PCA Tomography: how to extract information from data cubes

João Steiner

IAG - Universidade de São Paulo

+R. B. Menezes, T. V. Ricci & A.S. Oliveira

SOAR Telescopes data cubes

- SIFS – SOAR Integral Field Spectrometer: IFU with 1300 fibers (See: Antonio Cesar de Oliveira)

- BTFi – Brazilian Tunable Filter imager: Fabry-Perot + system of narrow filters (see: Claudia Mendes de Oliveira)

Data cubes

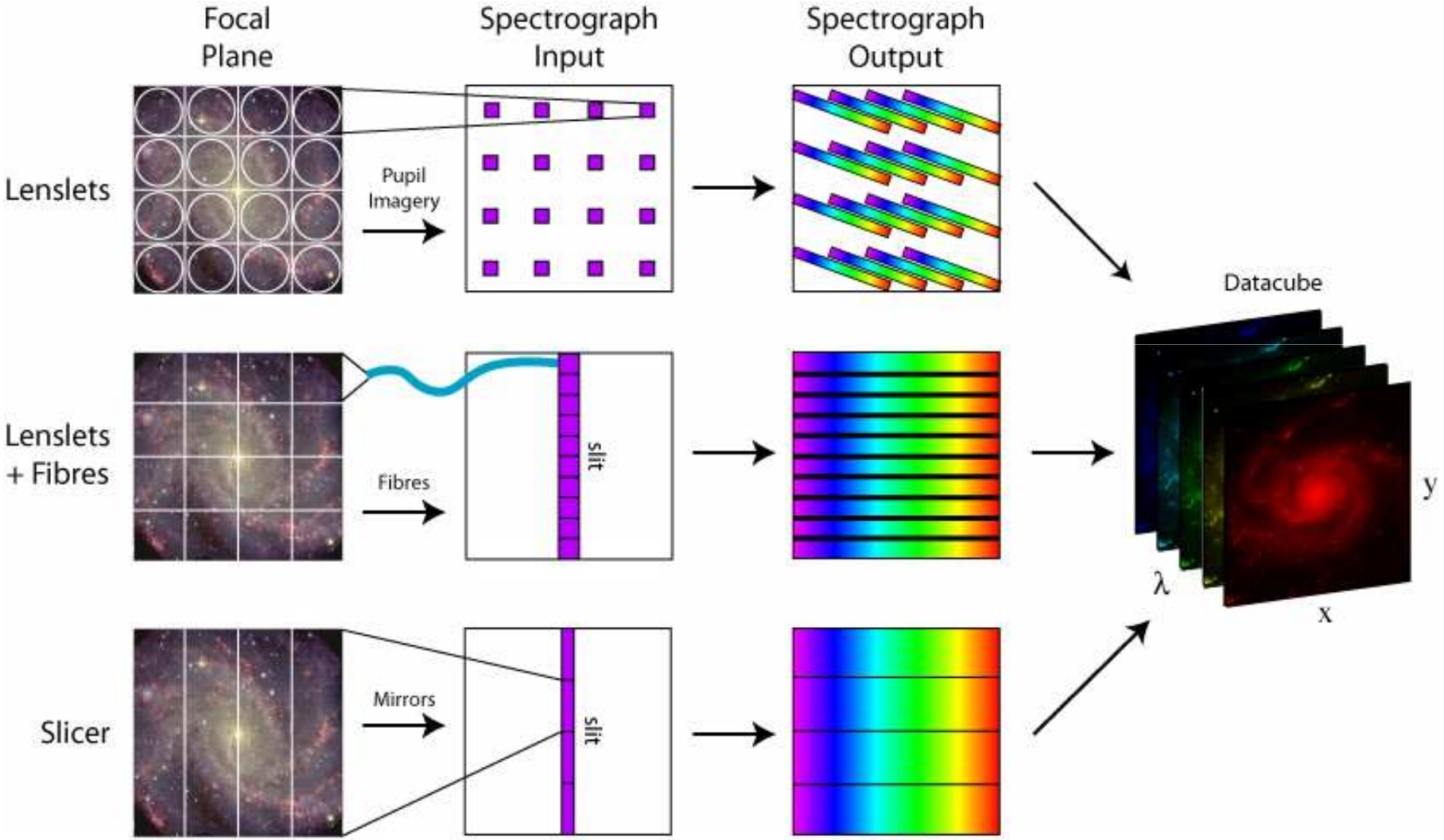
Traditional astronomical data:

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

Integral Field Unit (IFU) spectroscopy:

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

IFU spectrographs



Data Cubes

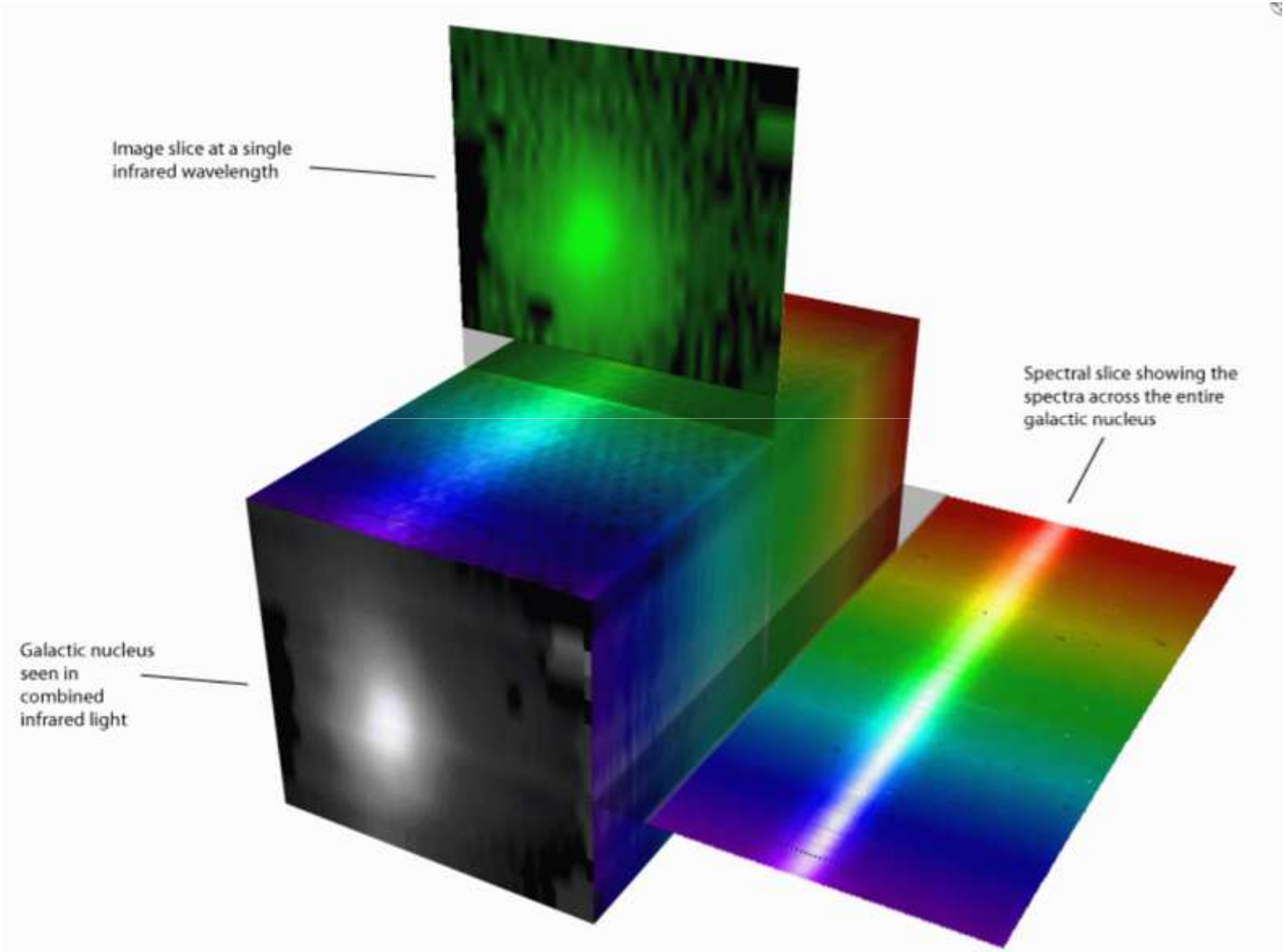


Image slice at a single infrared wavelength
Spectral slice showing the spectra across the entire galactic nucleus
Galactic nucleus seen in combined infrared light

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 = $\nu_x\mu$ spatial pixels and m spectral pixels.
- The mean intensity $$Q_\lambda = \frac{1}{n} \cdot \sum_{i=1}^{\mu} \sum_{j=1}^{\nu} \left(I_{ij\lambda} \right)_O$$
- The intensity adjusted $$I_{ij\lambda} = \left(I_{ij\lambda} \right)_O - Q_\lambda$$
- The data cube $I_{ij\lambda}$ has to be transformed into a matrix, $I_{\beta\lambda}$

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

PCA — Principal Component Analysis

- Covariance matrix $C_{\mathrm{cov}} = \frac{\left[I_{\beta\lambda}\right]^T \cdot I_{\beta\lambda}}{n-1}$
- Properties $C_{\mathrm{cov}} = \left[C_{\mathrm{cov}}\right]^T$
- PCA transformation: $T_{\beta k} = I_{\beta\lambda} \cdot E_{\lambda k}$
- D must be diagonal: $D_{\mathrm{cov}} = \frac{\left[T_{\beta k}\right]^T \cdot T_{\beta k}}{n-1}$

The diagonal elements of D are the eigen-values Λ_k

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.

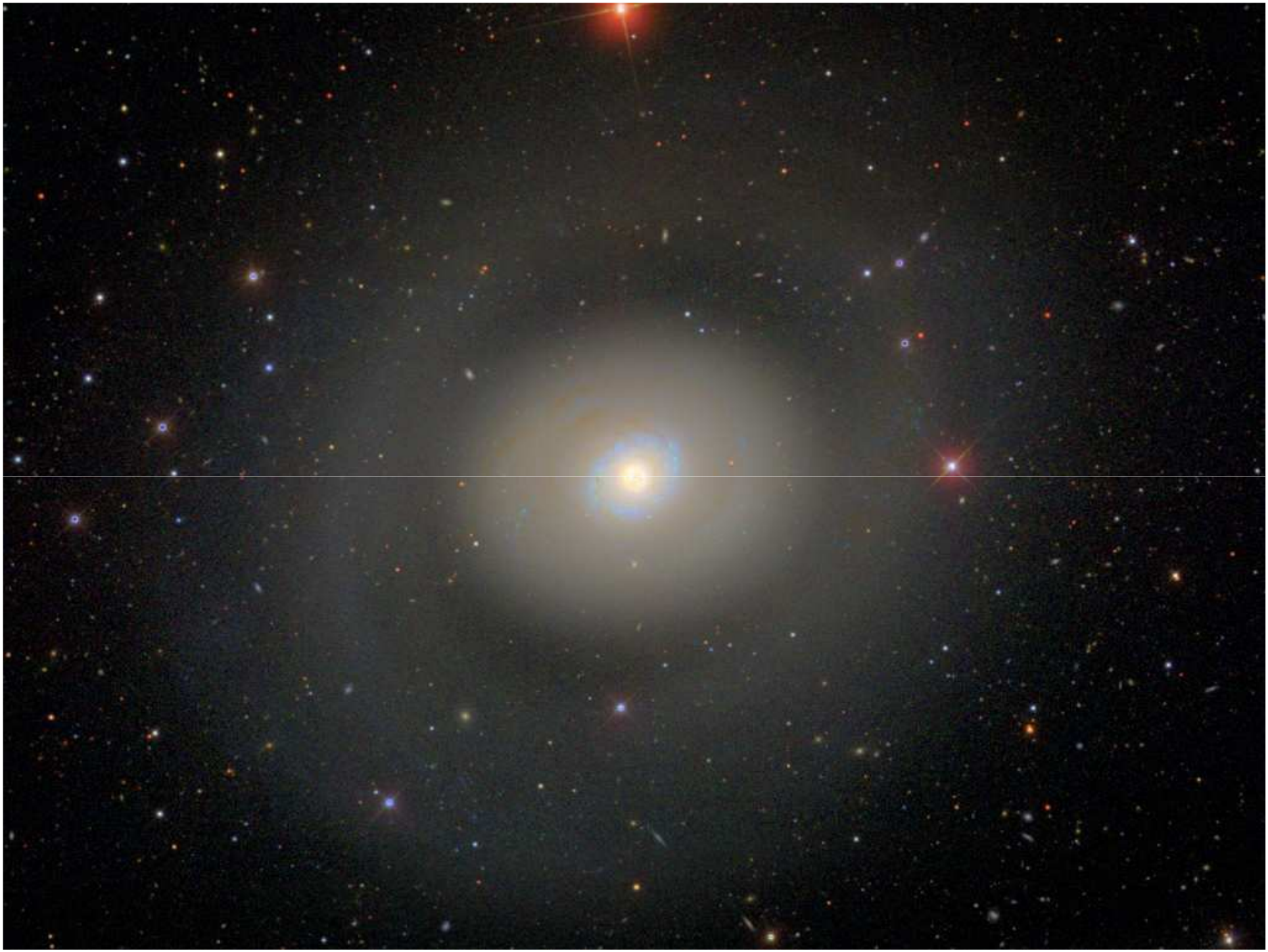
$$I'_{\beta\lambda}(\leq r) = T_{\beta k}(\leq r) \cdot \left[E_{\lambda k}(\leq r)\right]^T$$

- Callibrating flux back

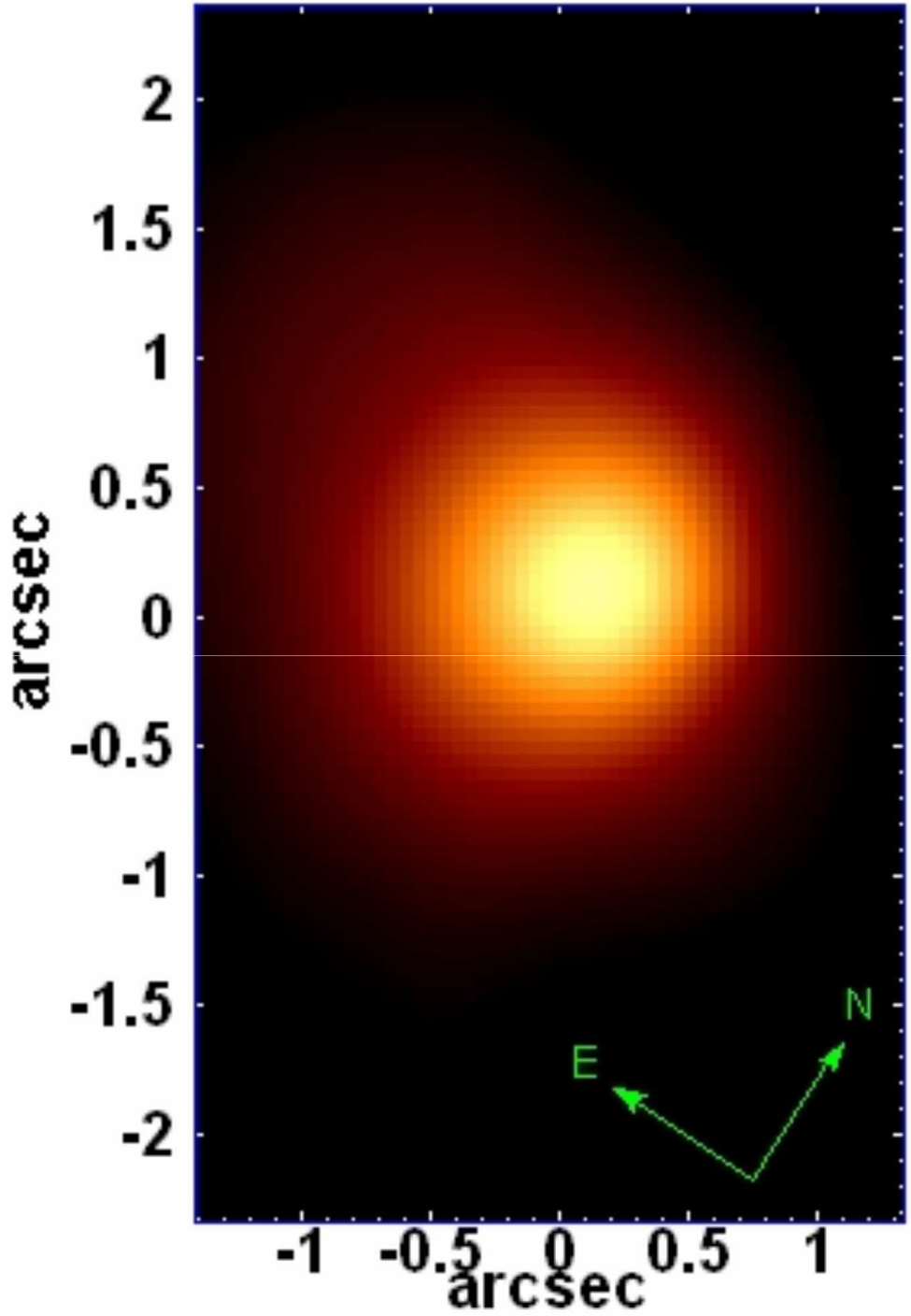
$$\left(I'_{ij\lambda}(\leq r)\right)_O = I'_{ij\lambda}(\leq r) + Q_\lambda$$

- Noise may be evaluated as

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

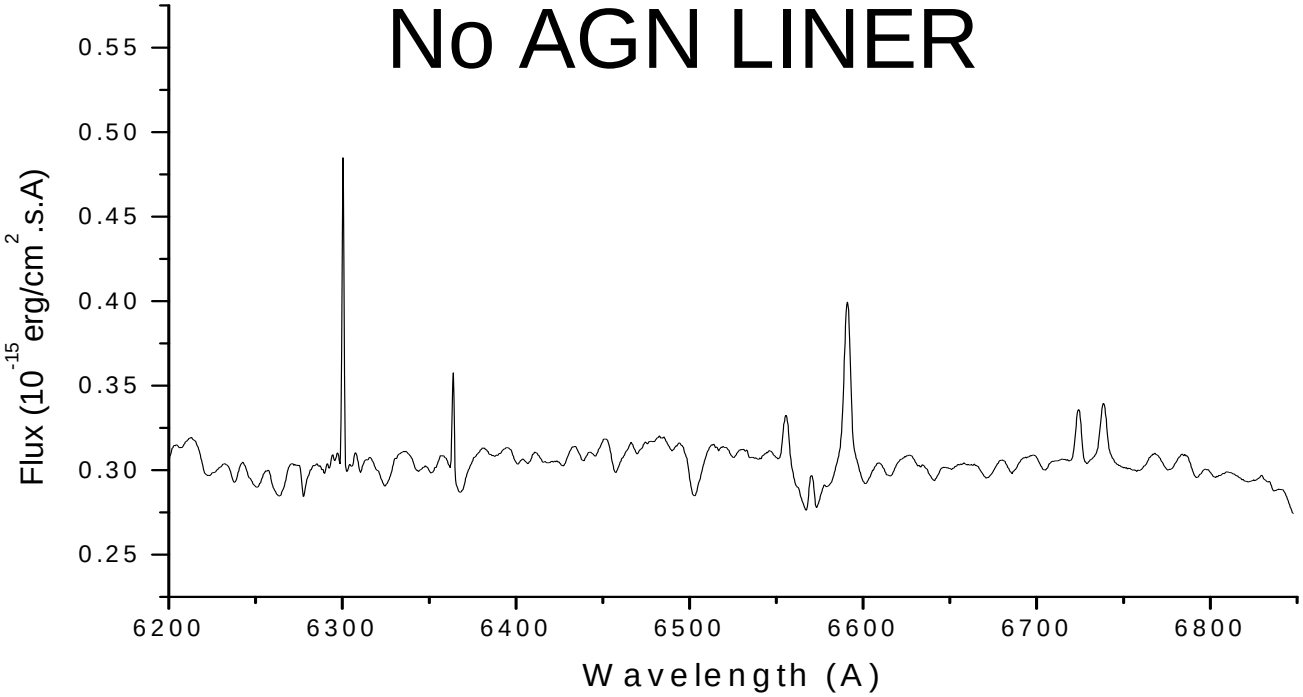


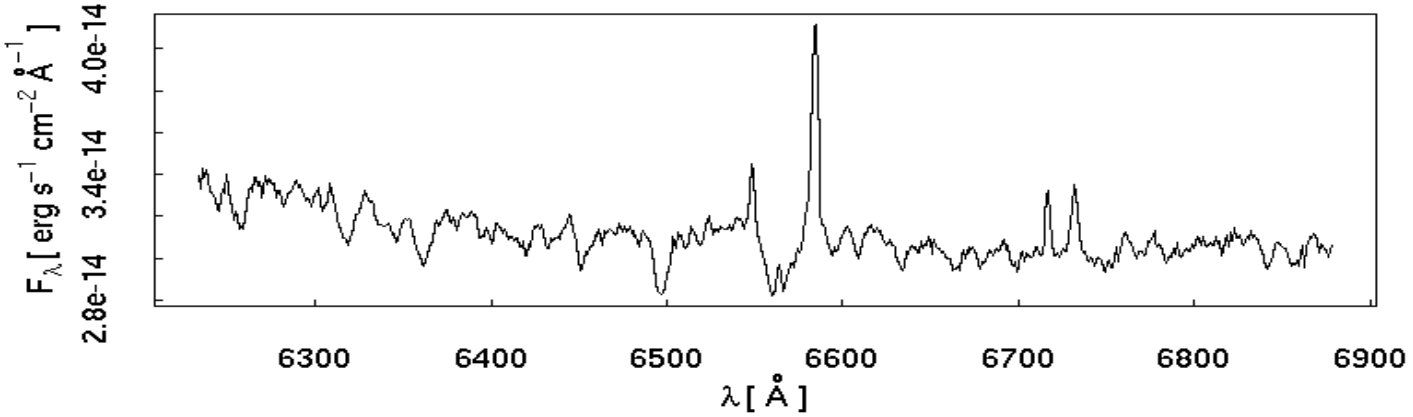
NGC 4736 (5 Mpc)
GMOS – IFU
Steiner et al 2009, MNRAS



NGC 4736 – Gemini 8m x Palomar 5m

No AGN LINER



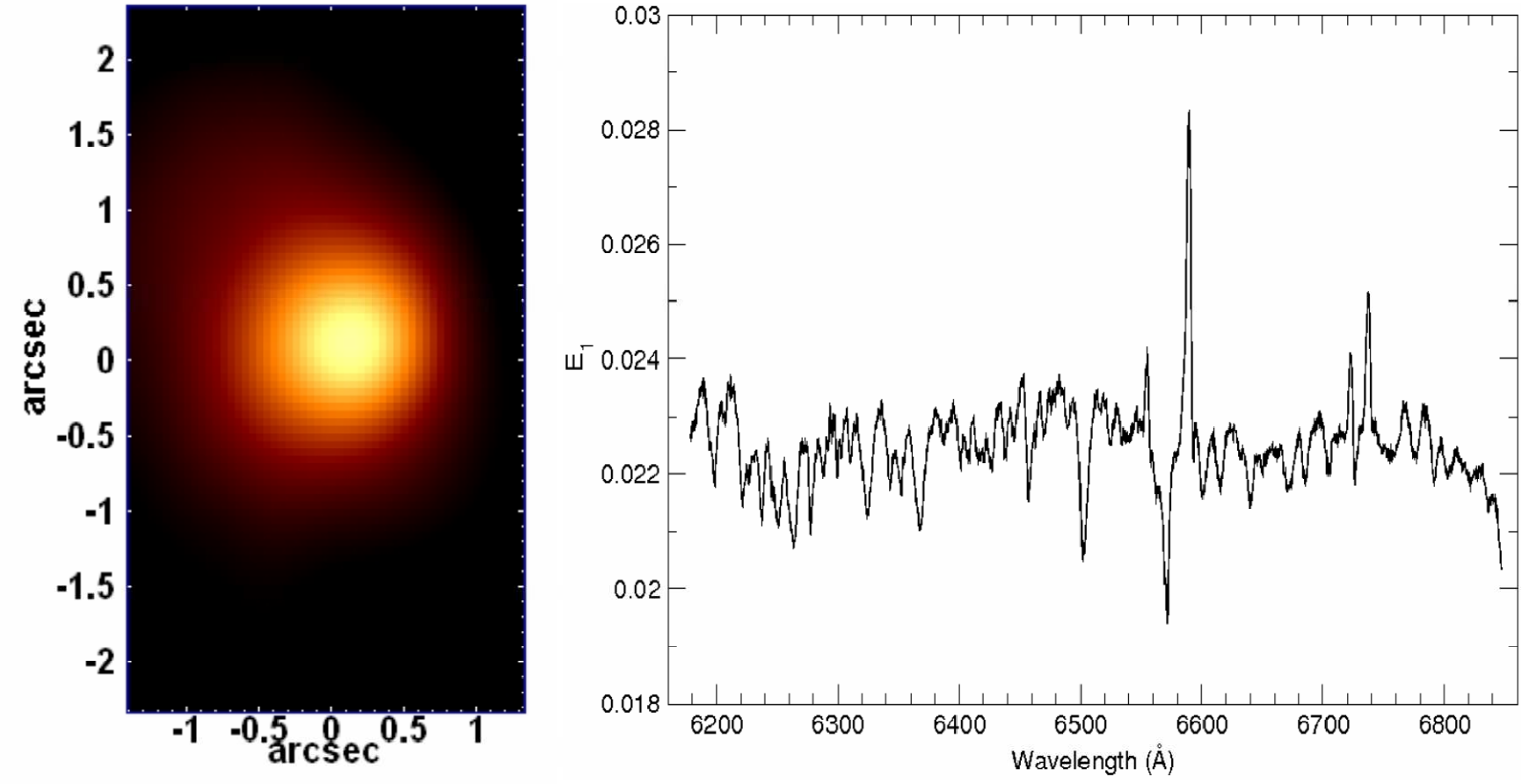


Eigenvalues and variance explained

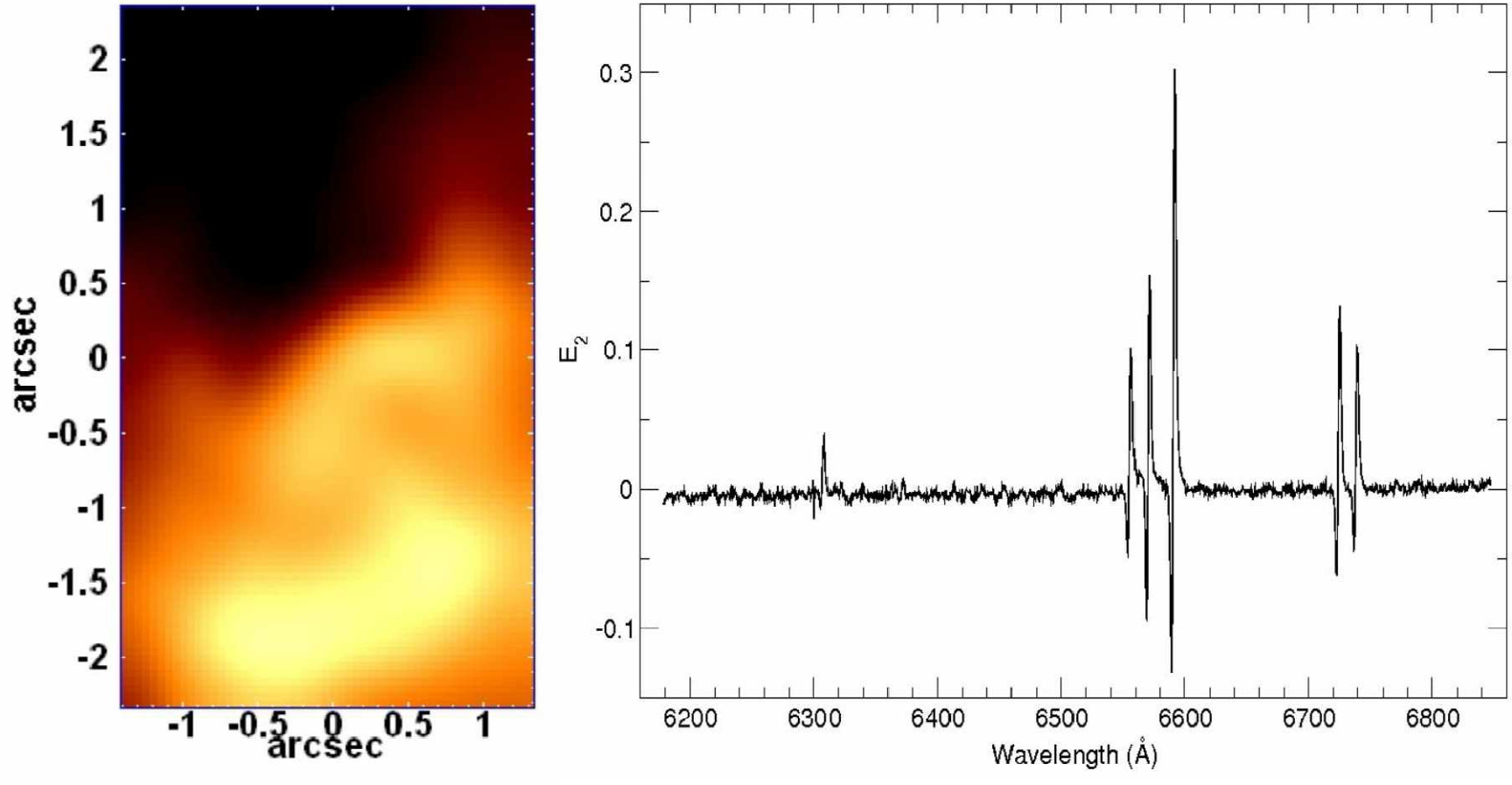
(Huge redundancy)

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

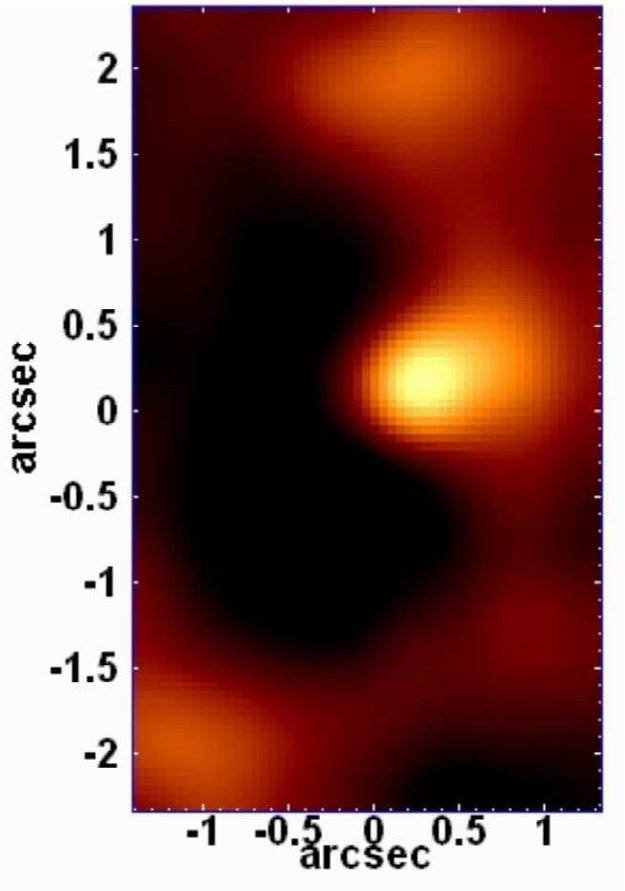
Tomogram 1 and eigenspectrum 1

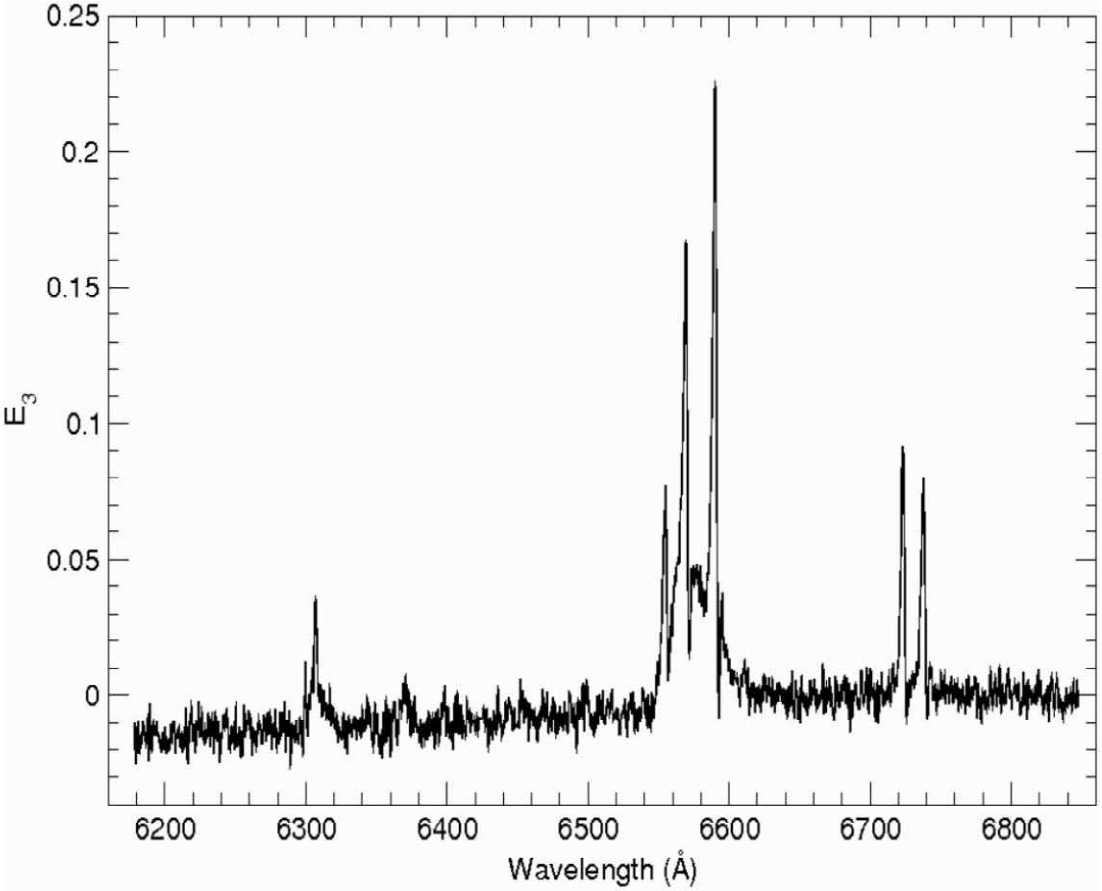


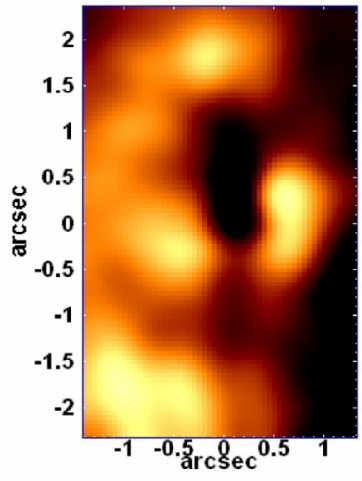
Tomogram 2 and eigenspectrum 2

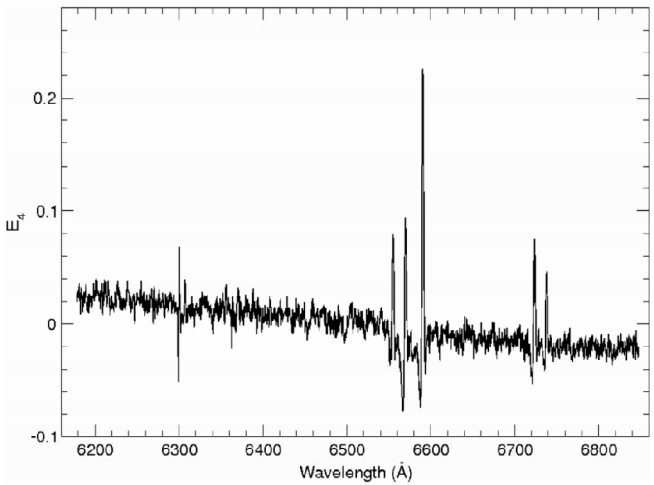


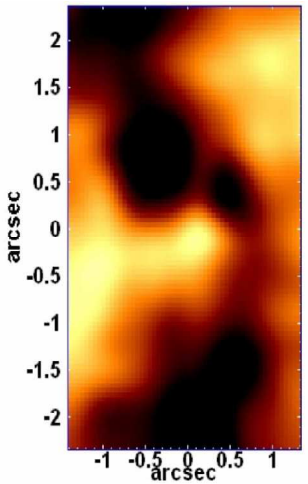
Tomogram 3 and eigenspectrum 3

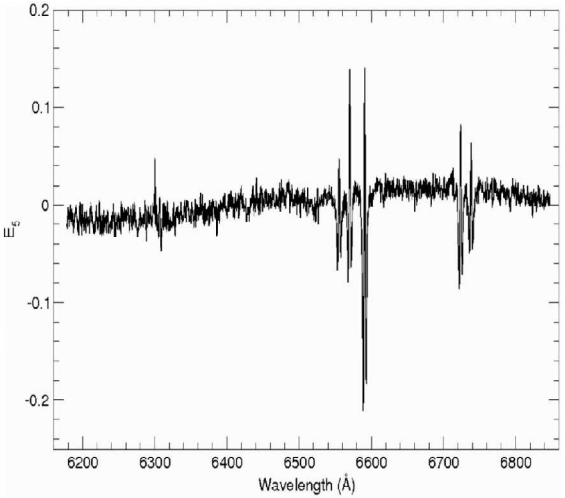


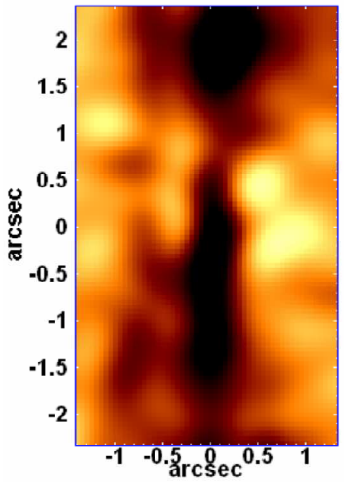


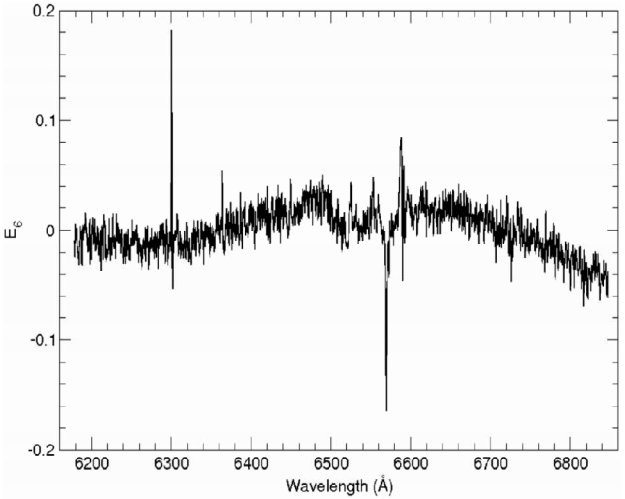

arcsec
arcsec

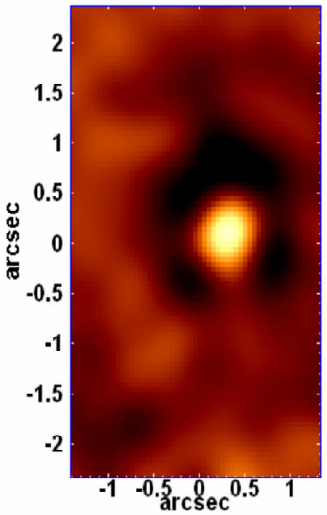

E4
Wavelength (Å)

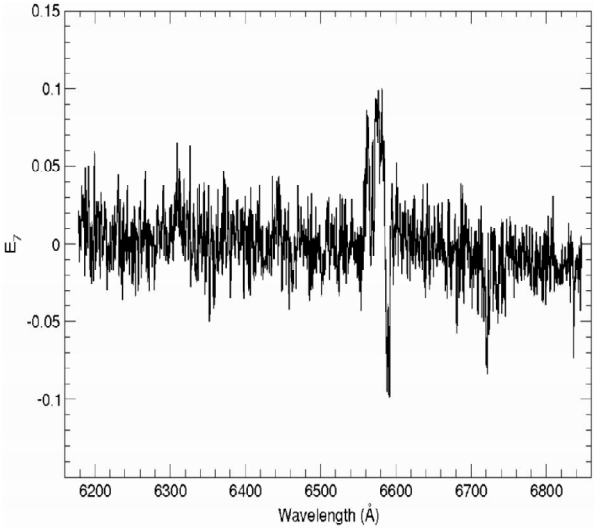

arcsec
arcsec


E5
Wavelength (Å)

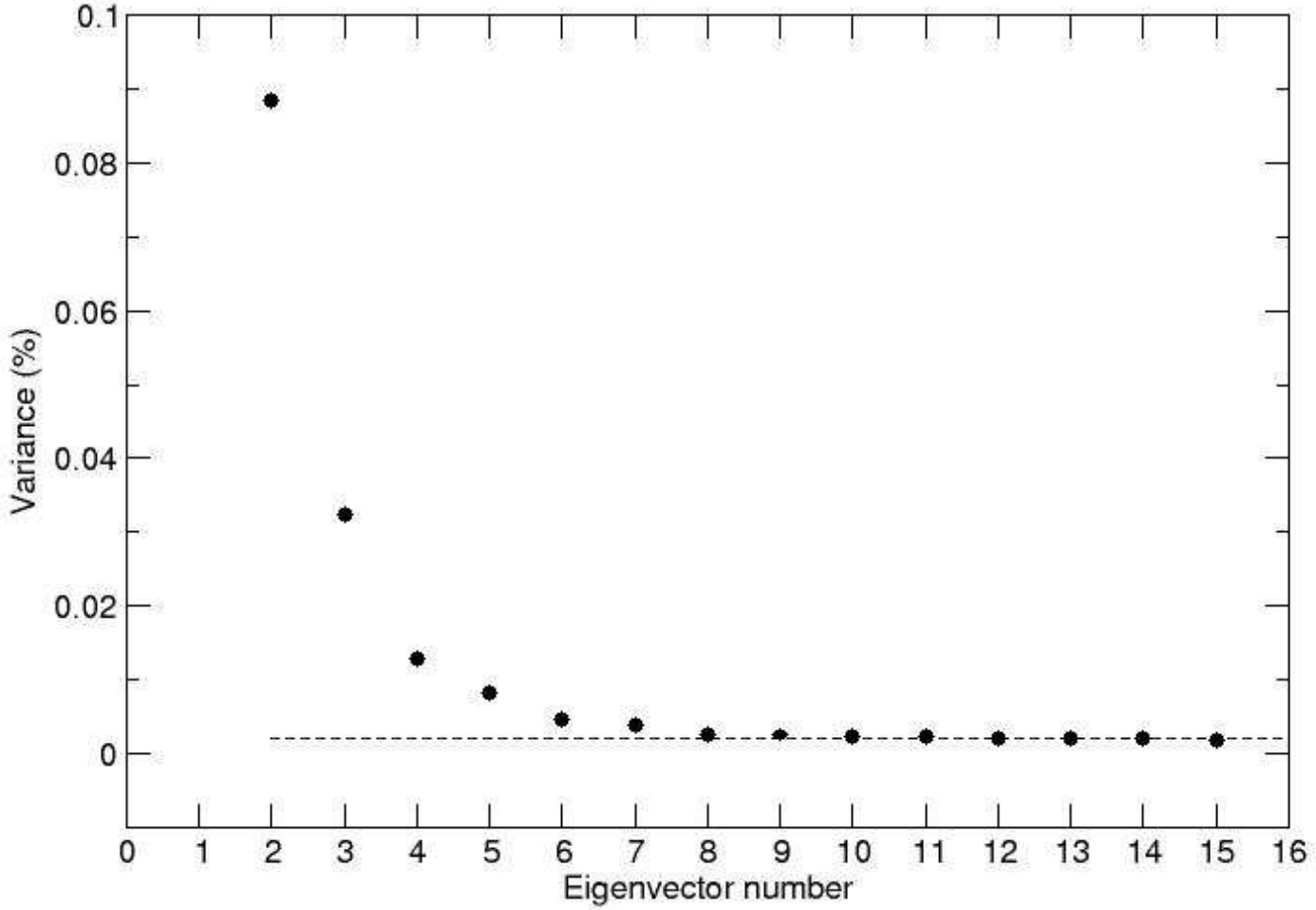

arcsec
arcsec


E6
Wavelength (Å)


arcsec
arcsec

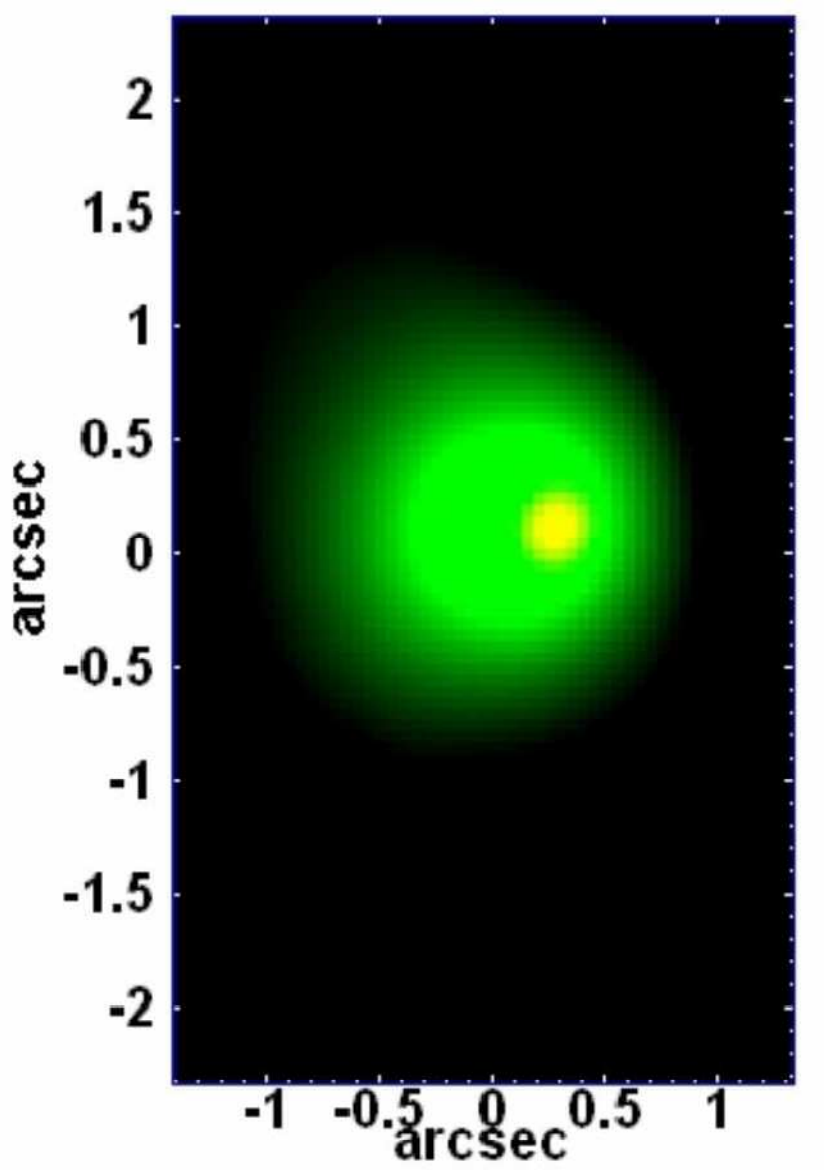

E7
Wavelength (Å)

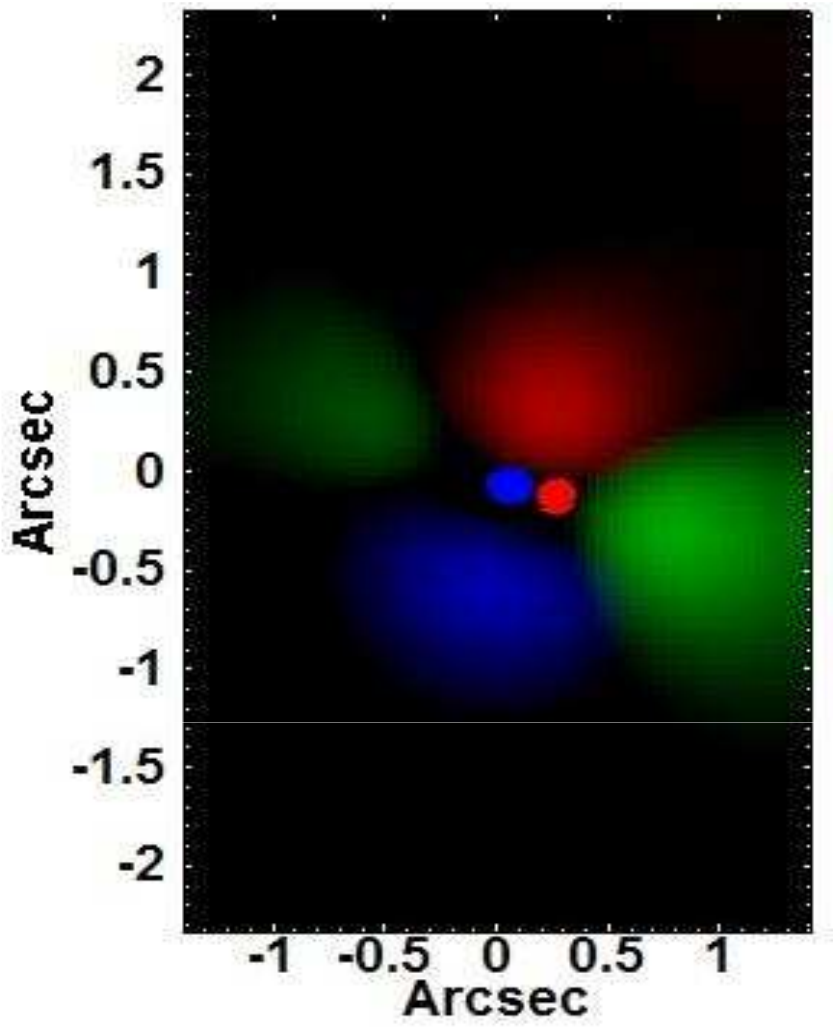
The “scree test”

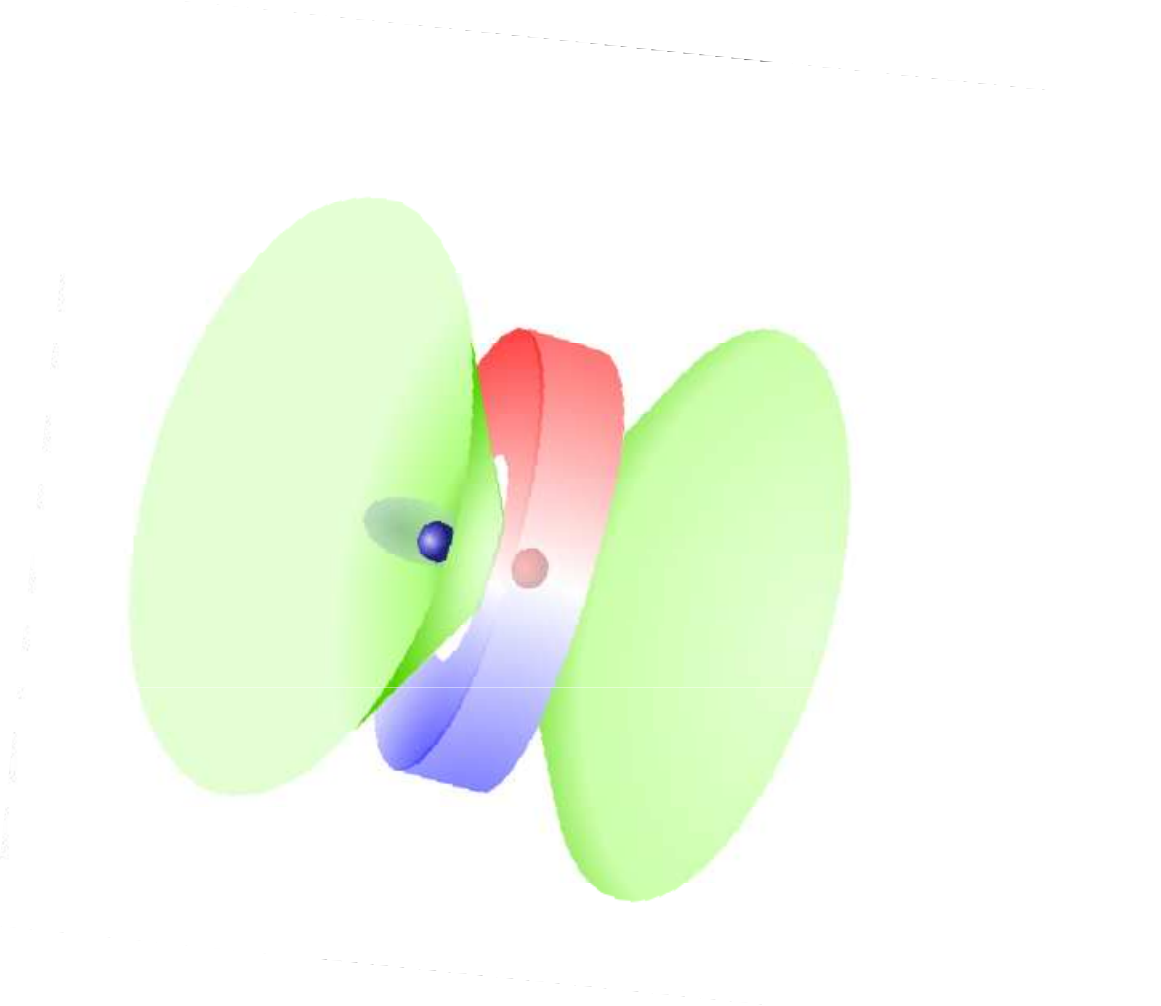


Tomogram 1 (green) + 2 (yellow)

The stellar bulge and the BLR



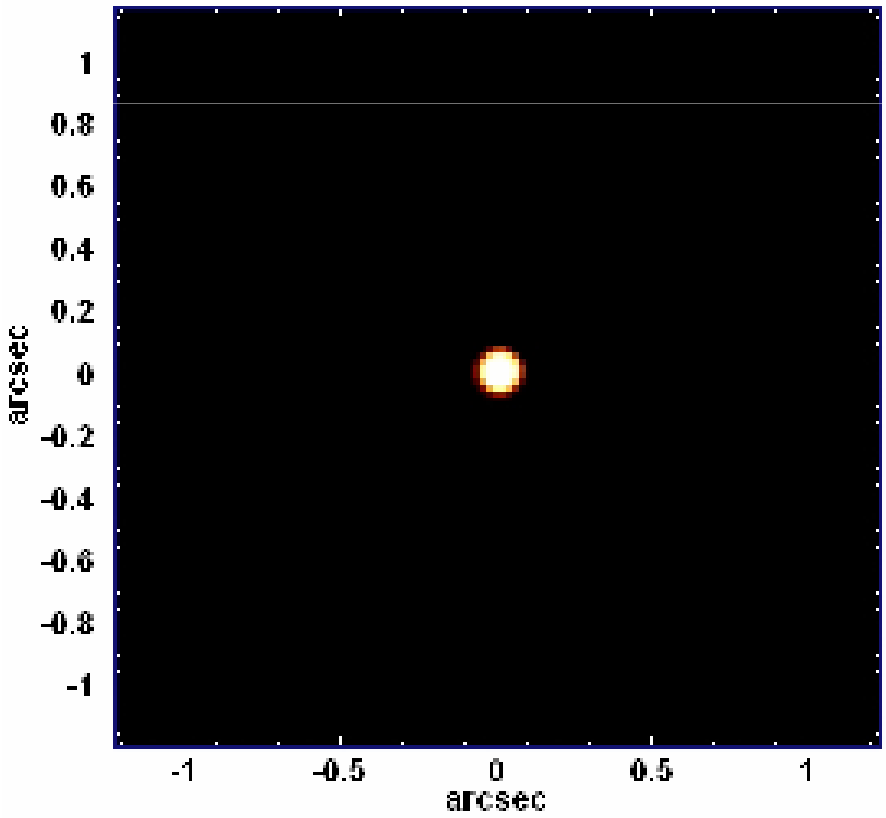


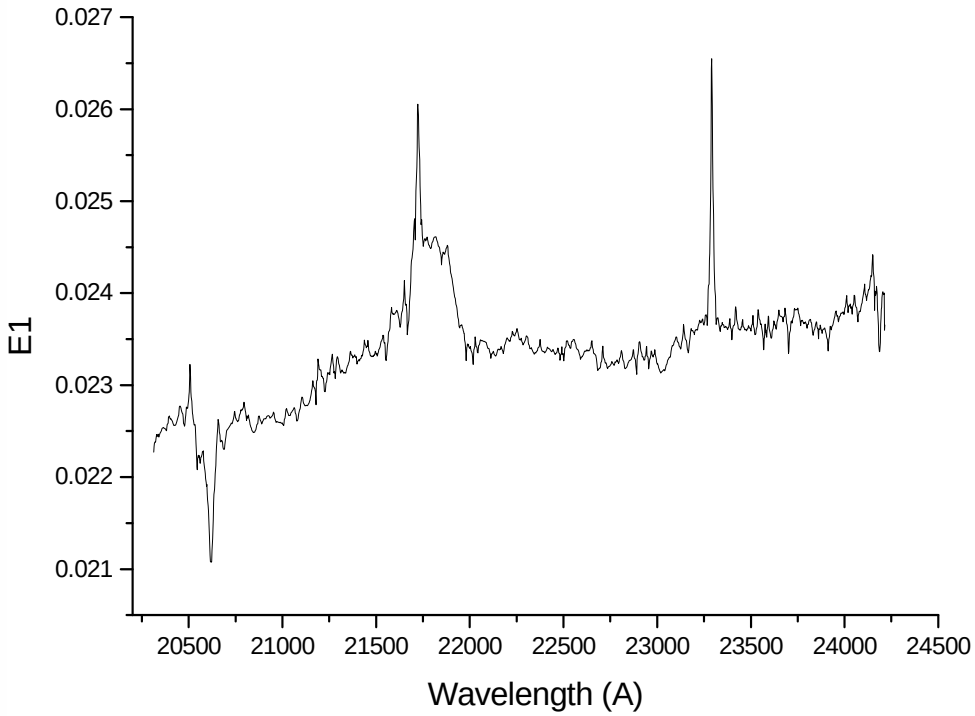


- NGC 7097 – GMOS IFU
- Ap J Lett, in press
- See poster: Tiago Ricci

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
-Strehal ratio after deconvolution = 0.15





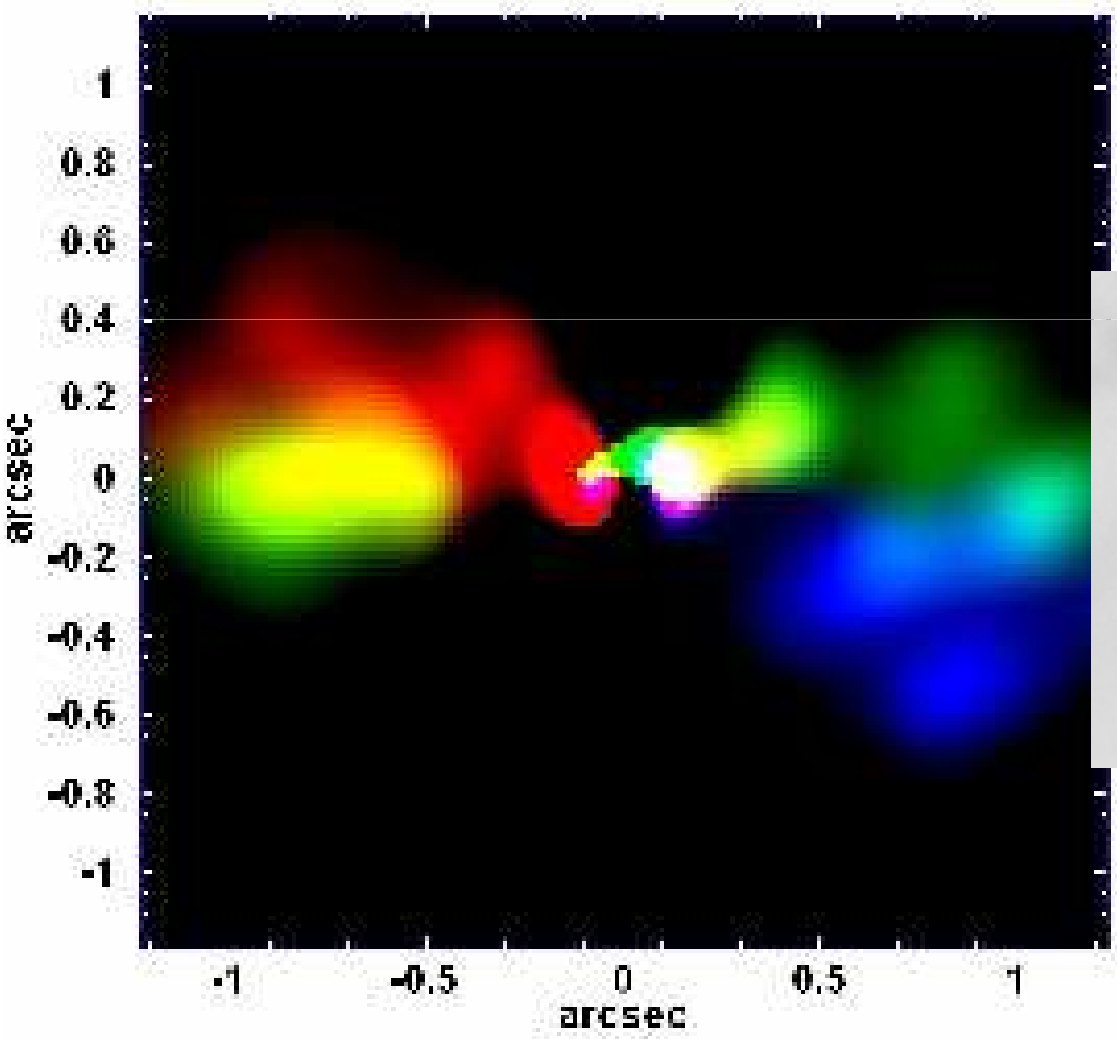
Feature supression and enhancement

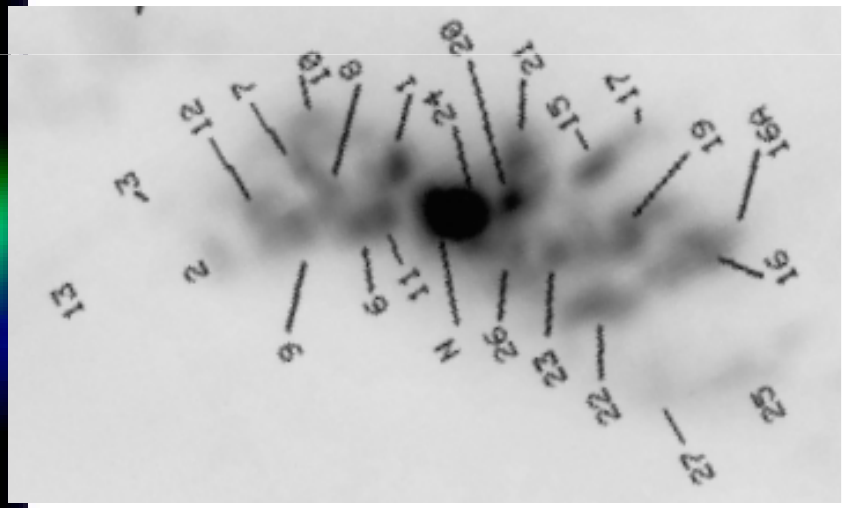
- Defining an object “A”: $\Gamma_k(A) = 1;0$
- Cube with enhanced feature: $I'_{ij\lambda}(A) = \sum_k \left[\left(I'_{ij\lambda}(k)\right) \cdot \Gamma_k(A)\right]$
- Or directly $I'_{\beta\lambda}(A) = T_{\beta k} \cdot \left[\left(E_{\lambda k}\right)_{\Gamma}\right]^T$

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

“software coronagraph”
Br γ after switching-off PC1+PC2

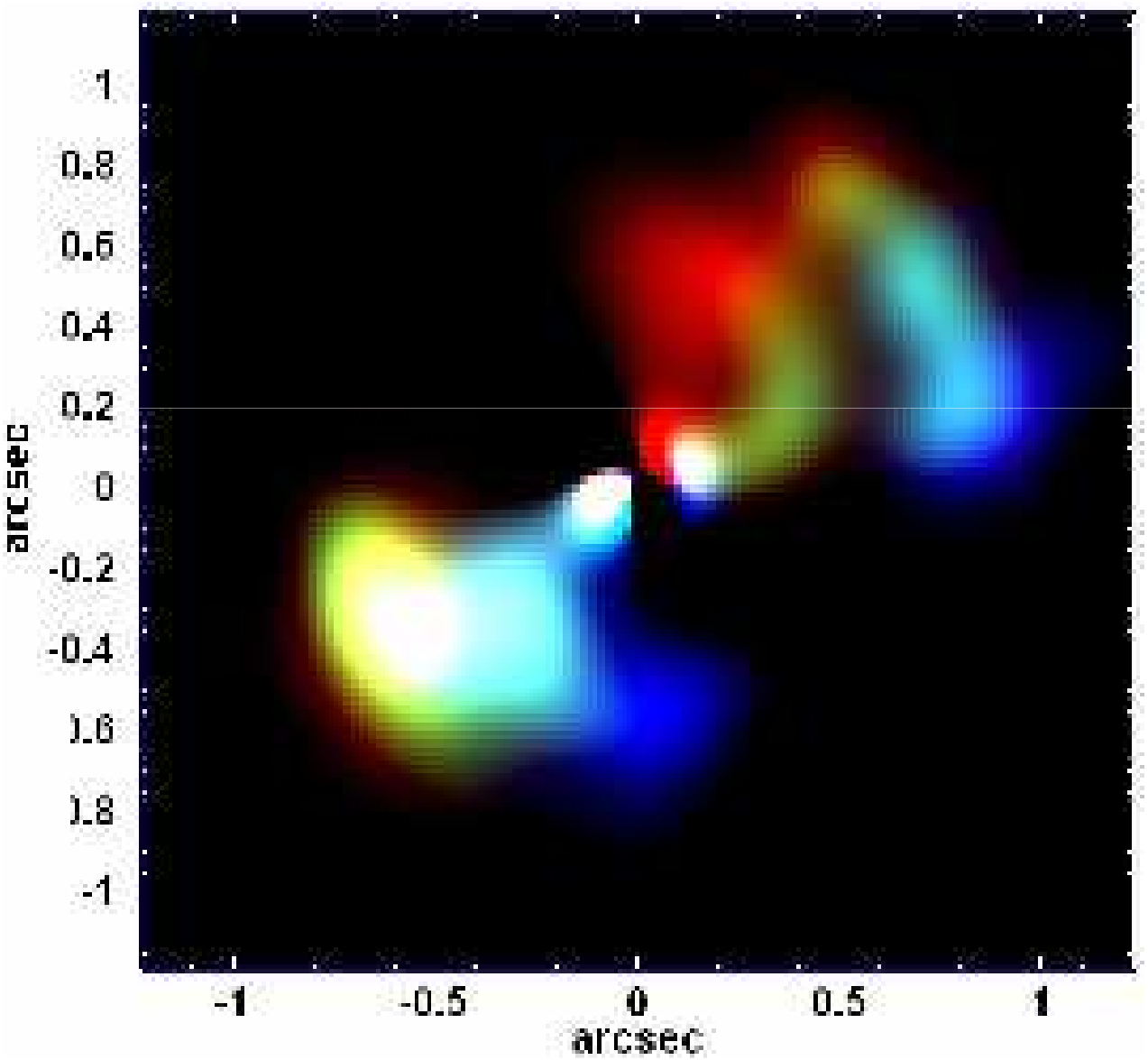
HST observations





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

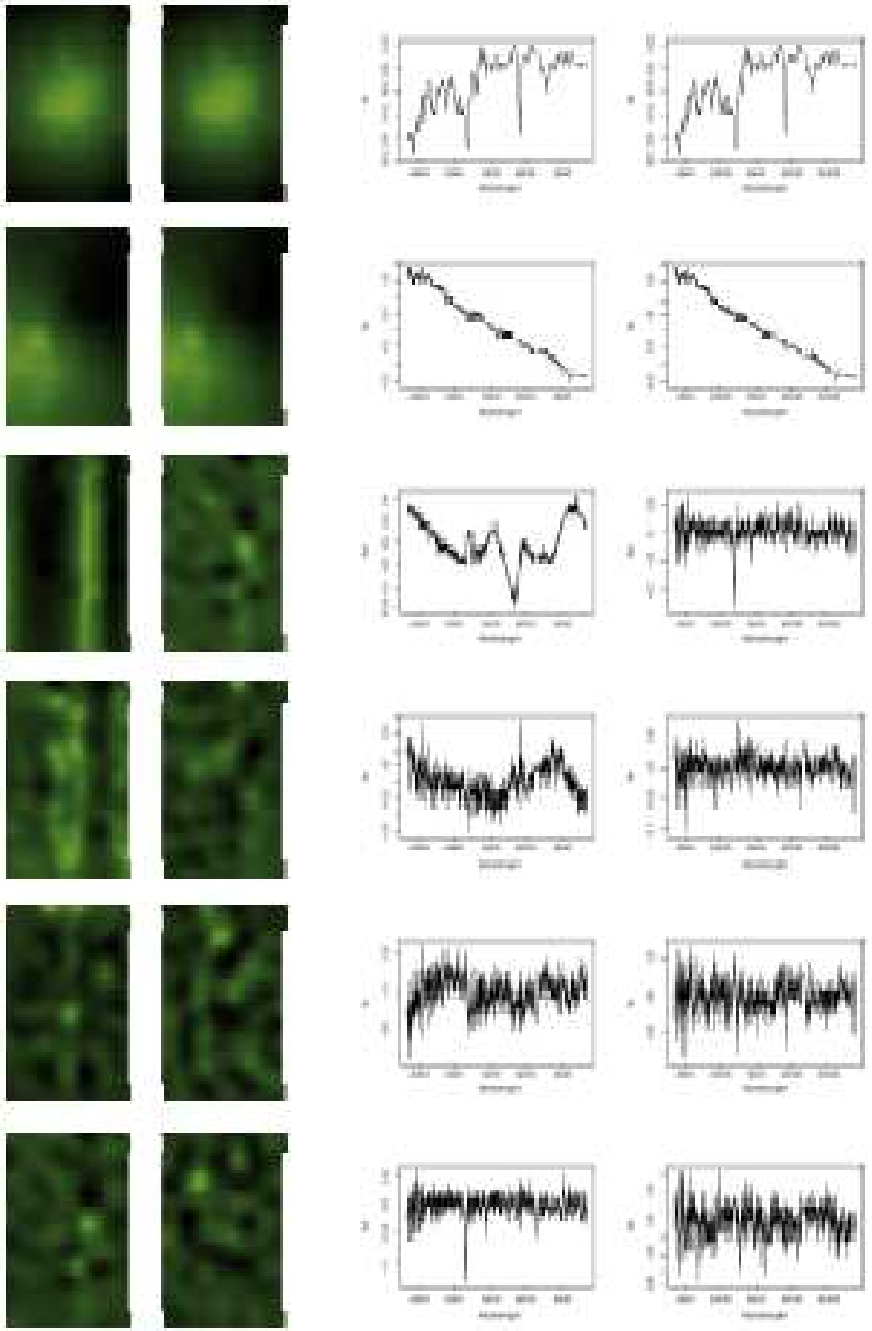
H2 lines – after switching-off PC1+PC2



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

The SOAR data cube toolbox

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

- *IDL (+Python)*

Thank you