Untangling the physical components of galaxies

SDAF: 19/10/2016

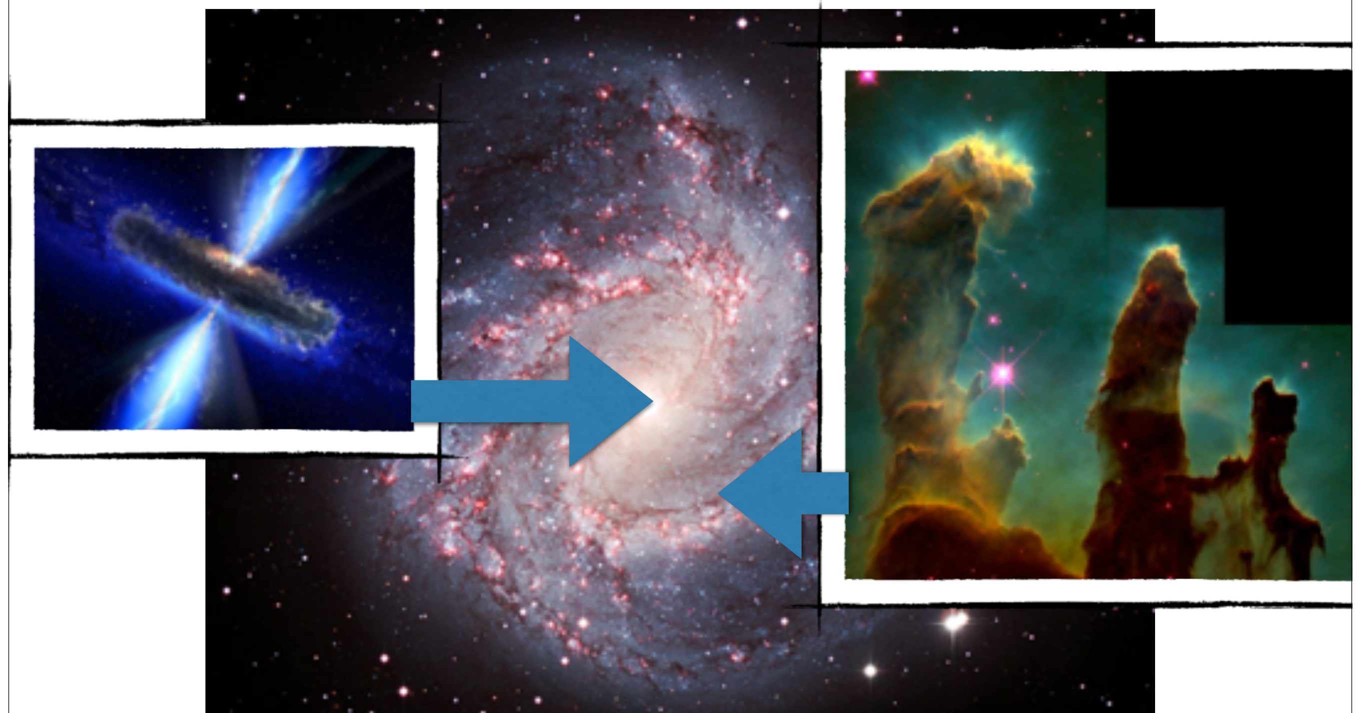
Peter Hurley

Purpose of talk: show how I have used techniques such as PCA and NMF in my work
Time at the end for discussion

What is a galaxy?

M83, hundreds of billions of stars

Galaxies have different regions



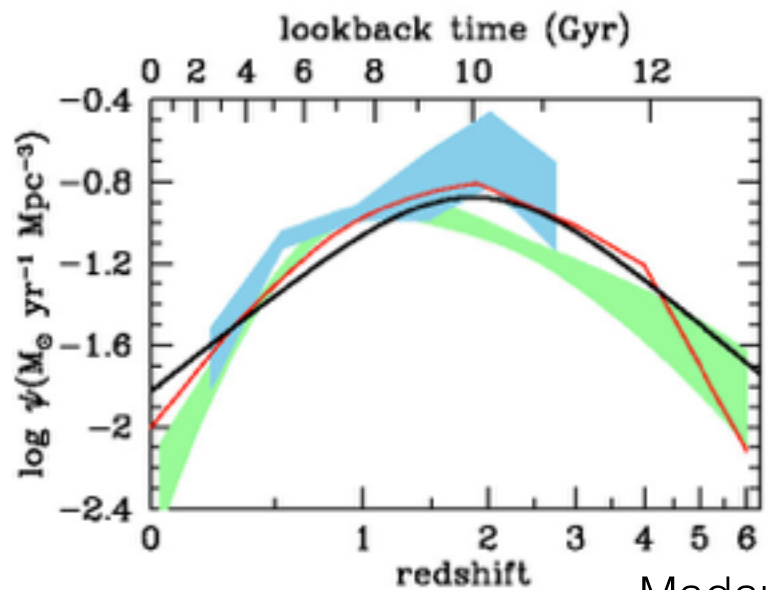
Galaxies have different physical environments:

Active Galactic Nuclei: super massive black hole, accretion disk, gives of huge amount of energy, XRAY radio, infra red dusty torus

Star forming regions: molecular clouds, collapse, form stars, these give of UV-optical light, but heat up surrounding material, e.g. gas and dust..

Galaxy evolution

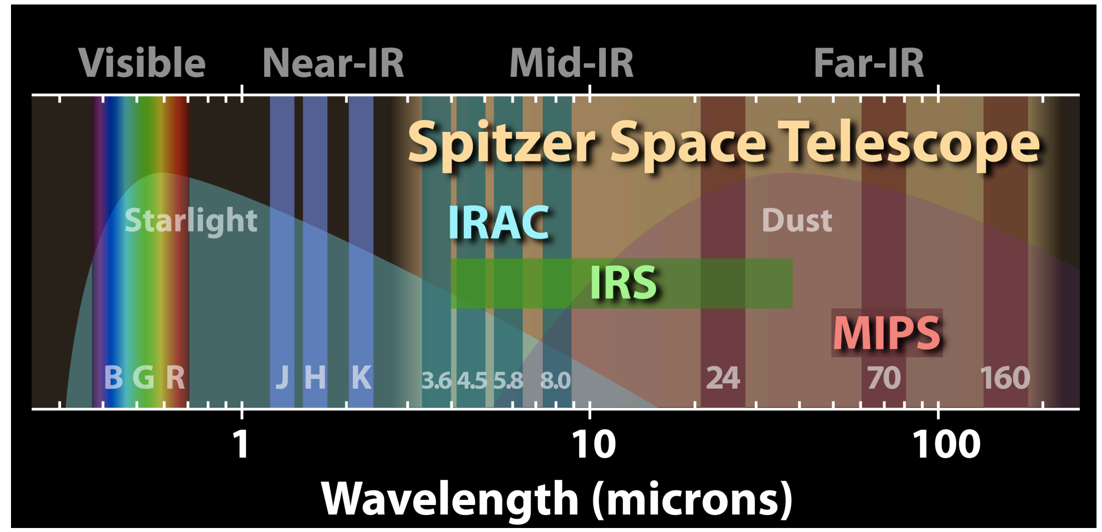
- How does SFR and AGN change over time?
- Connection between SF and AGN?



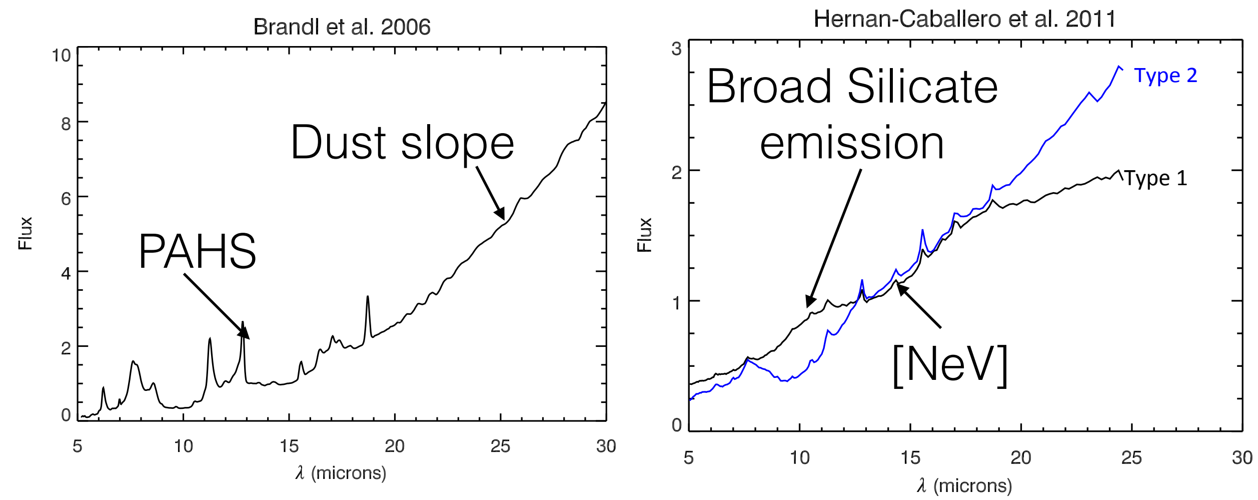
Madau et al. 2014

black curve SFR
red curve XRAY -AGN

Understanding connection between different regions requires multi-wavelength data

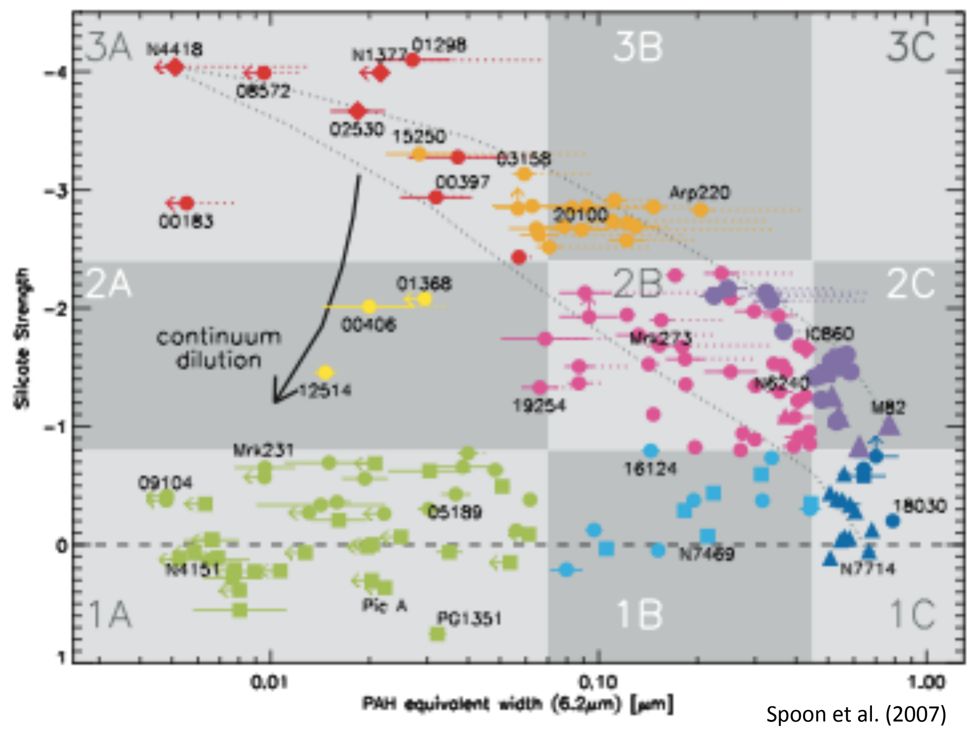


The MIR contains spectral signatures from starbursts and AGN



People tend to use only one or two specific features as signatures

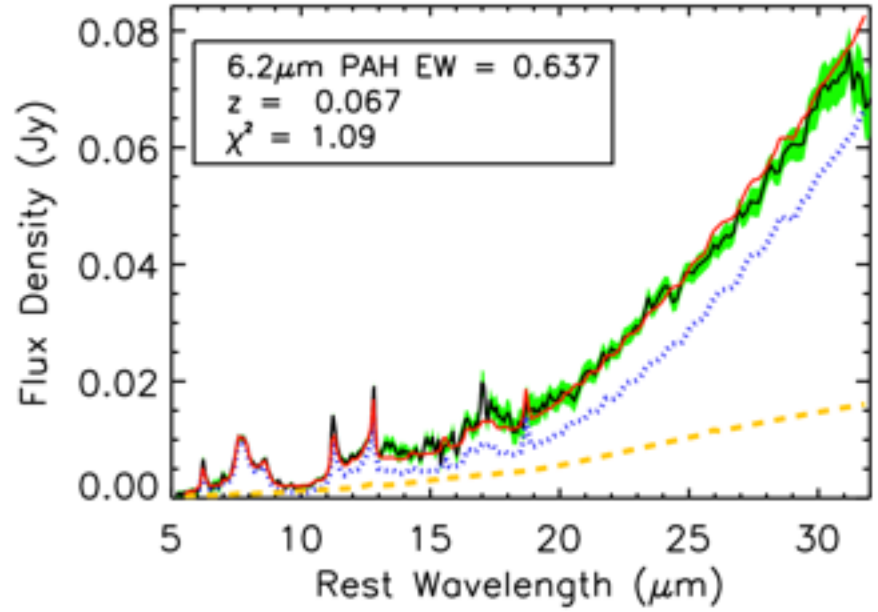
Spectral features are used as diagnostics, but they ignore information contained in the rest of the spectrum



Spoon et al. (2007)

A specific feature missing e.g. bad spectra?
What happens if spectra is noisy?

Spectral templates are often based on 'Average templates' and 'specific objects' rather than the physical environments



Wu et al. 2011

Ideally, we want a set of MIR spectral components that relate to the physical environments of galaxies.

What do the templates for these physical environments look like?

Can we use multivariate analysis techniques to:

1) Use the most of the data
2) Learn the templates from the data

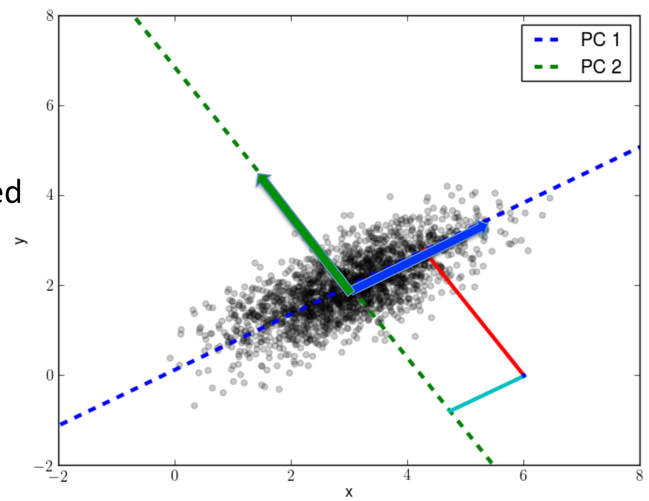
Multivariate Analysis Techniques can help us get the most out of high dimensional data

Principal Component Analysis (PCA)
is the most popular, also ICA, NMF etc

Used for:

- Investigating what features are correlated
- Data compression
- Source separation

PCA 2D example

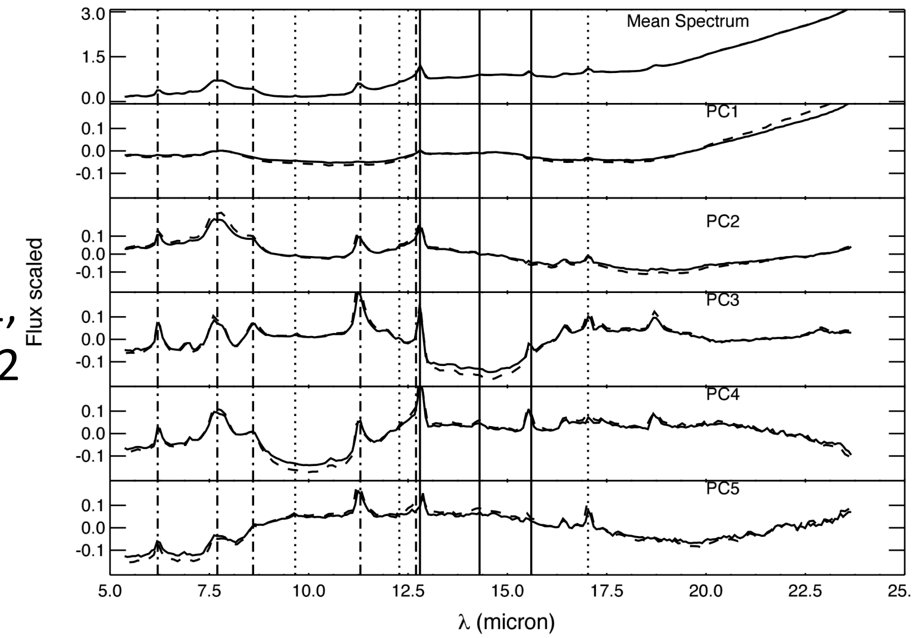


PCA: finds components that provide the most variance, and are orthogonal:.. i.e. modelling as multivariate gaussian

PCA applied to IRS spectra of ULIRGS: components have a statistical interpretation

119 local ULIRGs

- Wang et al 2011,
- Hurley et al 2012



$$F(\lambda) = Mean(\lambda) + c_1 \cdot PC1(\lambda) + c_2 \cdot PC2(\lambda) + c_3 \cdot PC3(\lambda) + c_4 \cdot PC4(\lambda) + c_5 \cdot PC5(\lambda)$$

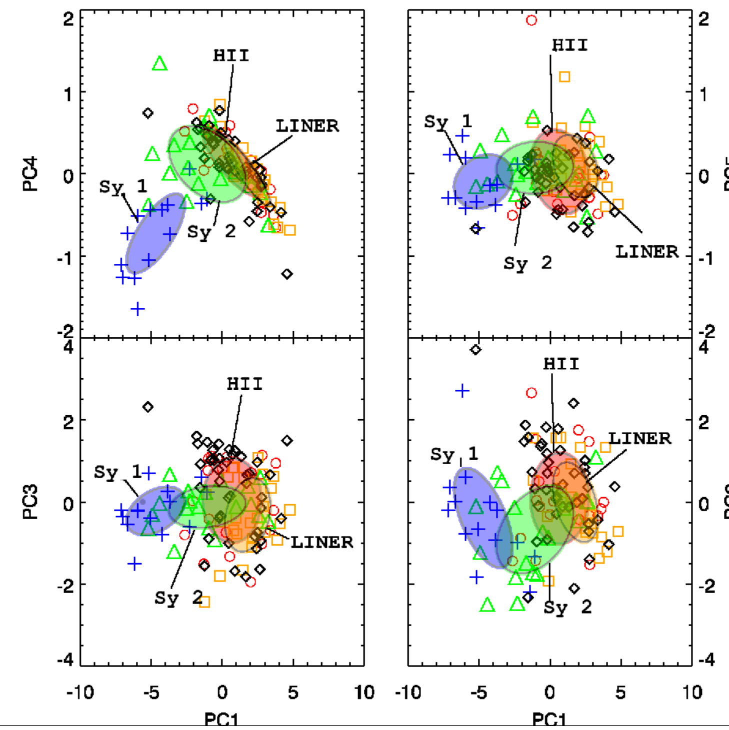
Features that vary most e.g. dust slope, then PAH features
NOTE: positive and negative weights and templates

PCA still useful for classifying objects

•Weights define co-ordinates in a 5D, PCA space

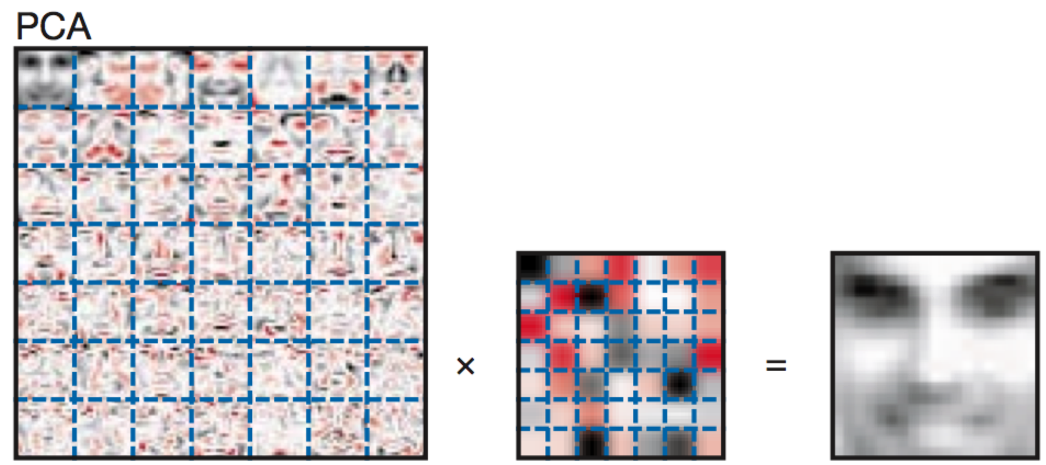
•Different types of object lie in different areas of PCA space

•Learnt through optical classification



PCA as a data compression technique:

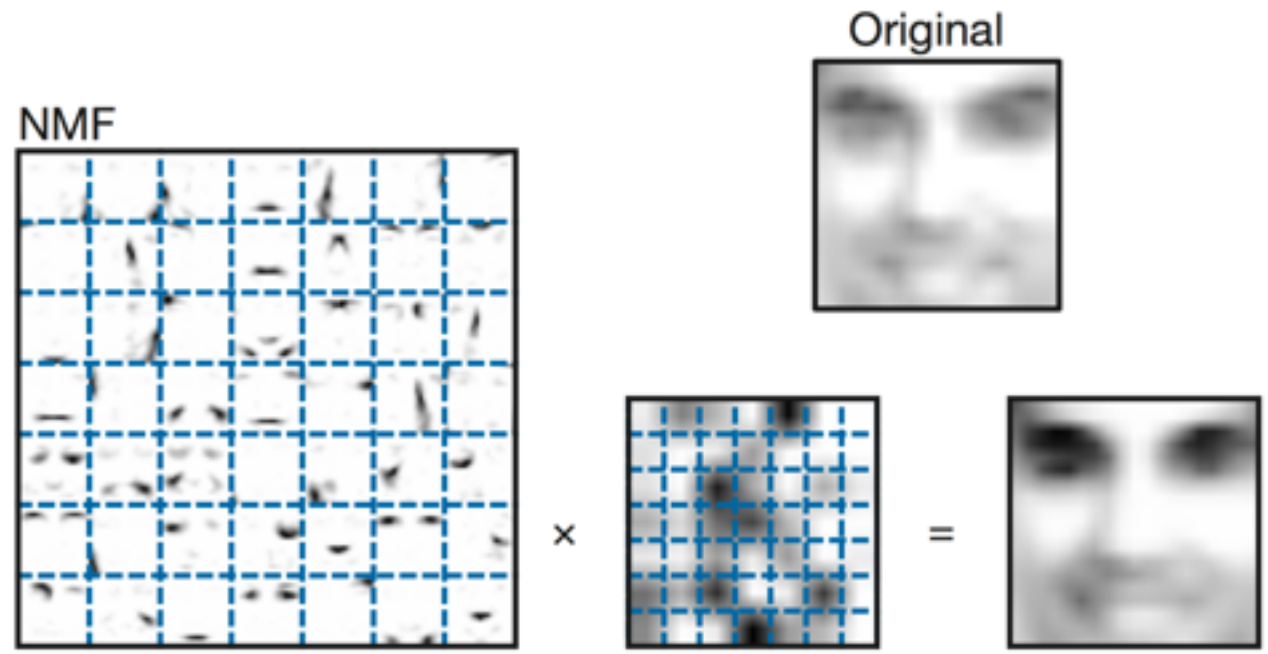
PCA provides a statistical interpretation



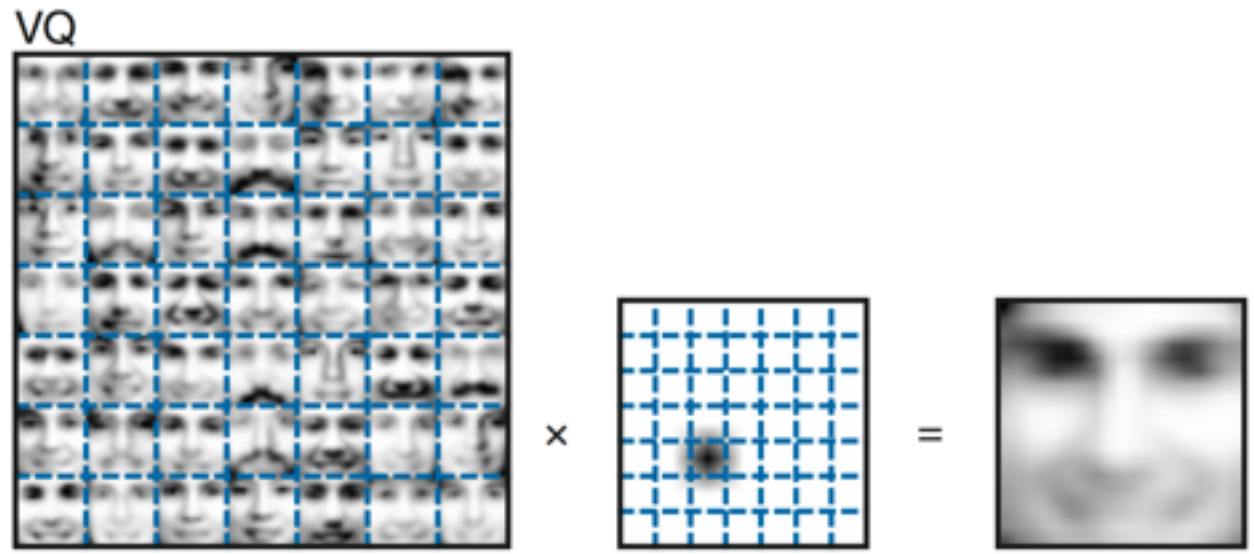
Lee and Seung (1999)

Take a bunch of faces, learn PCs

NMF provides a physical interpretation



Vector Quantisation provides a classification interpretation



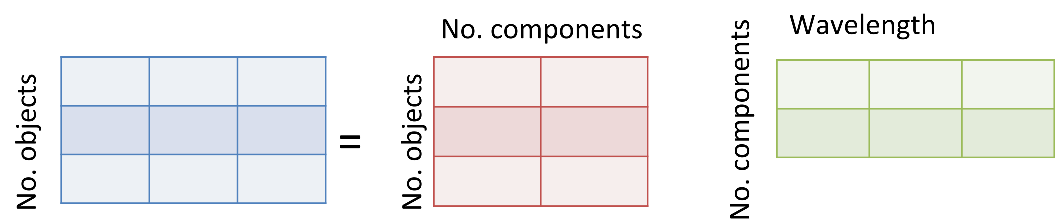
In VQ, each column of H is constrained to be a unary vector, with one element equal to unity and the other elements equal to zero. In other words, every face (column of V) is approximated by a single basis image (column of W) in the factorization V WH

NMF is just a matrix factorisation with positive constraints

Weights

Components

$$\underset{I\times J}{\boldsymbol{X}} = \underset{I\times K}{\boldsymbol{A}}\,\underset{K\times J}{\boldsymbol{B}} + \underset{I\times J}{\boldsymbol{N}}, \quad \text{or equivalently,} \quad x_{ij} = \sum_{k=1}^{K} a_{ik} b_{kj} + n_{ij},$$



By applying NMF to IRS spectra, we assume galaxy spectra are a linear combination of components

Weights

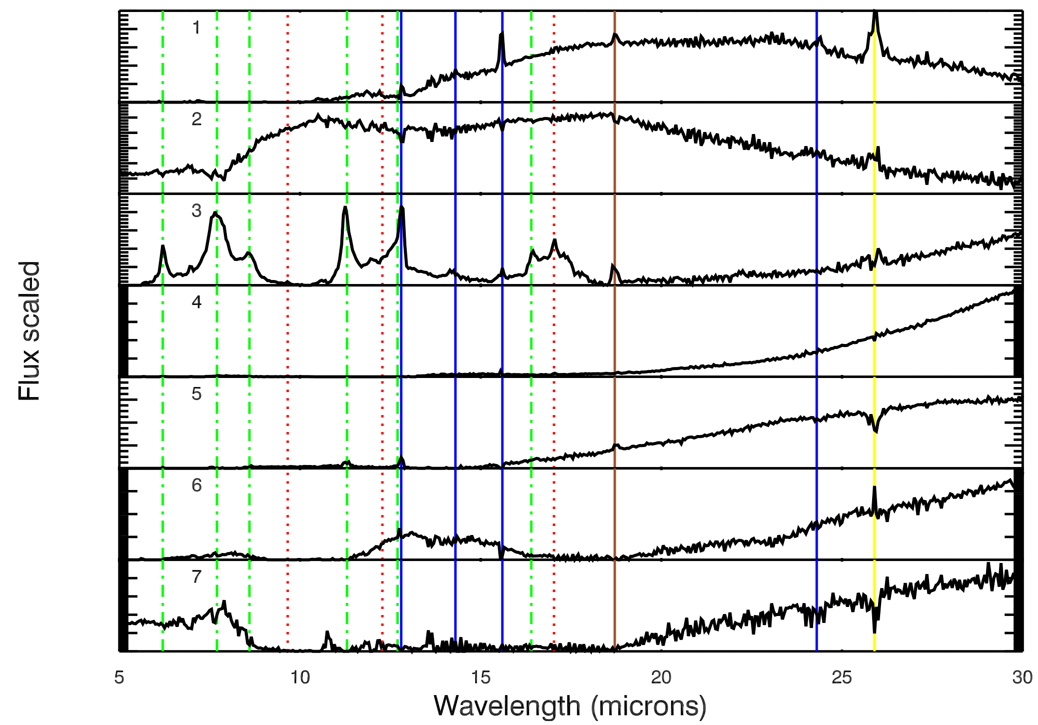
$$F(\lambda) = a \cdot T_{SF}(\lambda) + b \cdot T_{AGN}(\lambda) + c \cdot T_{MC}(\lambda) + d \cdot T_{C}(\lambda)$$

Components

More appropriate than PCA:

$$F(\lambda) = Mean(\lambda) + c_1 \cdot PC1(\lambda) + c_2 \cdot PC2(\lambda) + c_3 \cdot PC3(\lambda) + c_4 \cdot PC4(\lambda) + c_5 \cdot PC5(\lambda)$$

NMF with 7 components

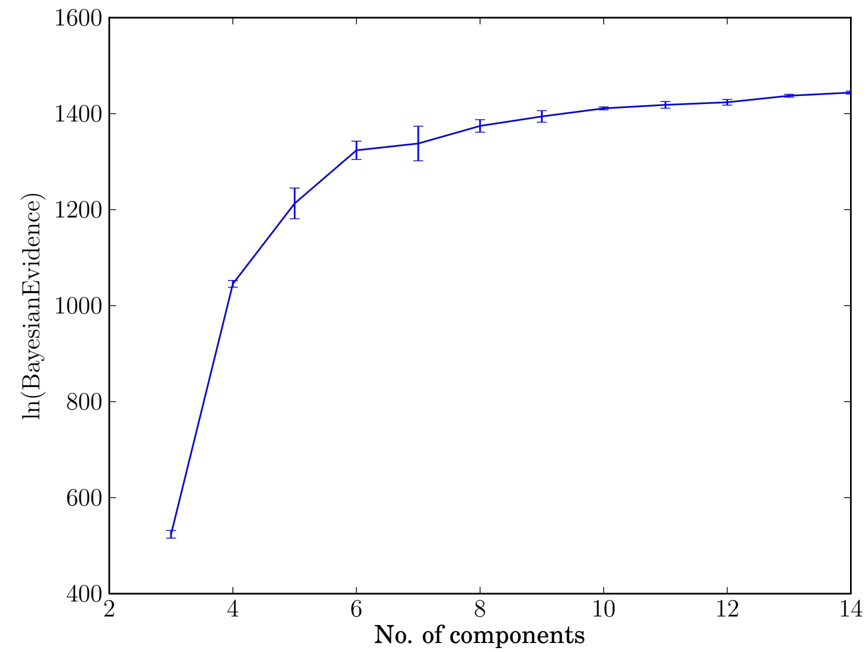


729 galaxies in total

How many components?

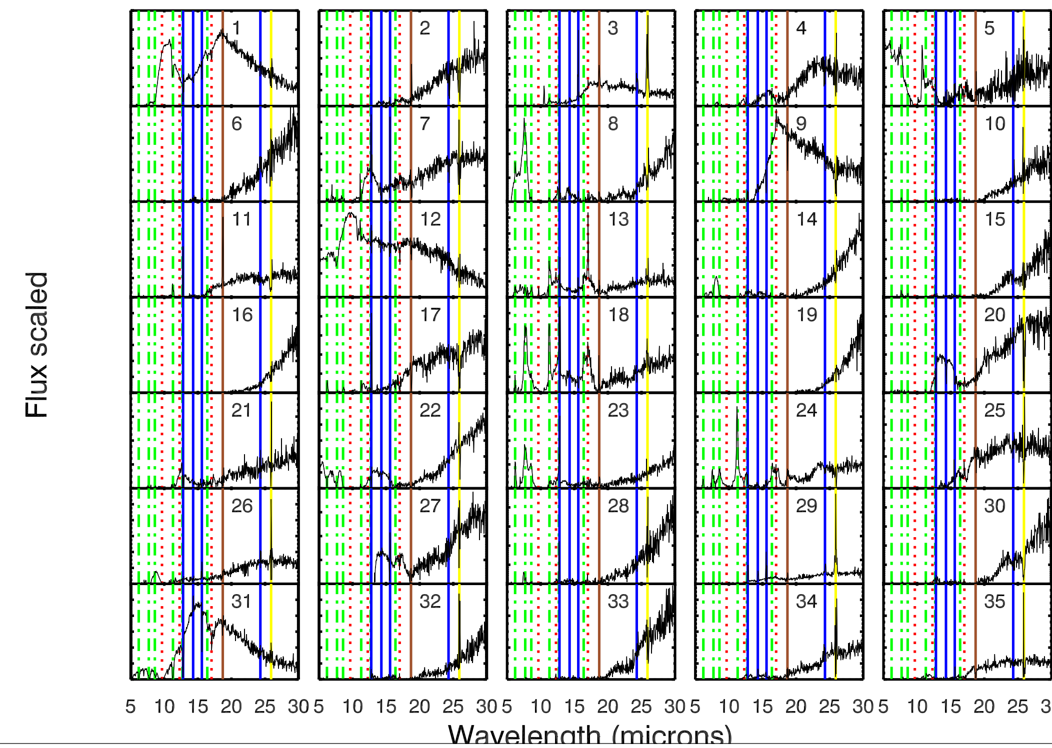
Use MULTINEST (Feroz et al. 2007) to calculate Bayesian Evidence

Why >15?
•Physics is not linear
•Uncertainties are too conservative



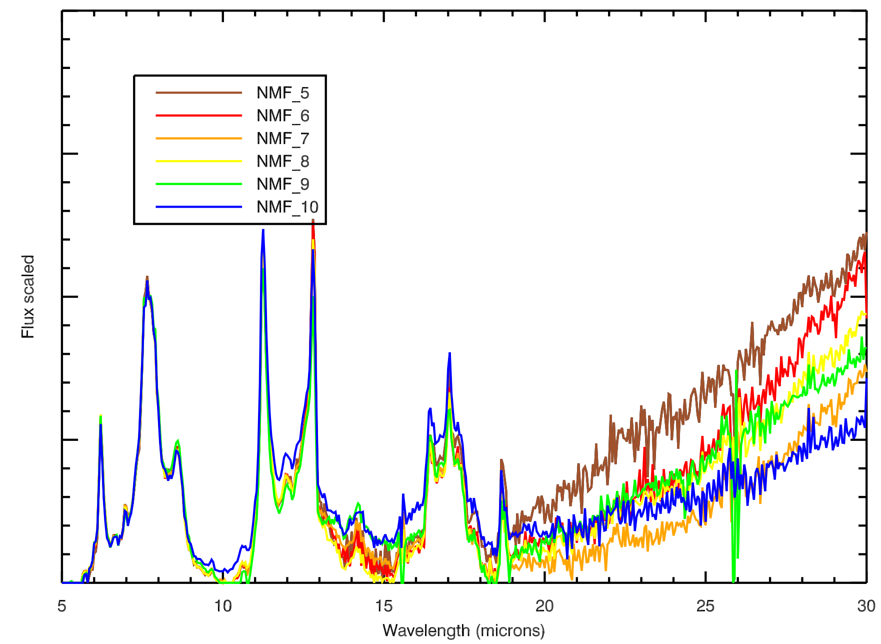
Fit all galaxies with template set, calculate Bayesian Evidence

An NMF set with > 15 components is not useful

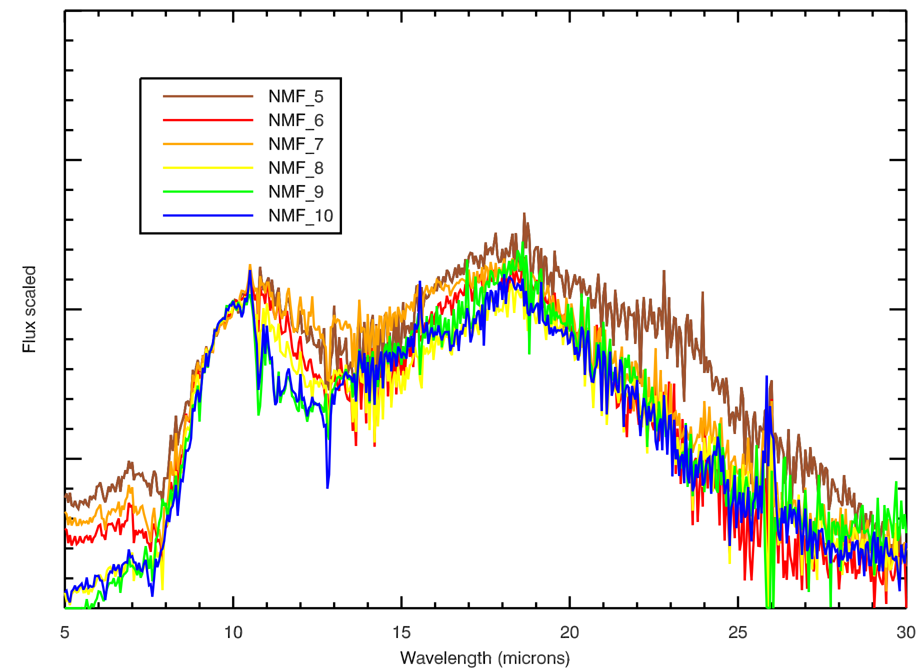


unphysical,

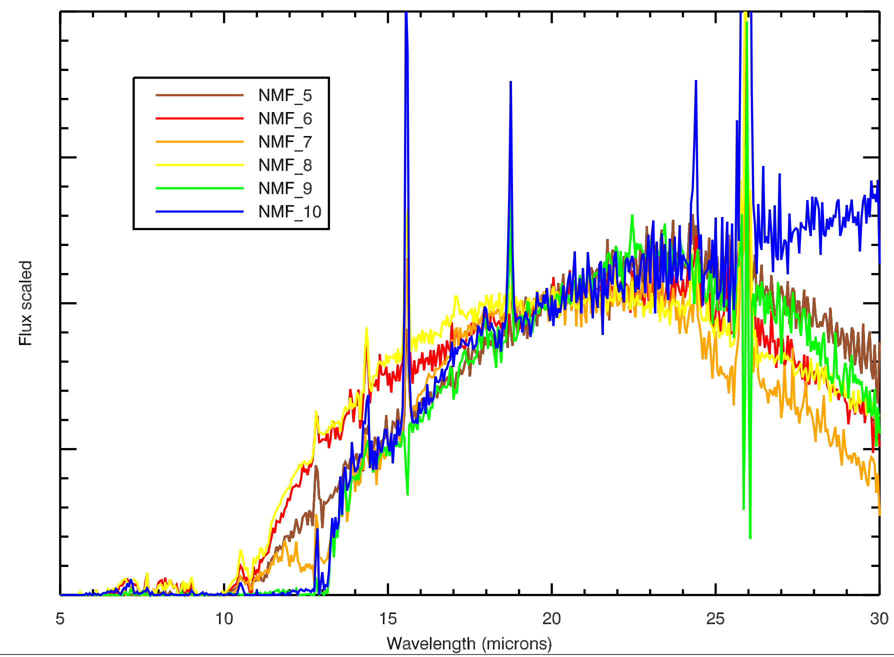
There is a common star formation component



There is a common 'hot silicate dust' component



There is a common 'hot dust and ionised gas' component

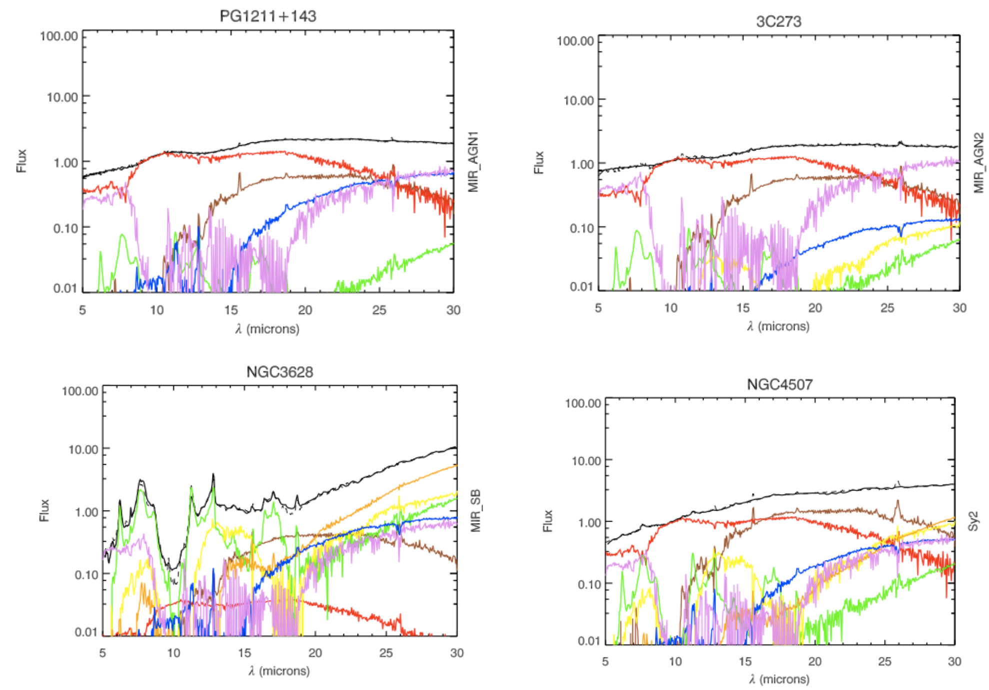


Finding common suggests the 3 components are fundamental spectral building blocks whose behaviour is linear

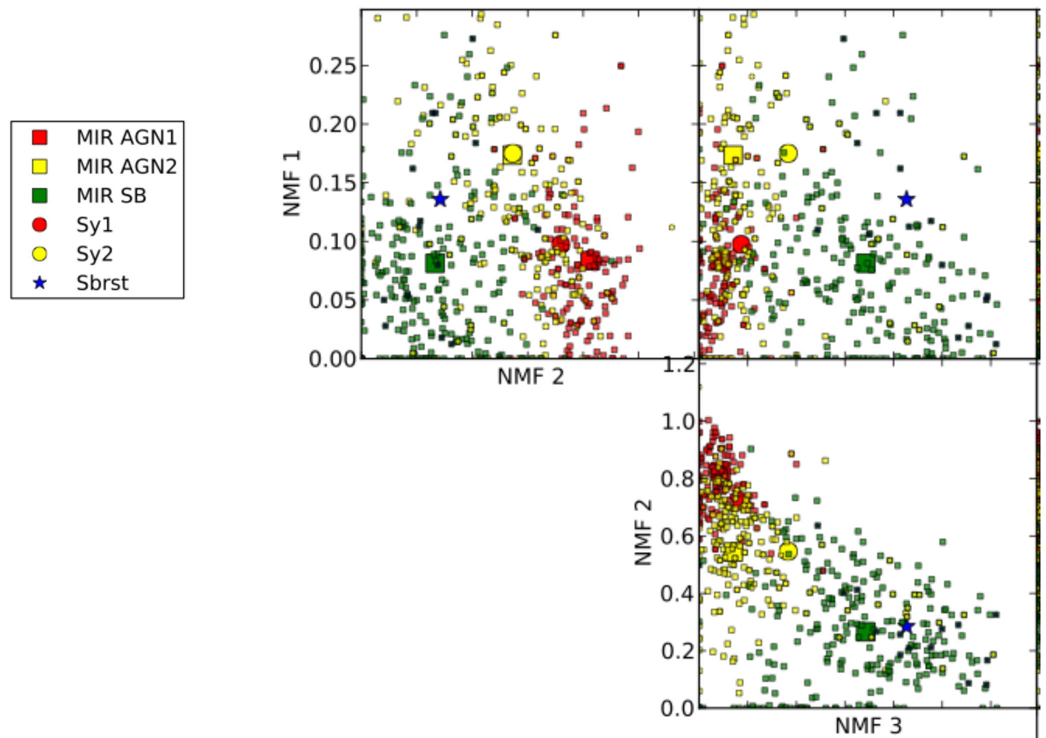
- Common star formation component:
  - contains all the PAH features,
  - star formation regions look similar amongst a range of galaxies

- Unified model of AGN predict silicate emission from type 1 AGN:
  - Silicate emission is a fundamental spectral component
  - Occurs in more than just type 1 AGN
  - Evidence for Clumpy torus?

We can use an NMF set with 7 components to fit galaxies



We can use our components to get the star formation and AGN contribution



Use a data compression technique..

Assumptions of technique affect results you get out

PCA assumptions:
- Orthogonal, linear components,
- Multi-variate Gaussian Distribution
- Variance comes from something interesting

NMF assumptions:
- Linear combination
- Positive constraint on weights
- Positive weight on templates