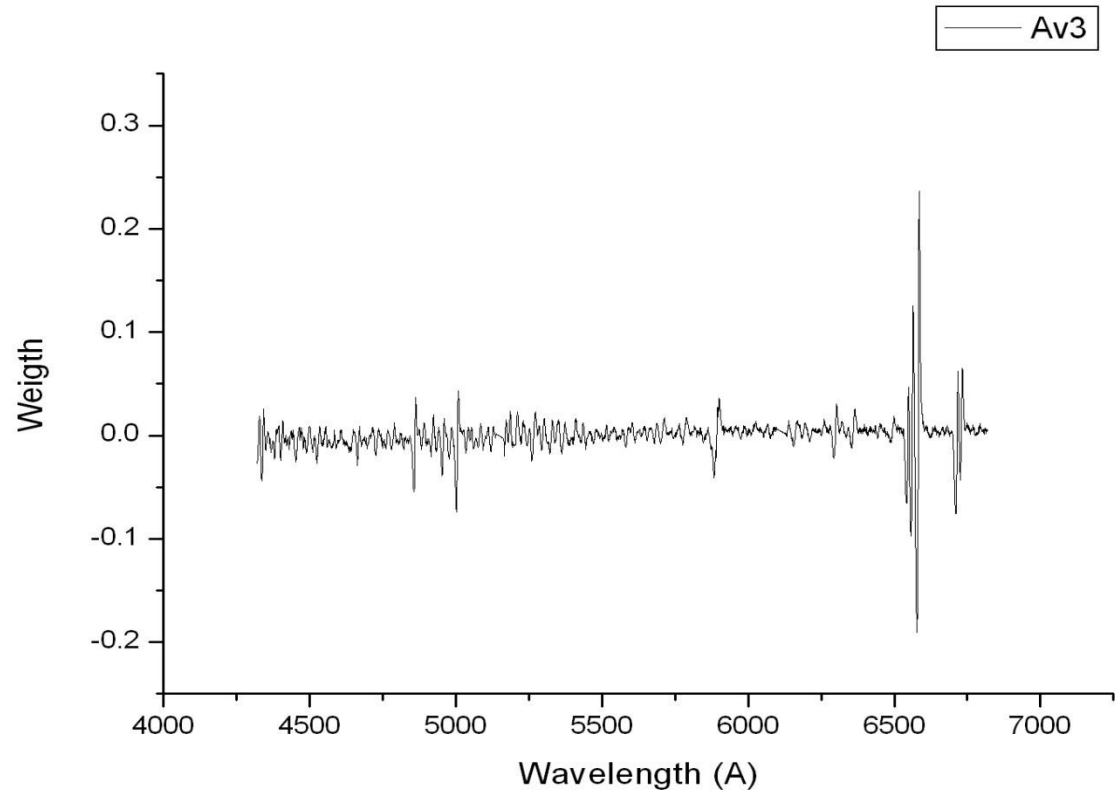
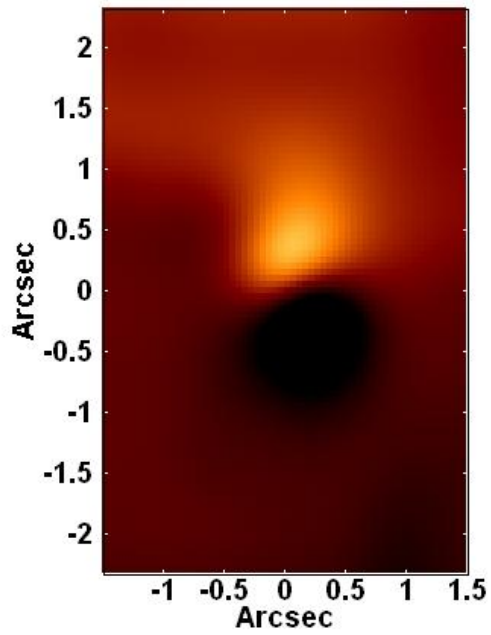
IC 1459 (cD; 18 Mpc)- Av 1

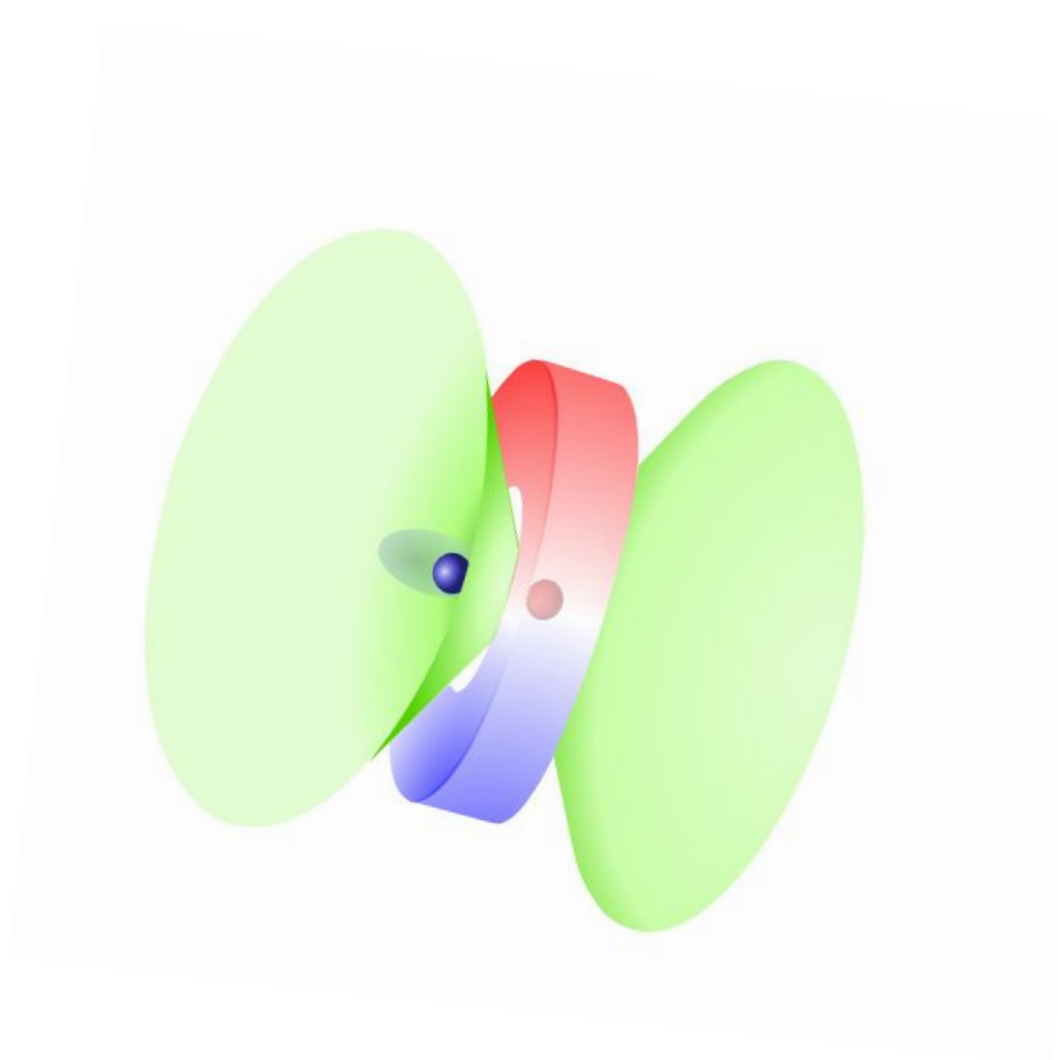
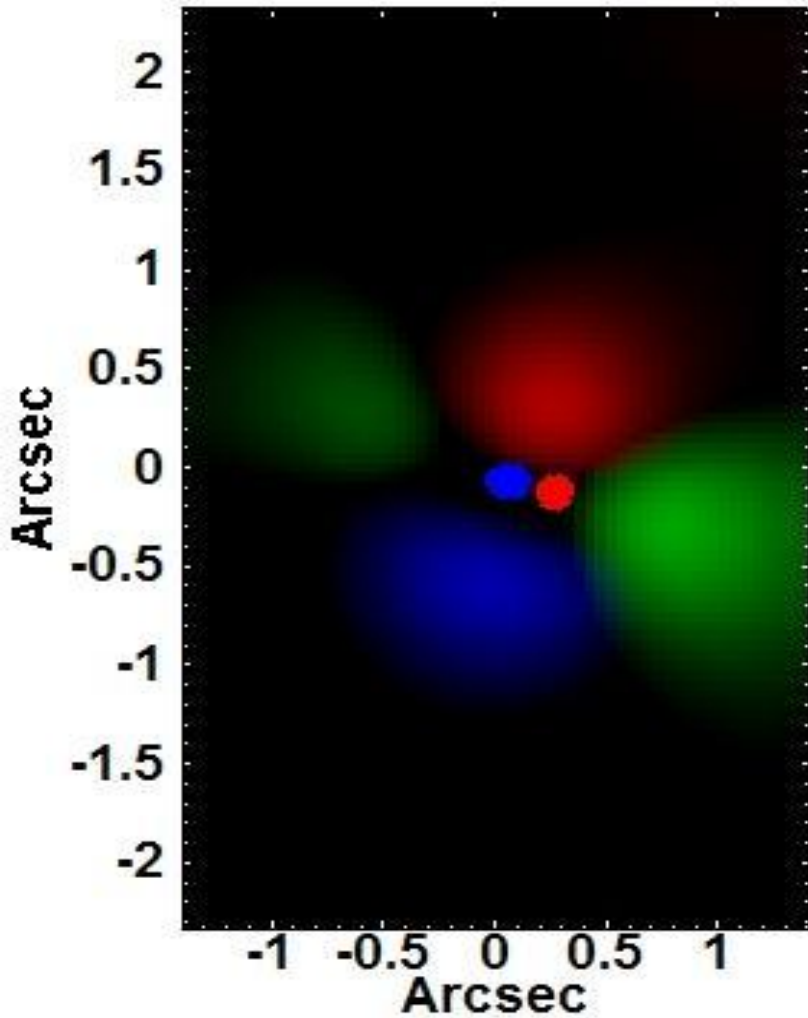


IC1459 – Av 2



IC1459 – Av 3

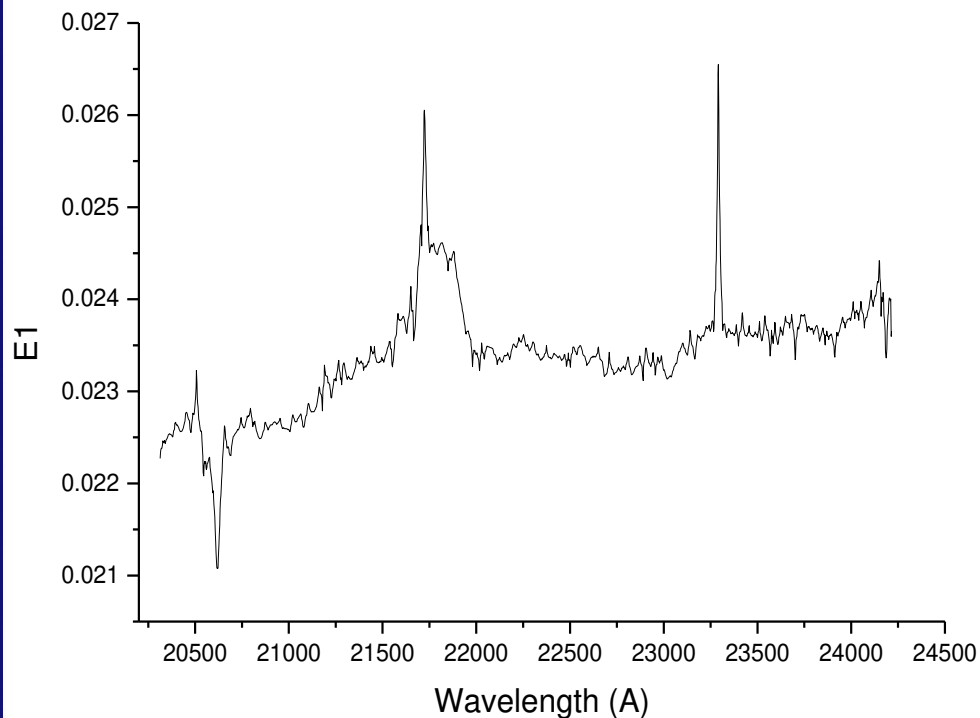
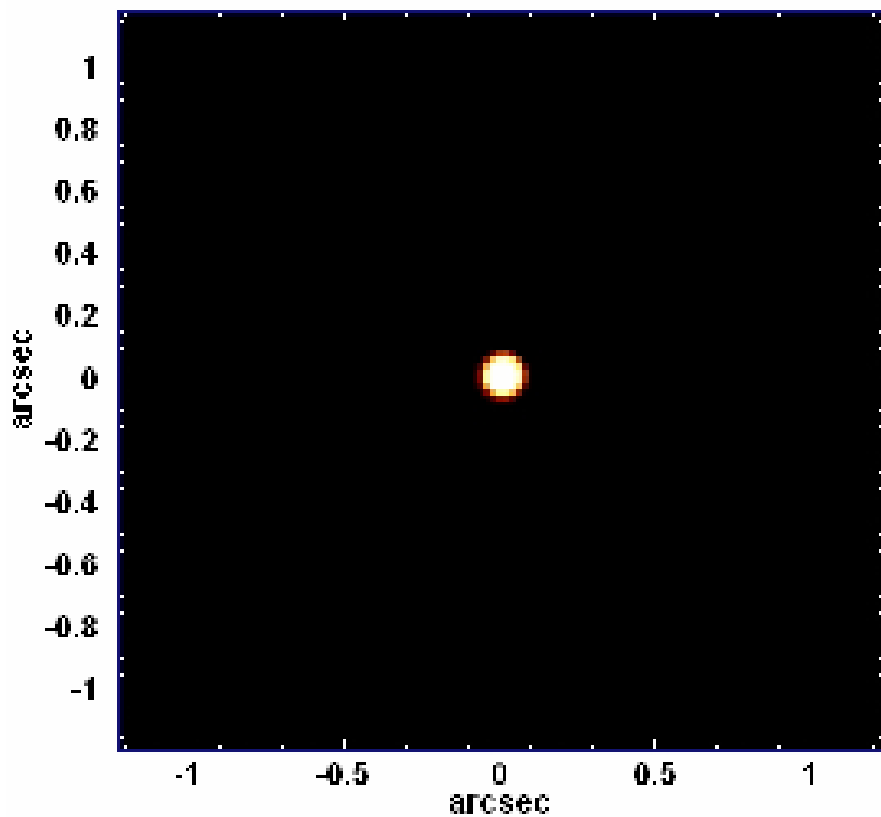




- NGC 7097 – GMOS IFU
- Ricci, Steiner & Menezes : The Astrophysical Journal Letters, 734, L10 (maio/2011).

NGC 4151 NIFS data (Storchi-Bergmann et al 2009)

- Re-sample the data to a pixel of 0.021"
- Butterworth filter in Fourier space to remove high spatial frequencies
- Tomogram 1 can be used as a reliable PSF of the AGN (point source)
- It can be used to deconvolve the data cube
- Strehl ratio before deconvolution = 0.05
- Strehl ratio after deconvolution = 0.15



Feature suppression and enhancement

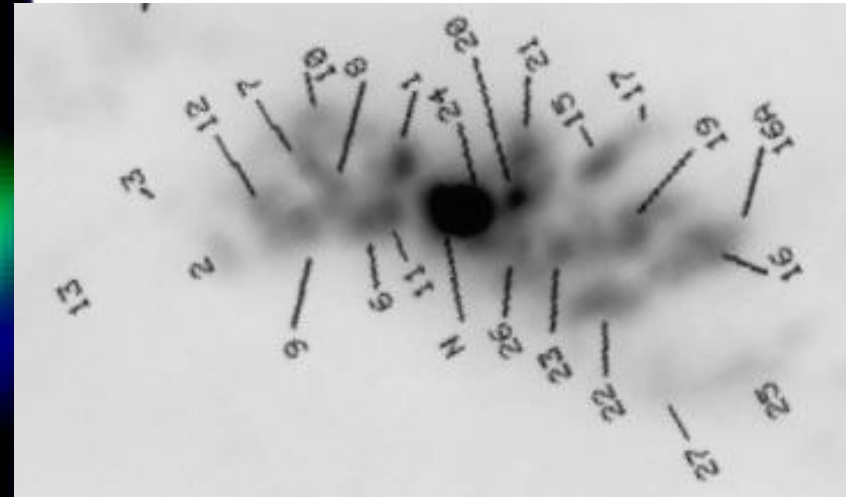
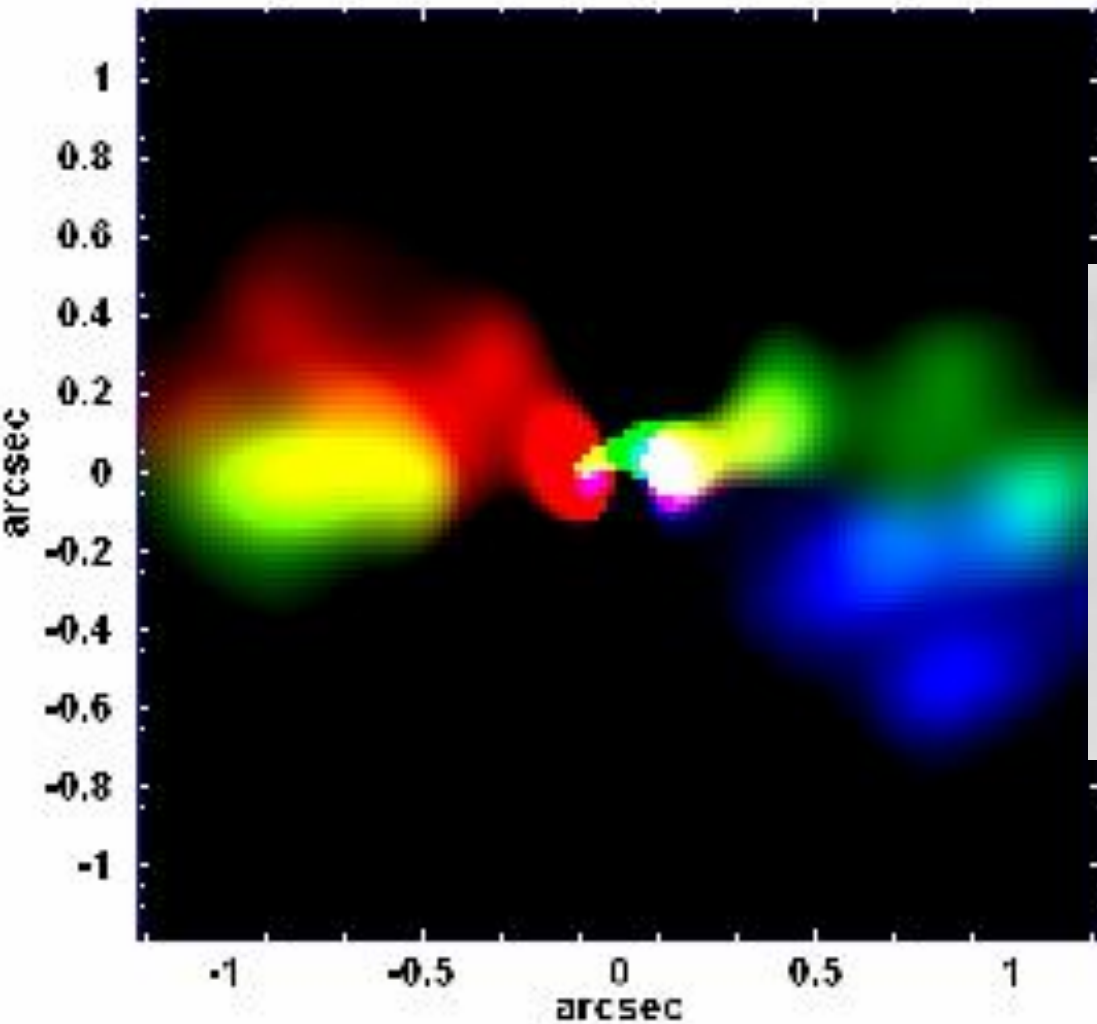
- Defining an object “A”: $\Gamma_k(A) = 1;0$
- Cube with enhanced feature: $I'_{ij\lambda}(A) = \sum_k [(I'_{ij\lambda}(k)) \cdot \Gamma_k(A)]$
- Or directly $I'_{\beta\lambda}(A) = T_{\beta k} \cdot [(E_{\lambda k})_{\Gamma}]^T$

NGC 4151-NIFS observations (Storchi-Bergmann et al 2009)

“software coronagraph”

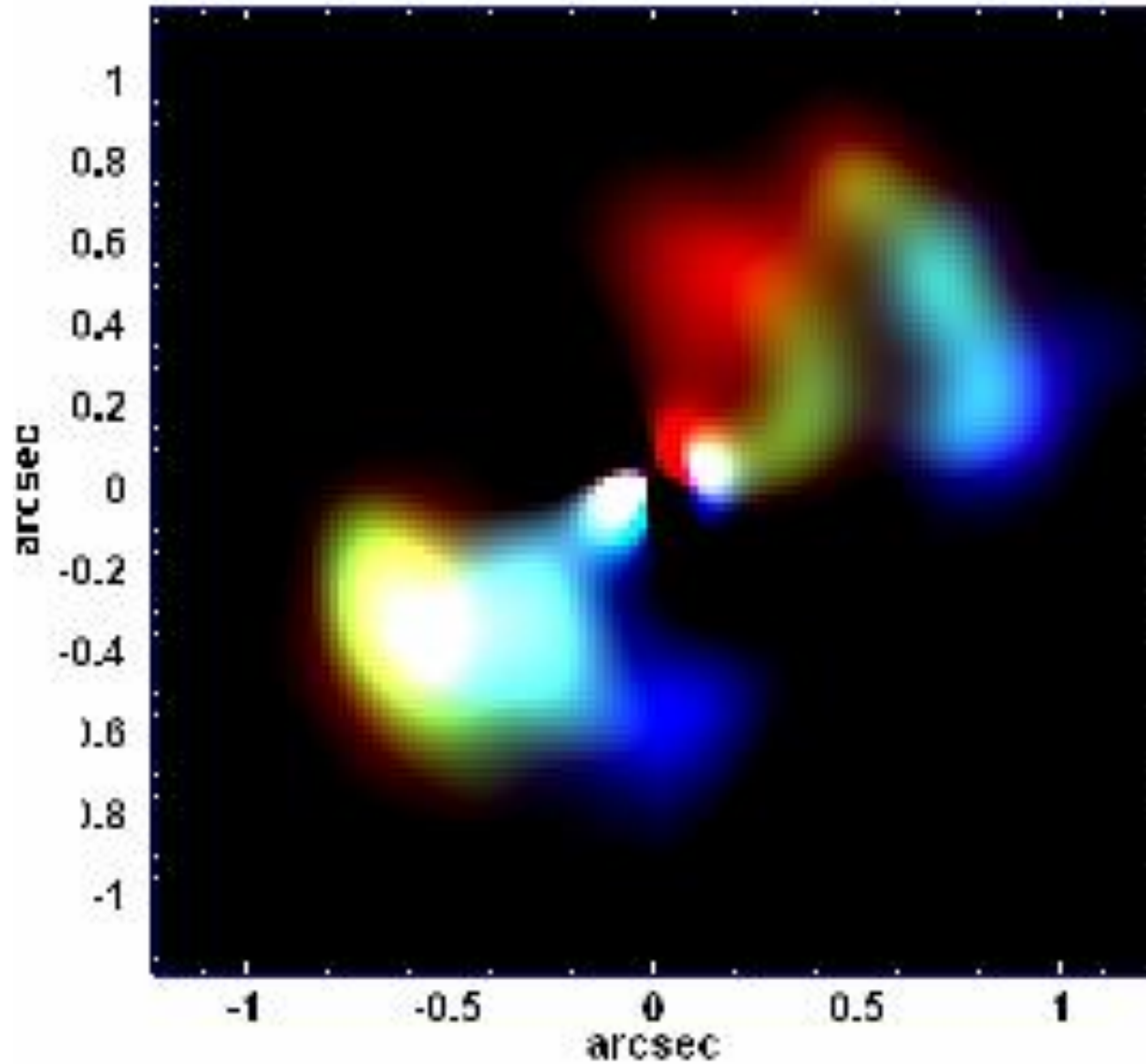
Br γ after switching-off PC1+PC2

HST observations



NGC 4151 – NIFS observations (Storchi-Bergmann et al 2009)

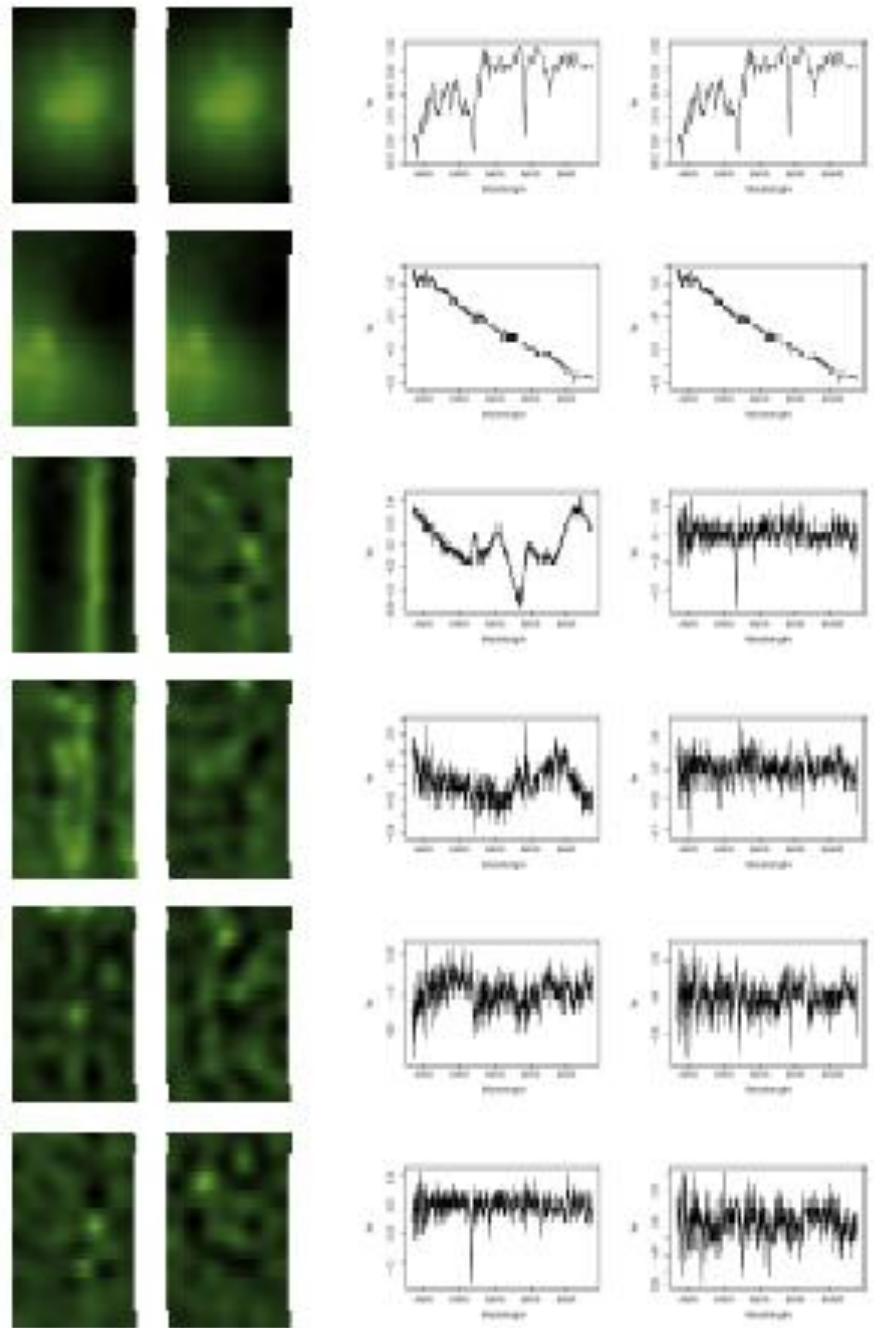
H2 lines – after switching-off PC1+PC2



NGC 1399 - GMOS IFU

Identifying and removing
instrument “fingerprints”

(see poster: Roberto Menezes)



Steiner et al (2009) MNRAS 395, 64

Relevant software can be found at
WWW.astro.iag.usp.br/~pcatomography

Thank you

The SOAR data cube toolbox

- Butterworth filtering in Fourier space
- Anscombe transform
- Differential atmospheric refraction correction
- Identification and removal of fingerprints
- Richardson-Lucy deconvolution
- PCA Tomography

- *IDL (+Python)*

T1 T2 T3 T4 T5 T6 T7 T8.....

W0



W1



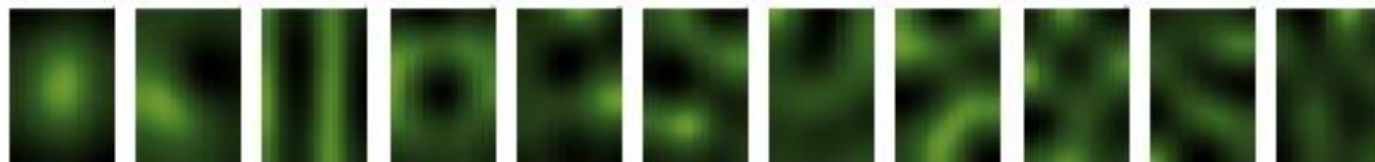
W2



W3



W4



WC



E1 E2 E3 E4 E5 E6

W0



W1



W2



W3



W4



WC



