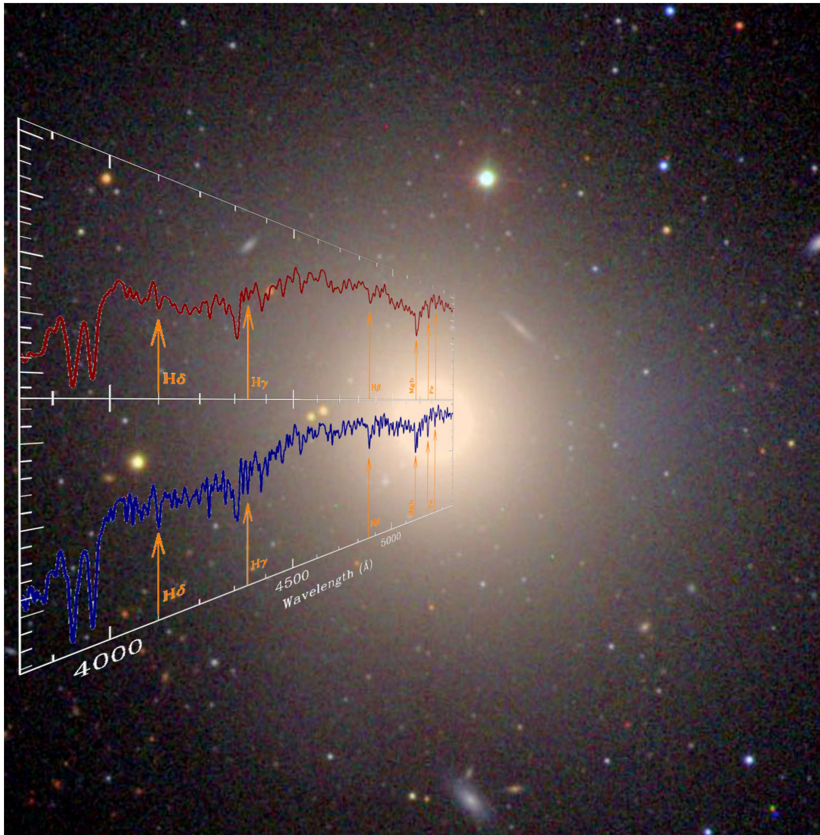
An information theory approach to stellar populations in galaxies

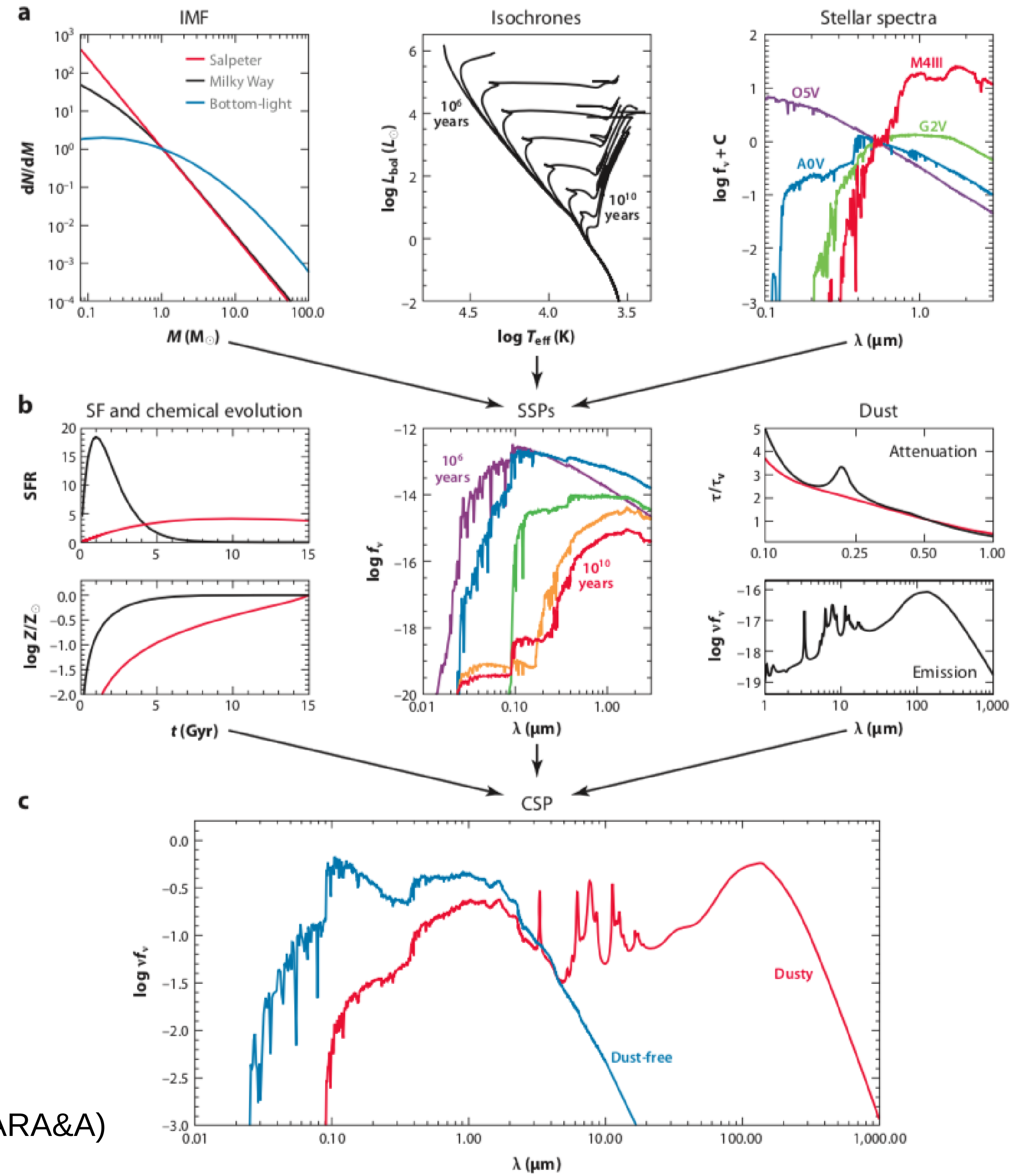
Ignacio Ferreras
IAC & UCL

Celebrating Ofer’s 60th
Cumberland Lodge Meeting
April 9, 2019



The star formation/galaxy formation connection

- Galaxy spectra encode a vast amount of information about the underlying stellar populations
- The age and metallicity distribution of these populations are a representation of the past star formation history and chemical enrichment history



Conroy (2013 ARA&A)

A model driven approach

The spectrum of a galaxy is the result of a superposition of individual stellar populations produced by the effective star formation history: given by the star formation rate ψ(t), and the chemical enrichment track Z(t)

If we assume dustless populations, the problem is linear

$$F(\lambda) = \int_0^{t_G} \psi(t)\Phi[\lambda; t_G - t, Z(t)]dt,$$

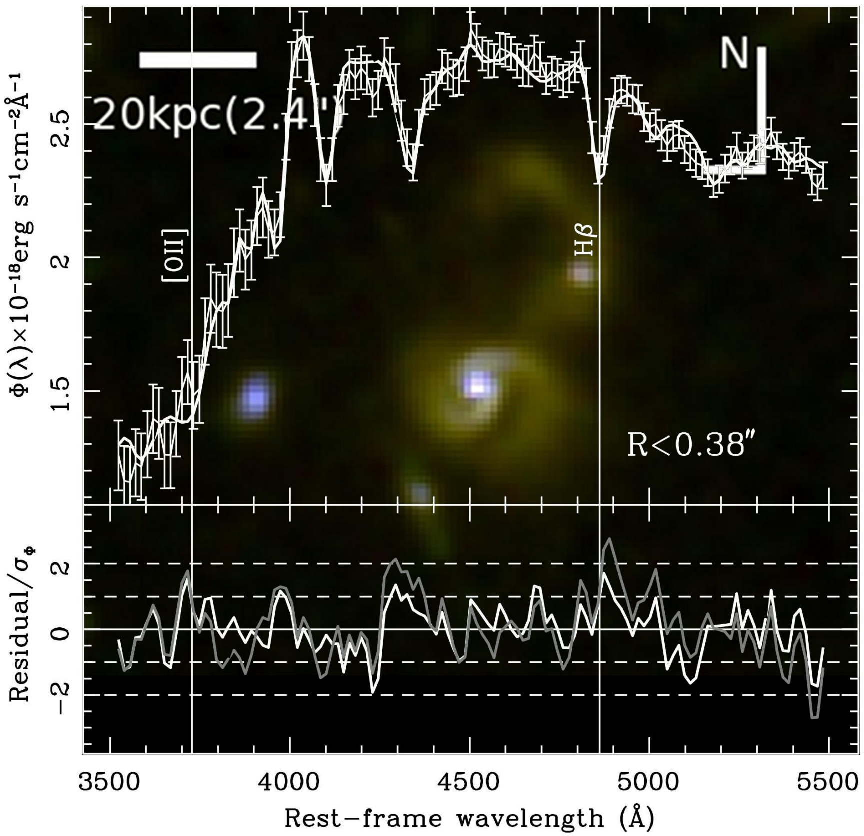
where Φ(λ; τ,Z) are the spectra of simple stellar populations (SSPs), corresponding to stellar ensembles at fixed age (τ) and chemical composition (Z, etc).

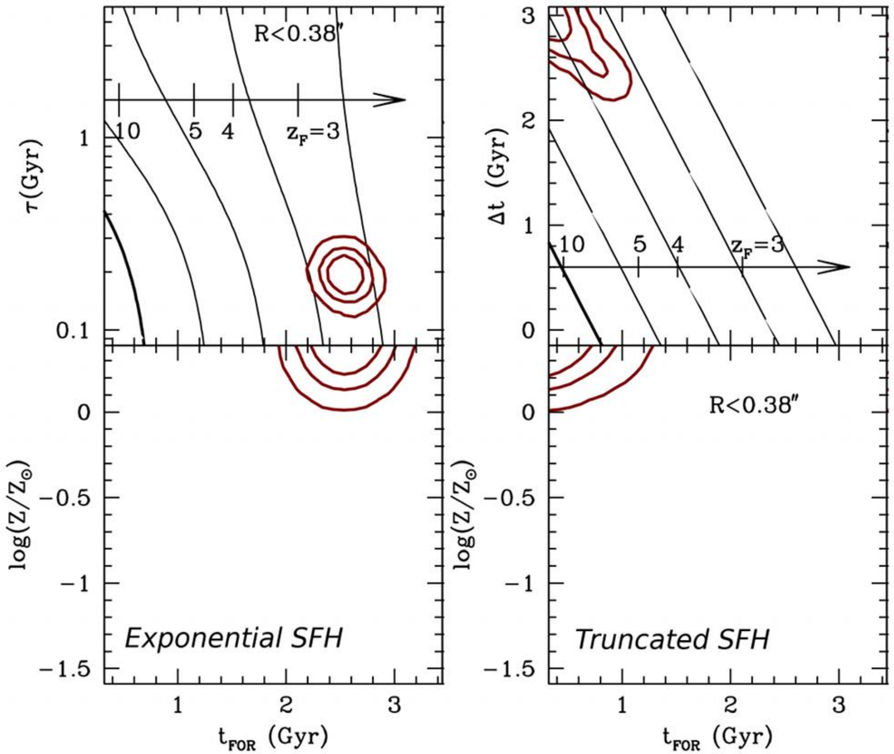
- Population synthesis models (B&C, Vazdekis, Conroy, Maraston, …)
- Spectral fitting (Starlight, StecMAP, ULySS, pPXF, FSPS, ...)
- Line strengths (Lick system and extensions, BMC, …)
- Dealing with degeneracies (Bayesian approach)

A model-driven approach

This example shows a model-fitting approach, defining a posterior based on a comparison of models and data. In this case, a z=1.9 massive galaxy (WFC3-IR SED) can be found to undergo two very different formation histories:

- Exponential, with a short (0.2Gyr), intense (1,400 $M_\odot$/yr) star formation
- Truncated, with an extended (2.7Gyr) formation at ~100 $M_\odot$/yr



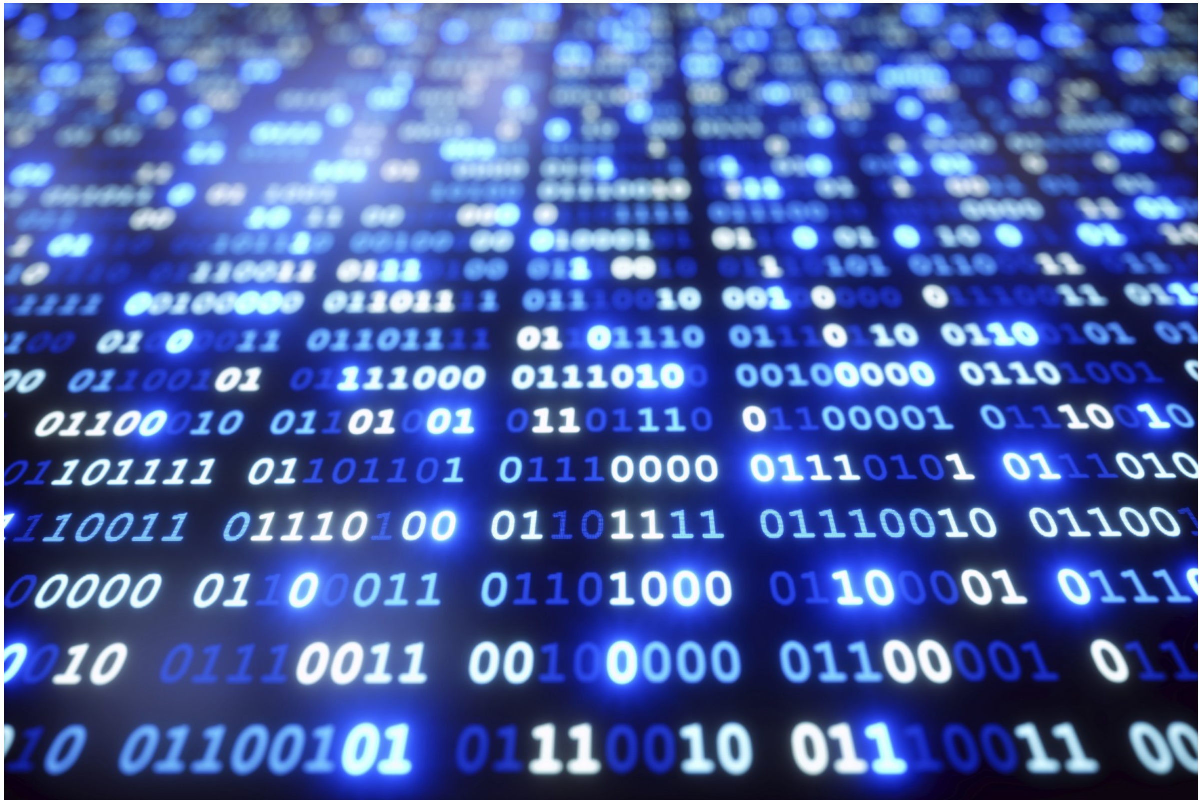


IF et al. (2012)

A data-driven approach

Change of tack: Instead of fitting data with models, we can use the statistical properties of the data to inform us about the properties of the sample.

Models are only needed *a posteriori*, i.e. once the analysis produces significant trends, we can go back to the models to “make physical sense” of these trends.



Galaxy spectra as the mother of all blind source separation problems

- One data-driven approach is to define an inverse problem: the observed photo+spectroscopic observations of galaxies are due to a (hopefully) reduced set of inputs (sources).
- This is an oversized Blind Source Separation (BSS) problem, where the observations (y) are the spectra and the sources (x) are the SSPs. The mixing matrix ($\mathcal{W}$) would provide the star formation history (modulo noise, ε).

$$\mathbf{y} = \mathcal{W}\mathbf{x}(+\epsilon) \longrightarrow \mathbf{x} = \mathcal{W}^{-1}\mathbf{y}$$

How can we invert the mixing matrix ?

Standard approach to BSS

Principal Component Analysis (Ronen +OL et al 1999; Madgwick +OL et al. 2002,2003; IF+OL et al., Wild & Hewett 2005, Rogers +OL et al 2007;2010)

Independent Component Analysis (Kaban et al 2005)

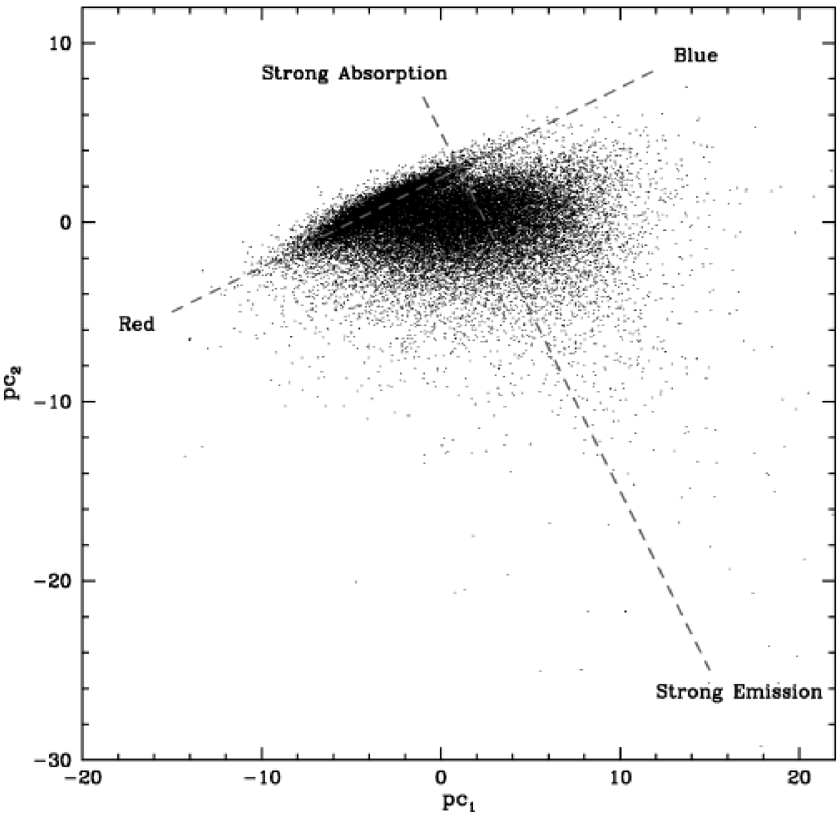
Fisher information matrix (Heavens, Jimenez & OL 2000, Panter et al. 2003)

Artificial Neural Networks (Folkes, OL, Maddox 1996)

Factor Analysis (Nolan et al 2007)

Clustering methods (Sánchez-Almeida et al. 2012)

Madgwick et al. (2002)



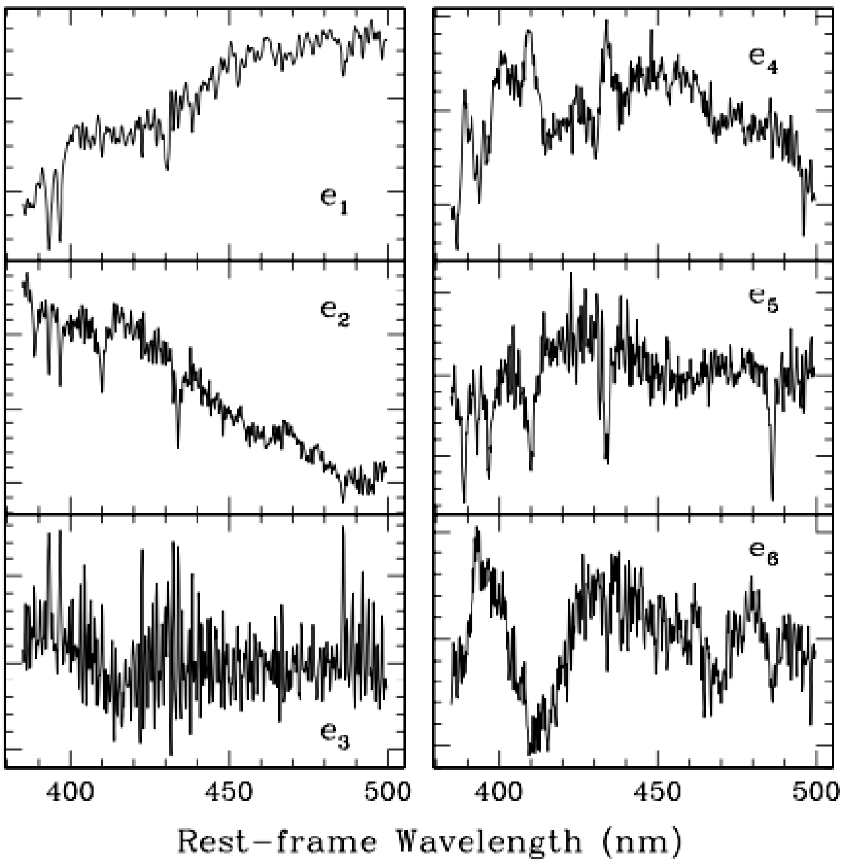
An example: PCA

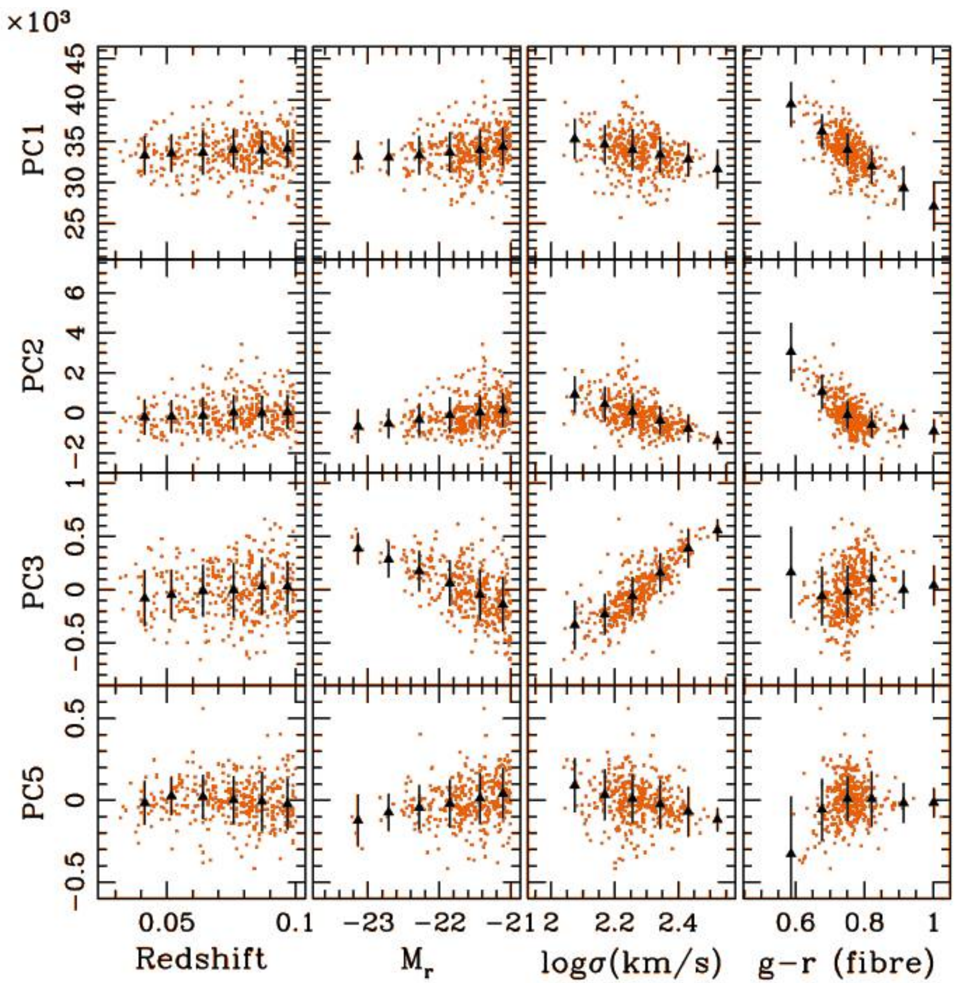
Use the large volume of high quality spectra from SDSS to produce a clean sample to analyse along the lines of BSS.

PCA is a variance-based method, so a carefully defined, “simple” sample is best.

Early-type galaxies: mostly quiescent populations, small contribution from recent SF and dust.

Rogers (+IF, OL) et al. 2007



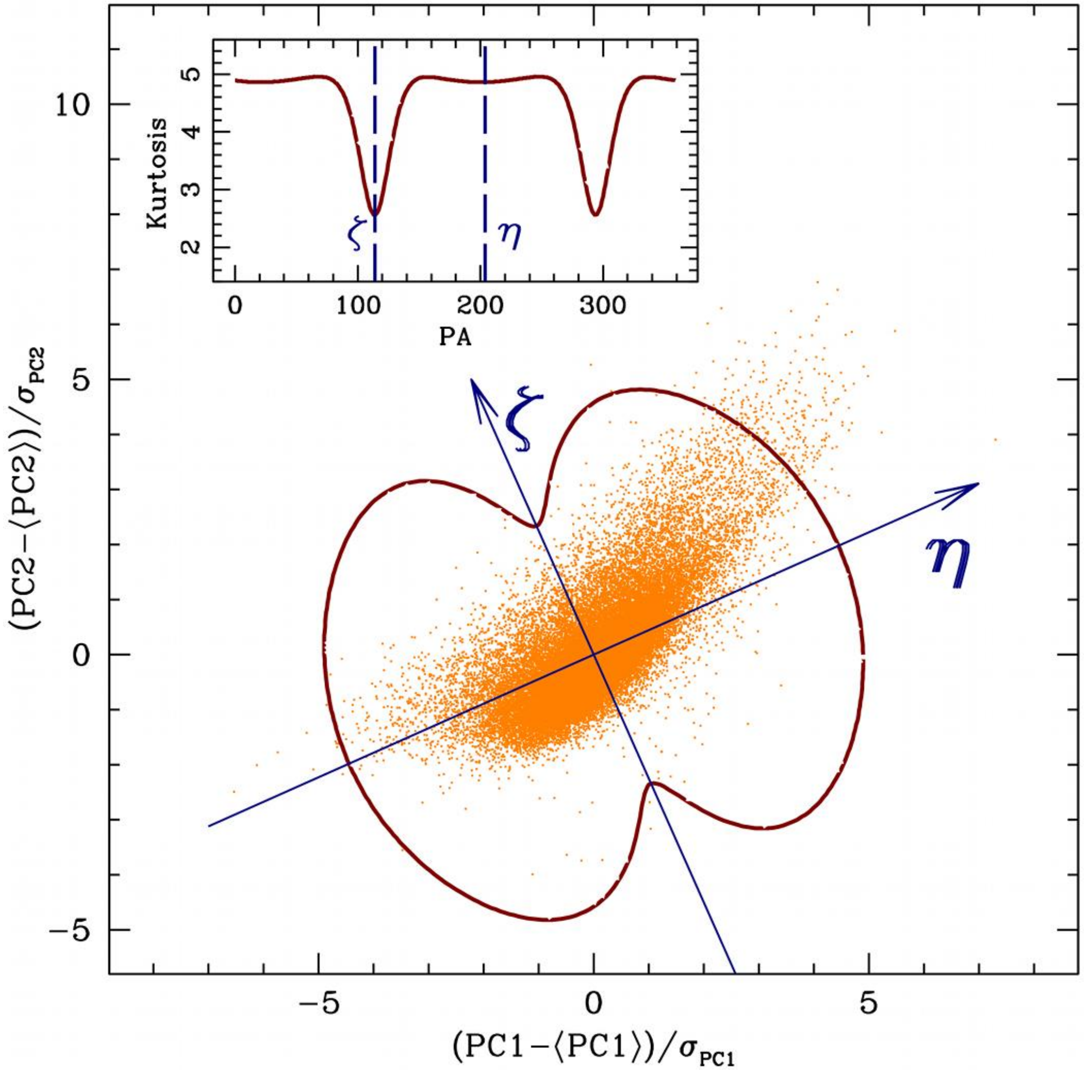


An ICA-inspired definition

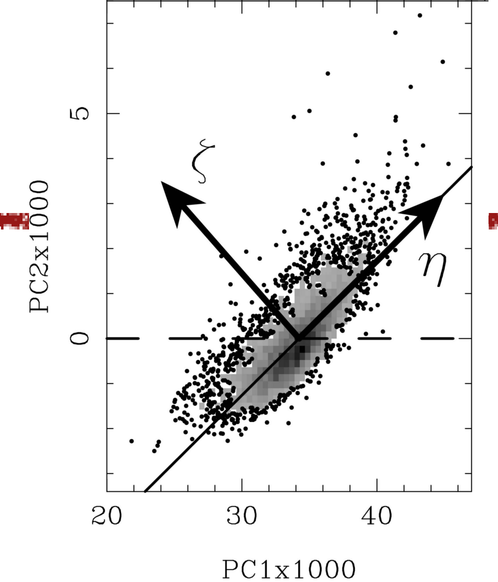
According to the central limit theorem, subsequent mixtures of random variables become more Gaussian. Hence, extracting information based on non-Gaussianity can yield more fundamental components.

In this figure, the definition of the η and ζ parameters is motivated by a search for the extrema of kurtosis (i.e. non-Gaussianity) in the distribution of PC1-PC2 projections.

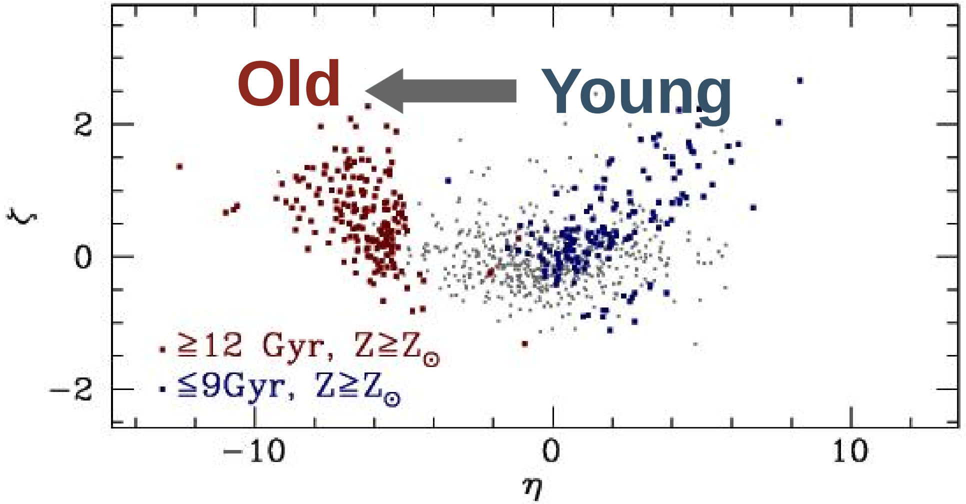
IF 2011

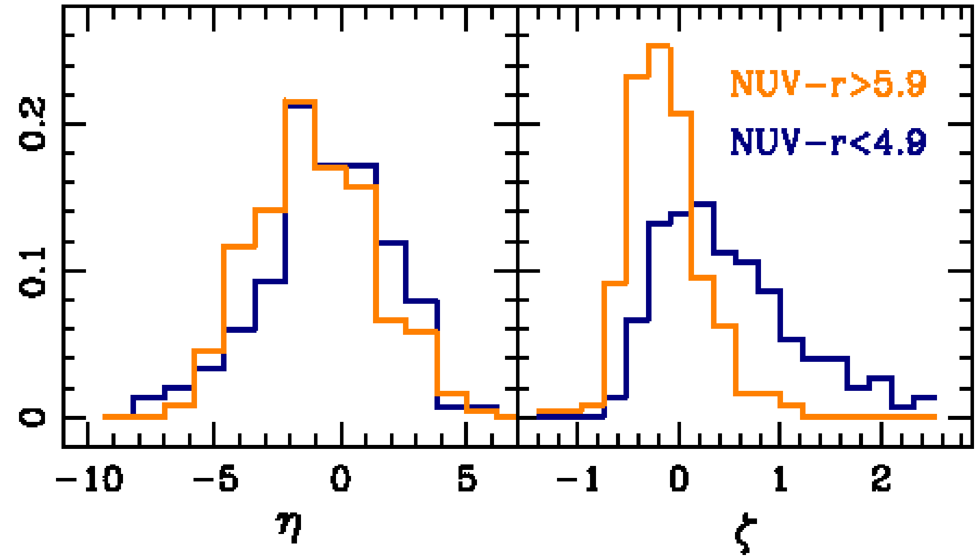


Meaning of the rotated PCA components (η,ζ)



Rogers (+IF, OL) et al. 2007



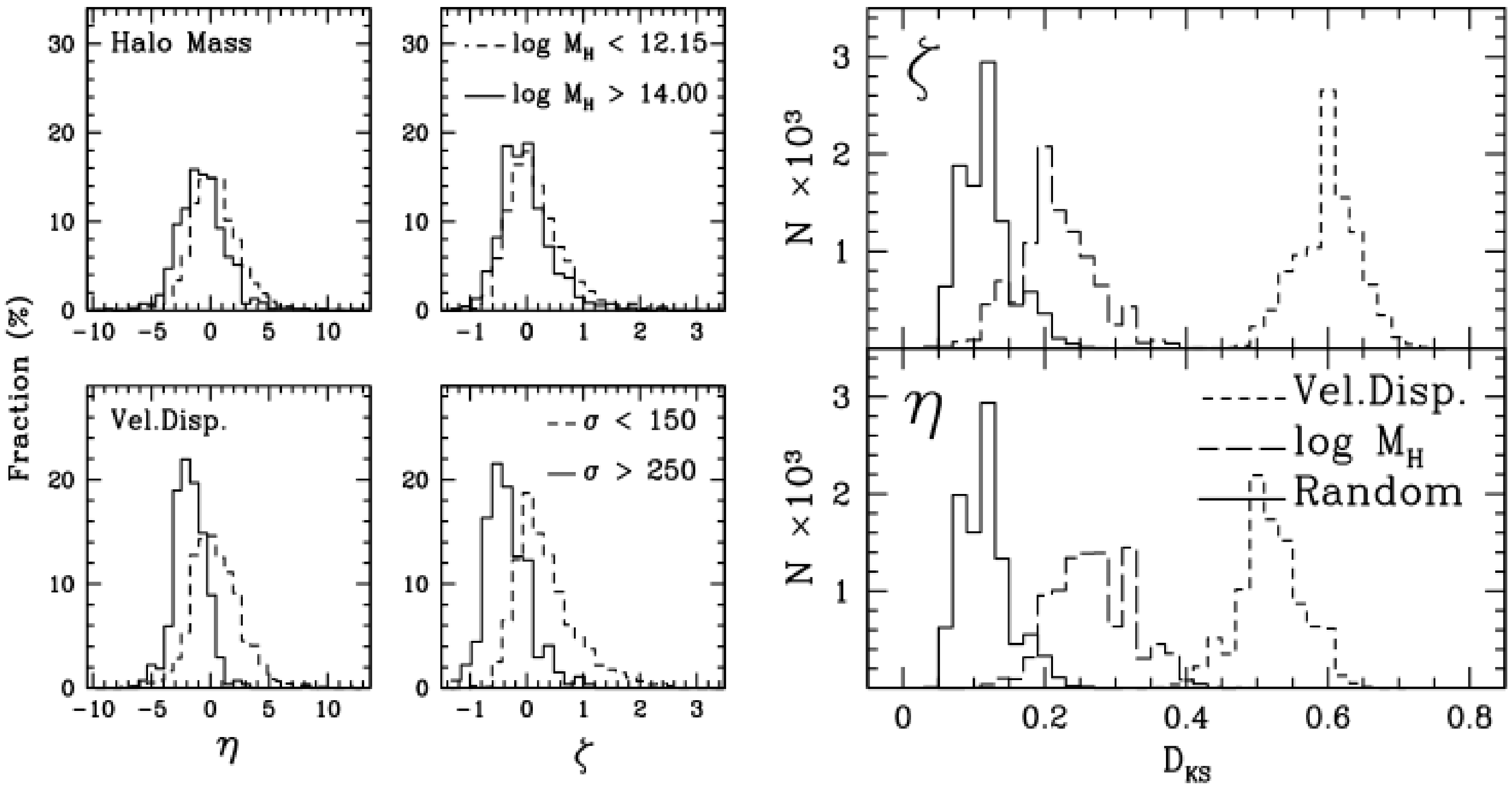


η: Average Age

ζ: Recent Star Formation

An application:
The role of environment in the formation of ETGs

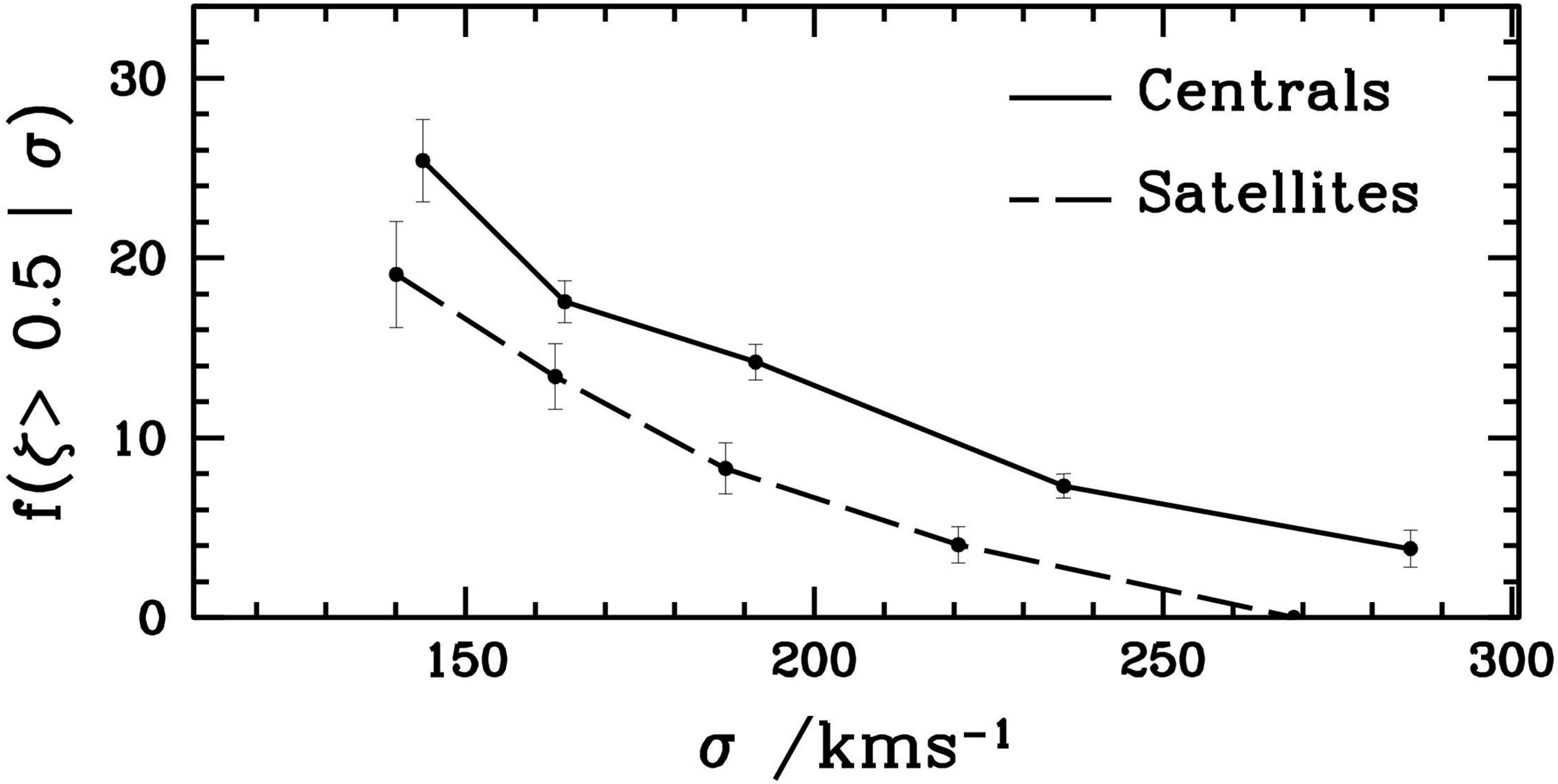
Volume limited (z<0.1) sample of SDSS ETGs (~7,000)
Groups catalogue of Yang et al. (2007)
The most important global observable controlling the SFH is velocity dispersion



Rogers (+IF, OL) et al. 2010

Environment: Centrals v Satellites

At fixed velocity dispersion, central galaxies present a higher amount of recent star formation than satellite galaxies.



Rogers (+IF, OL) et al. 2010

An information theory approach

One can give different interpretations to galaxy spectra:

1) A wavelength-dependent function (standard approach, model fitting)

2) A multi-dimensional vector (PCA/ICA/FA)

3) A probability distribution function: We observe a galaxy (g) with a photon counter. The (conditional) probability that a given photon has wavelength λ is $\mathcal{P}(\lambda|g)$

The latter allows us to consider the analysis of galaxy spectra as an information theory problem.

Ofer already explored this territory with Slonim+ (2001) as we will see next

An information theory approach

Let us assume we have a family of galaxies {g} with spectra P(λ|g)

The entropy associated to the galaxy sample is $H(G)=-\sum p(g)\log p(g)$, where we assume that p(g) is uniform over the sample.

If we take into account the information provided by the spectra (the "value" of the galaxy sample G at the wavelengths in some Λ), we can define a conditional entropy:

$$H(G|\Lambda) = -\sum_{\lambda} p(\lambda) \sum_{g} p(g|\lambda) \log p(g|\lambda)$$

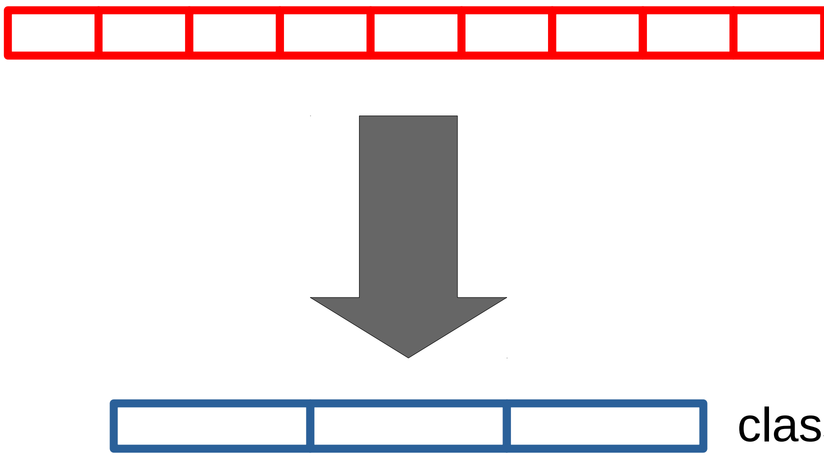
The amount of reduction of uncertainty is the mutual information:

$$I(G;\Lambda) \equiv H(G) - H(G|\Lambda) = \sum_{g,\lambda} p(g)p(\lambda|g) \log \frac{p(\lambda|g)}{p(\lambda)}$$

We now bin the G sample into a smaller set of classes C, with the corresponding loss of information: $I(C;\Lambda) \leq I(G;\Lambda)$: passing the galaxy data through an ***information bottleneck***.

The goal is to minimise the number of classes, keeping $I(C;\Lambda)/I(G;\Lambda)$ as high as possible

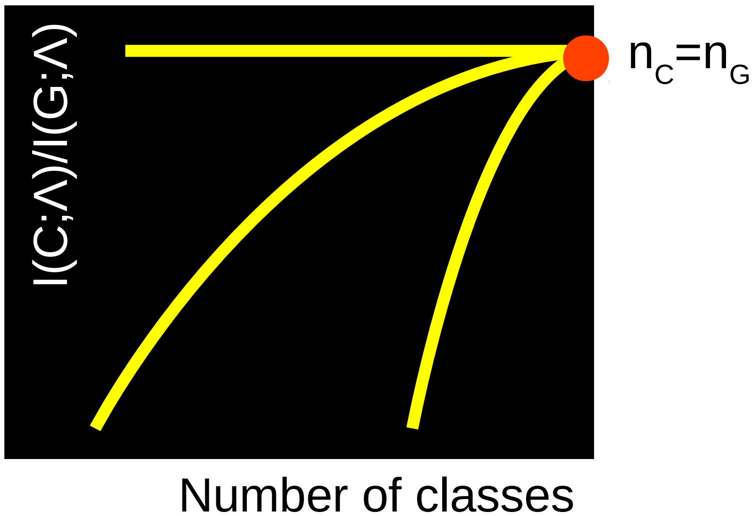
Producing the IB classes



galaxies
G={g_1,g_2,…}

classes
C={c_1,c_2,…}

Agglomerative procedure starting from the trivial $n_C=n_G$, merging spectra into classes and computing the loss of information.



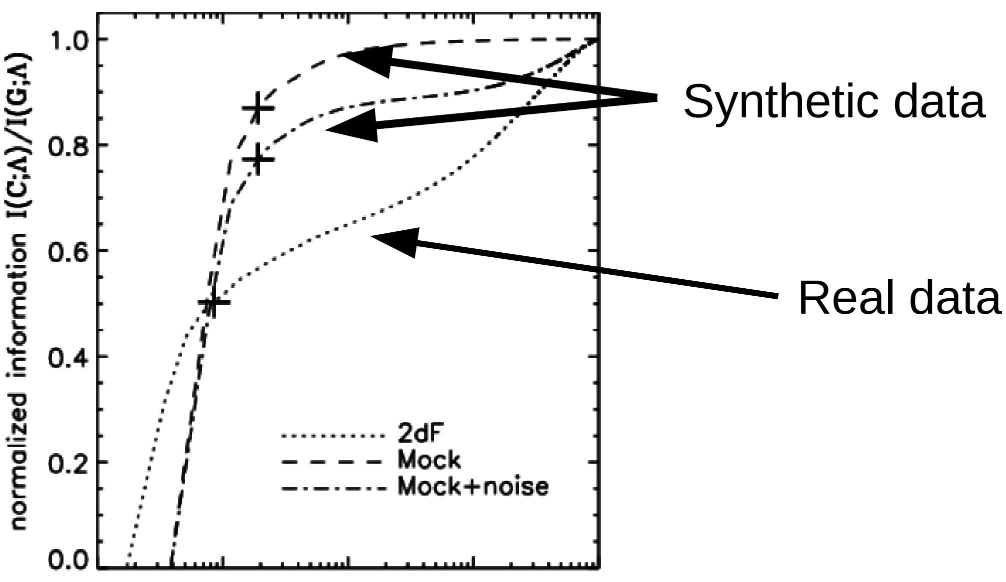
Cross-entropy (Kullback-Leibler divergence)

When classifying a large dataset, one can define the amount of information lost when the data are arranged into a small set of classes. This is called cross-entropy, or KL divergence, D_{KL}.

$$D_{\mathrm{KL}}(g \in G; c \in C)_\Lambda = \sum_\lambda p(\lambda|g) \log \frac{p(\lambda|g)}{p(\lambda|c)}$$

If D_{KL} is very small, we then know that the system can be simplified into a reduced number of "sources" from which the observed data originate.

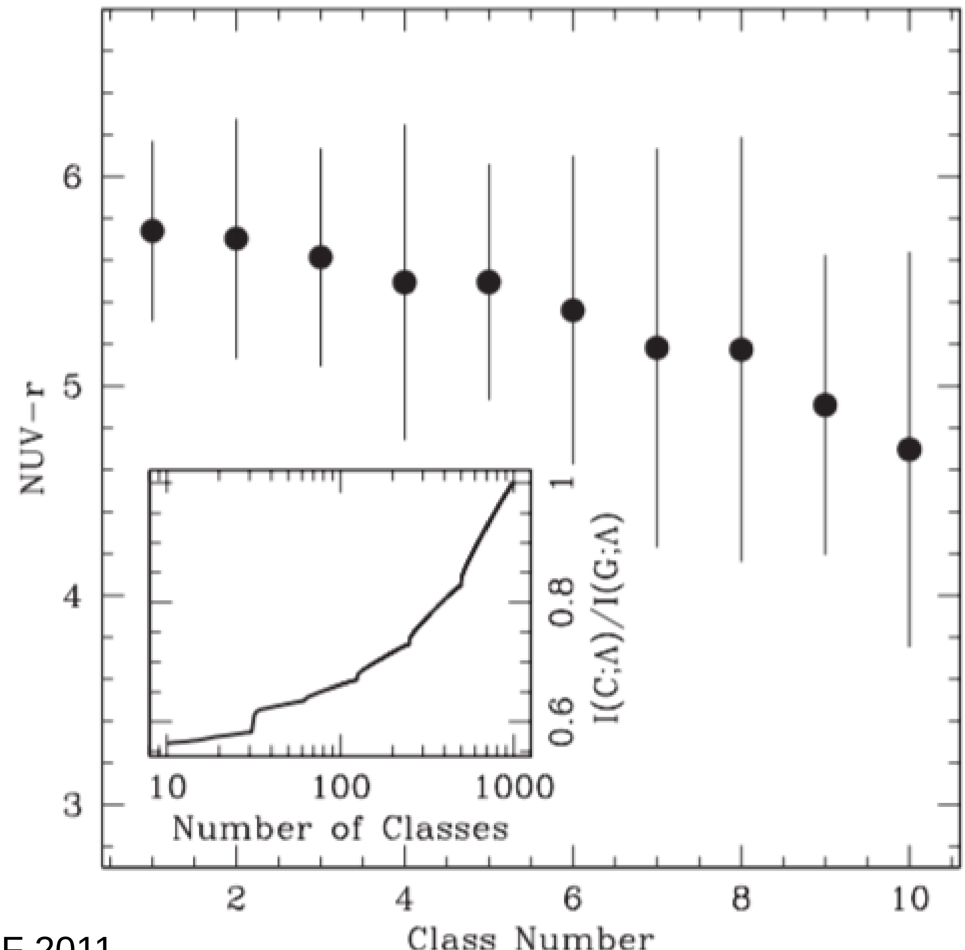
In our case, can we classify a large set of galaxy spectra (say from the 2dFGRS) into a small number of classes?

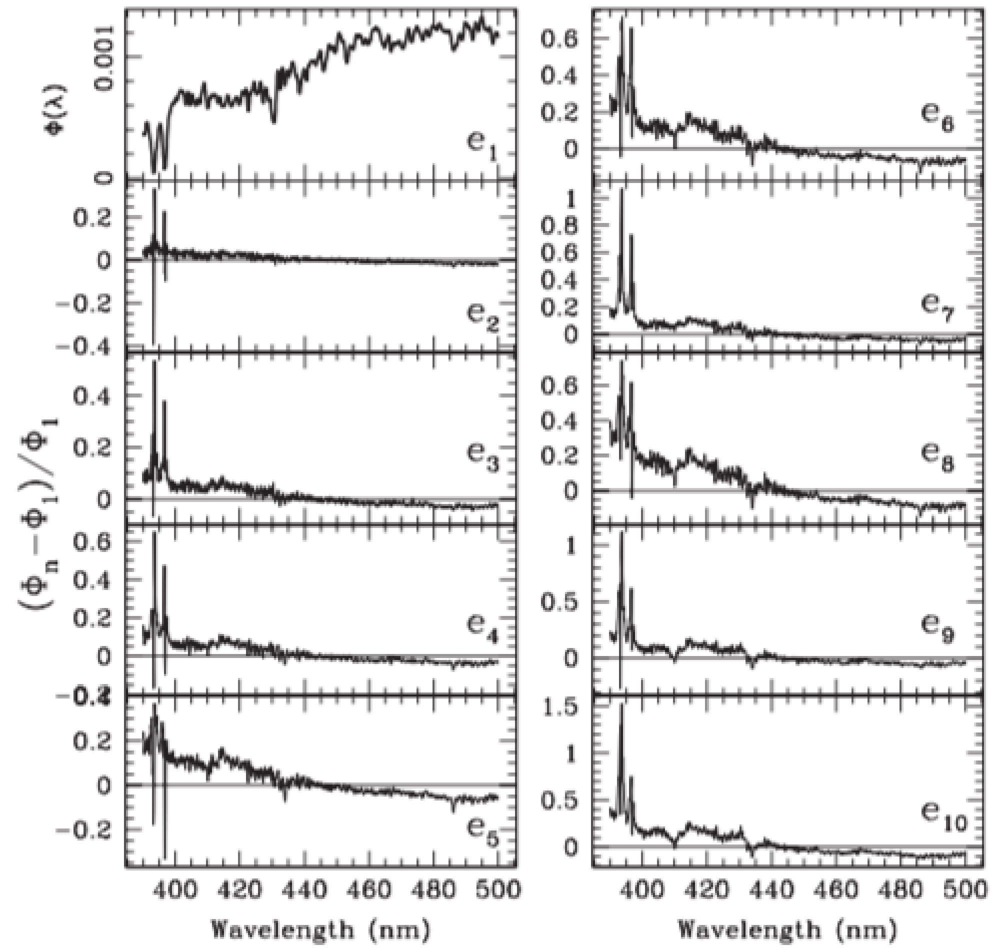


Slonim (+OL) et al. 2001

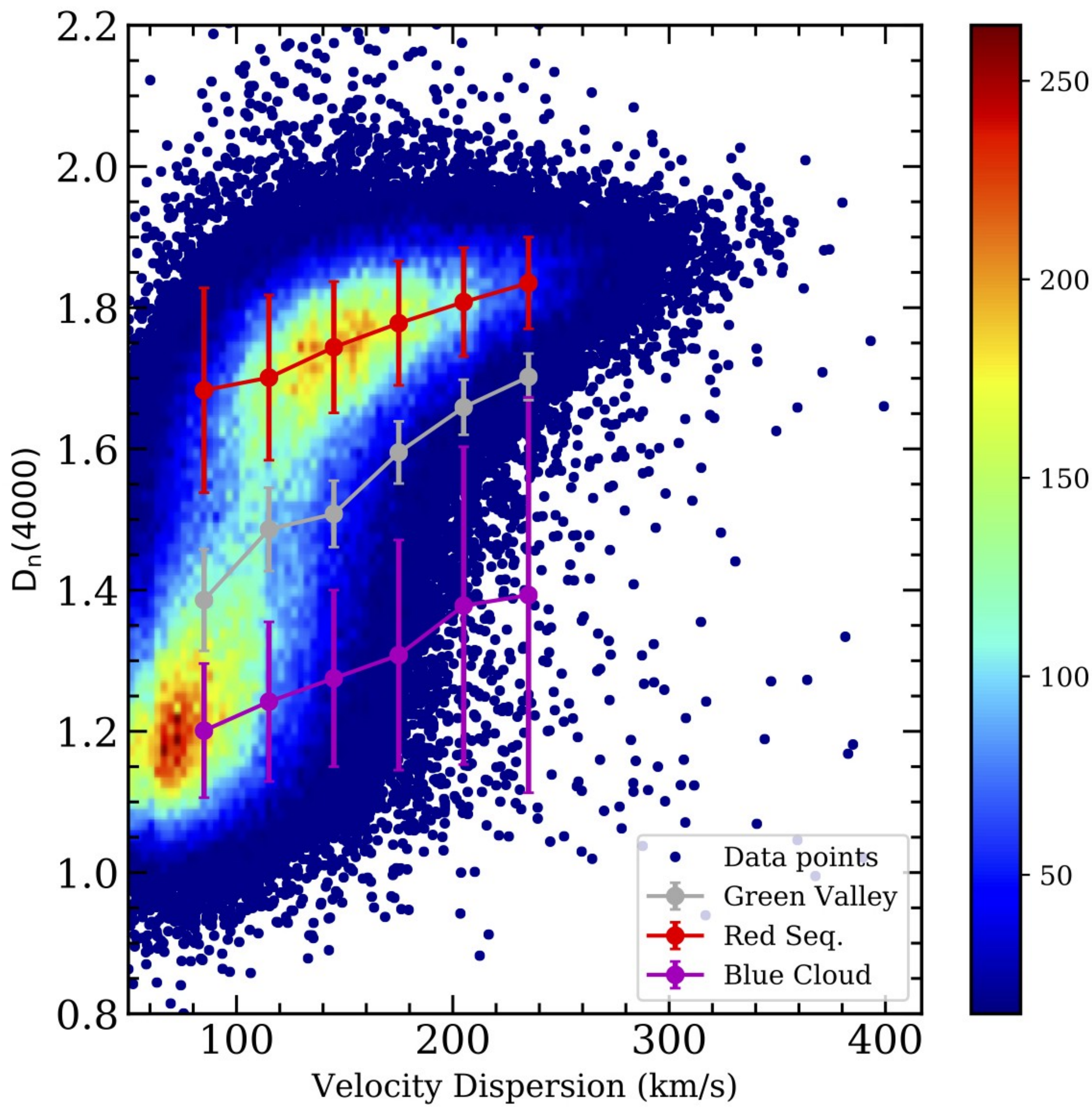
Cross-entropy (Kullback-Leibler divergence)

IB applied to SDSS early-type galaxy spectra (compared a posteriori with NUV-optical colours)



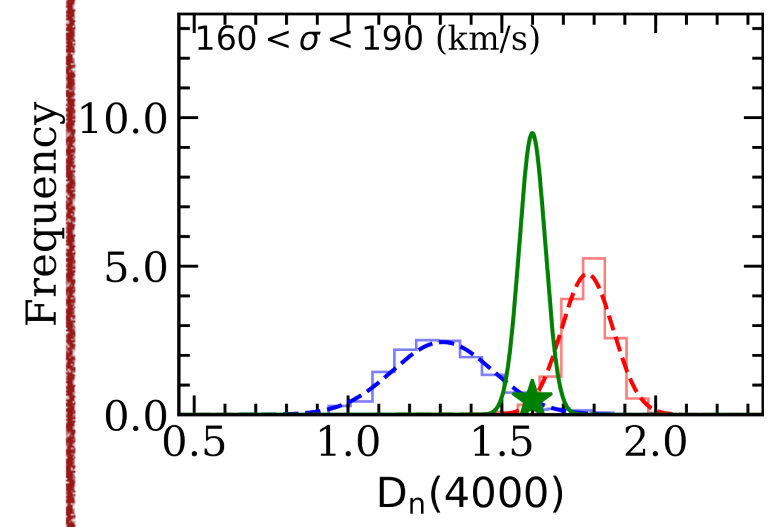


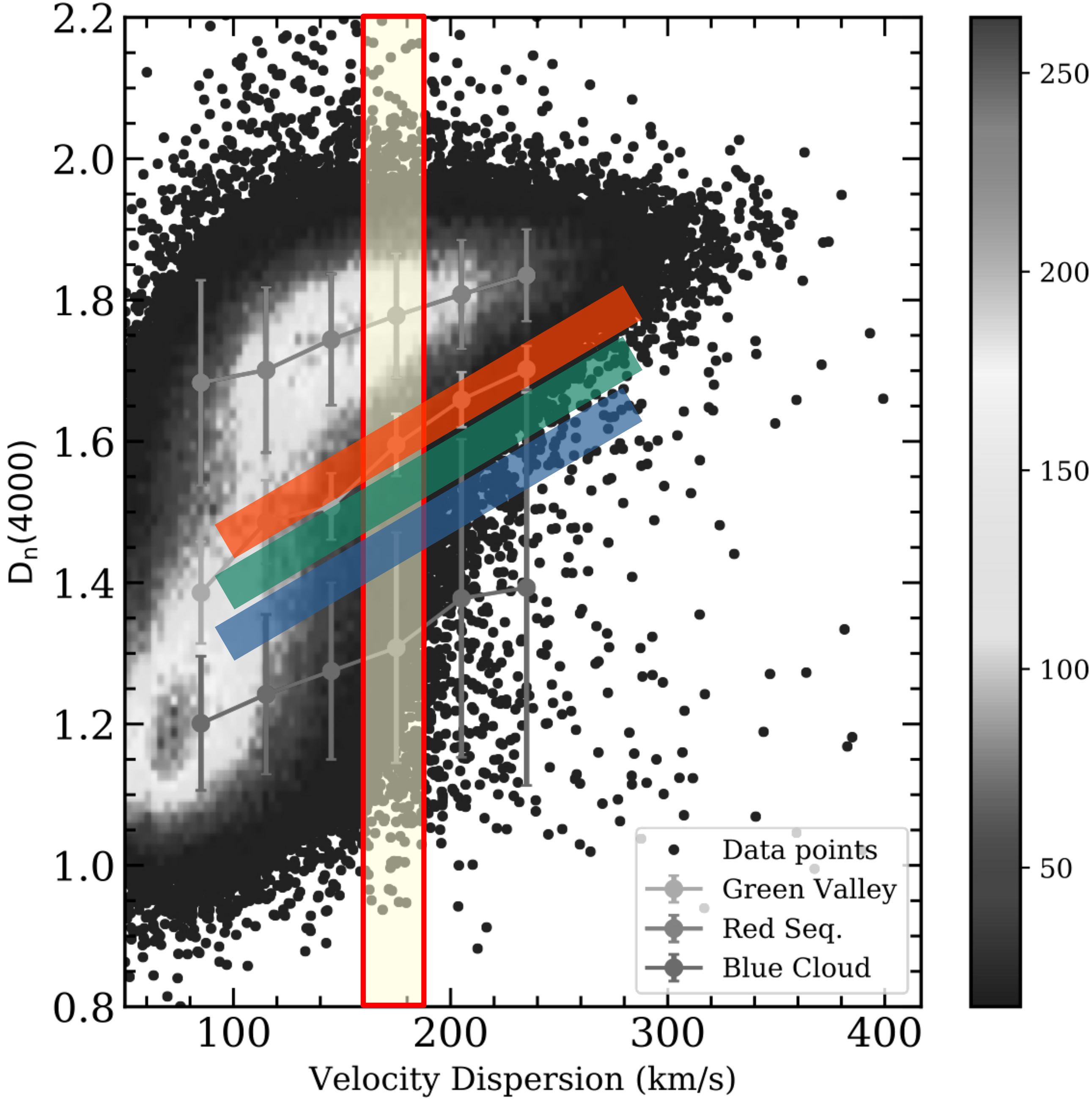
IF 2011



Angthopo, IF & Silk (2019)

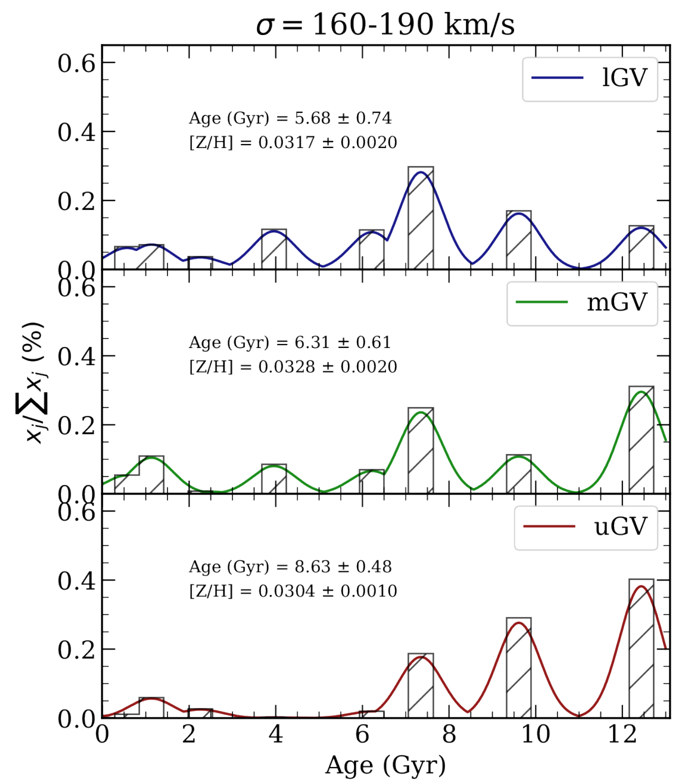
A test case: Green Valley galaxies



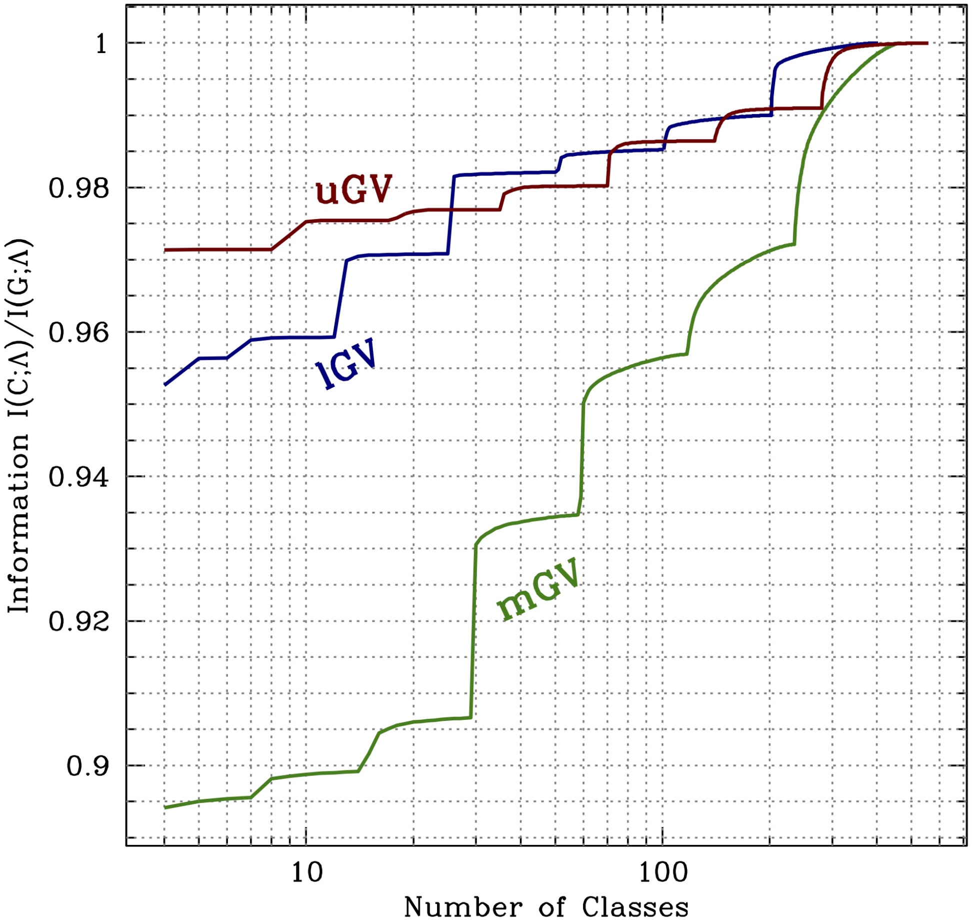


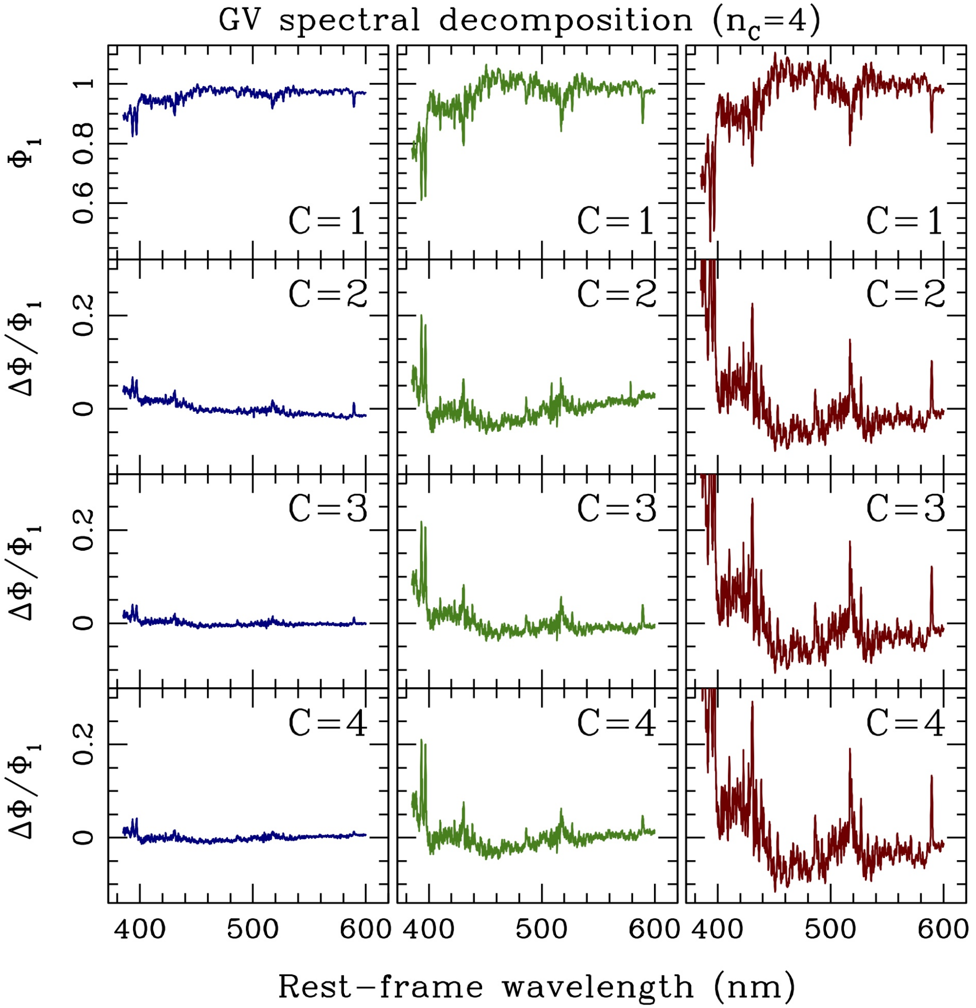
Angthopo, IF & Silk (2019)

A test case: Green Valley galaxies

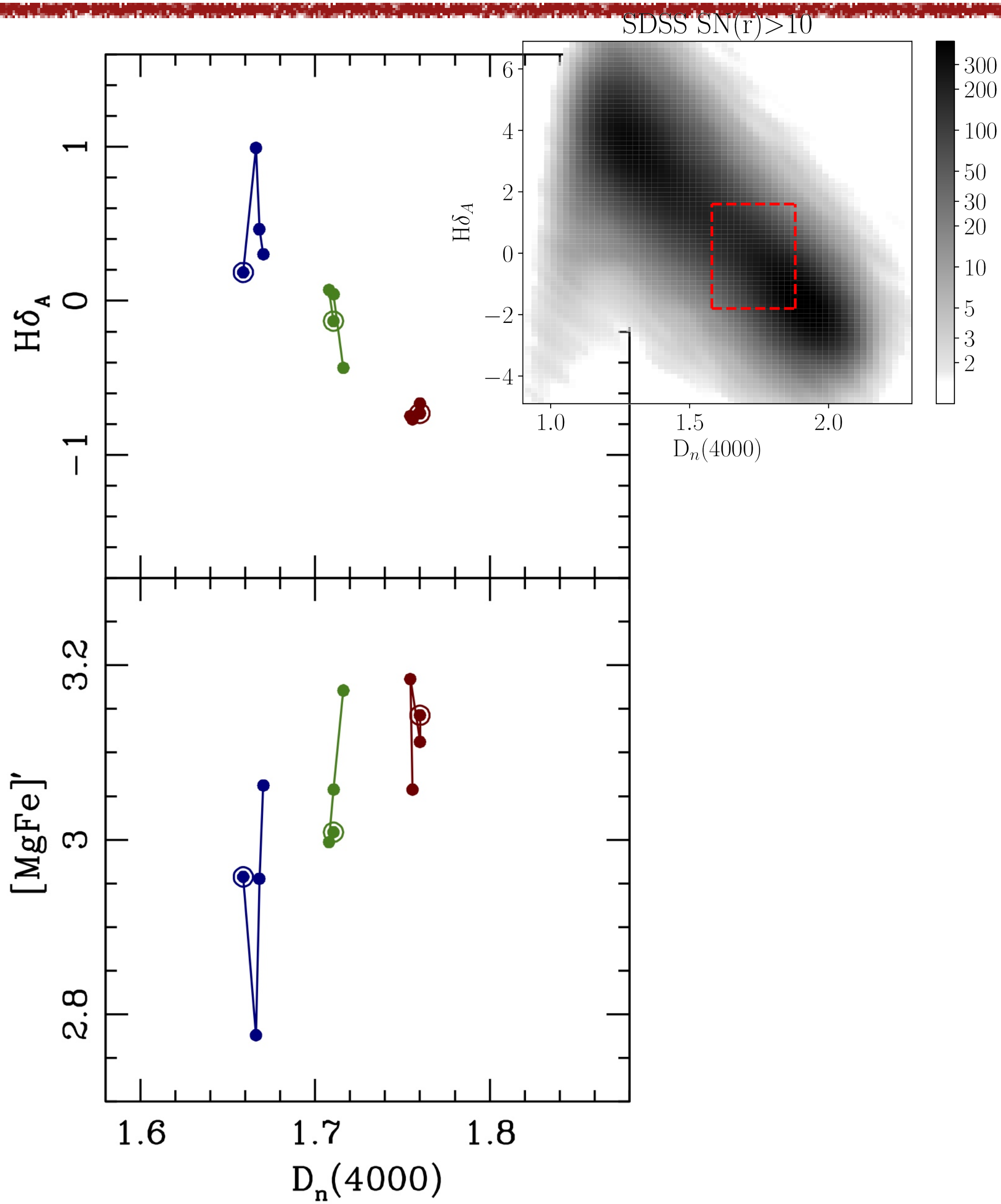


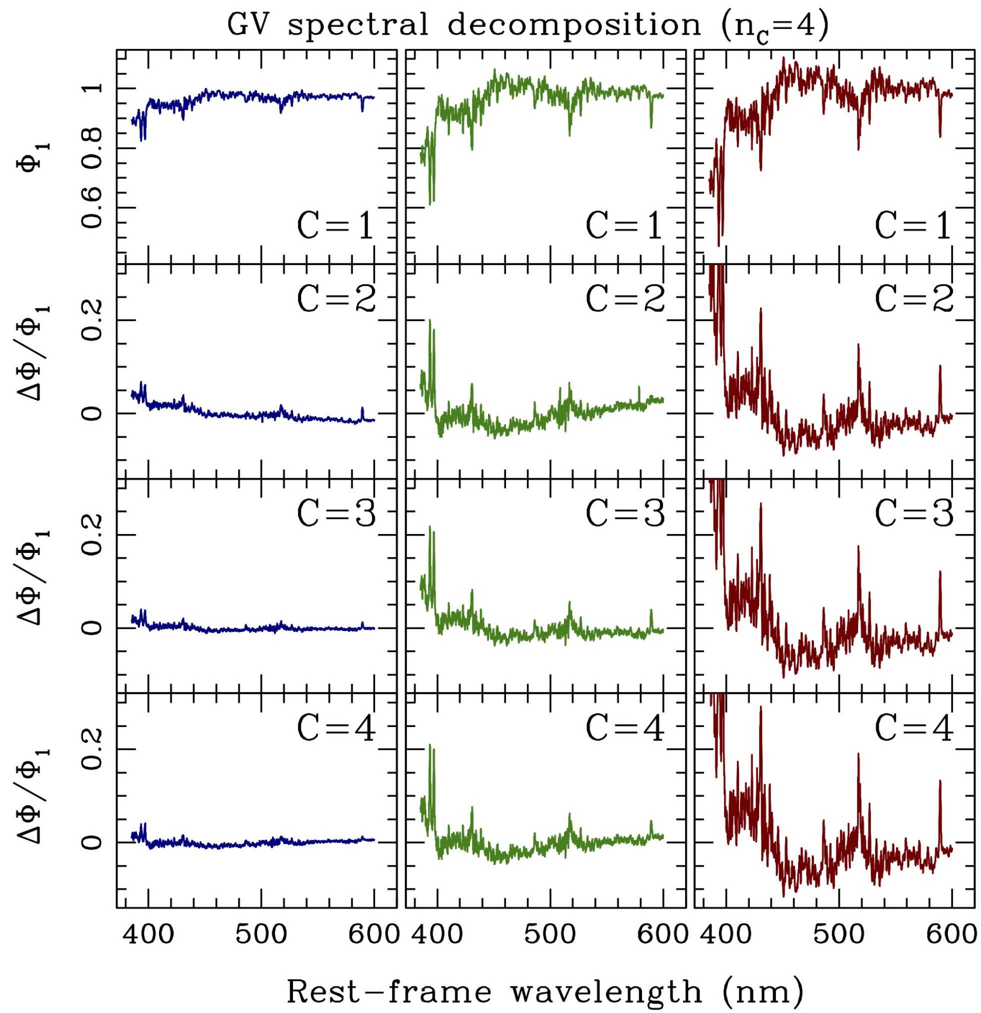
Passing the Green Valley through the IB





Passing the Green Valley through the IB





Stellar populations and Big Data

- The availability of large volumes of high quality spectroscopic data allows us to explore stellar populations in galaxies in a “data-driven” way.
- The high level of correlation makes this a much harder problem with respect to equivalent 1D streams such as time series, typically used, e.g. in audio signals, financial data, medical imaging.
- Machine Learning, multivariate analysis and information-based methods represent approaches that can mitigate the shortcomings of population synthesis modelling.
- Ofer has produced pioneering work in this field many years ago.
- Happy Birthday Ofer !! יום הולדת שמח !!