Deciphering the hidden side of galaxies with full-spectral fitting

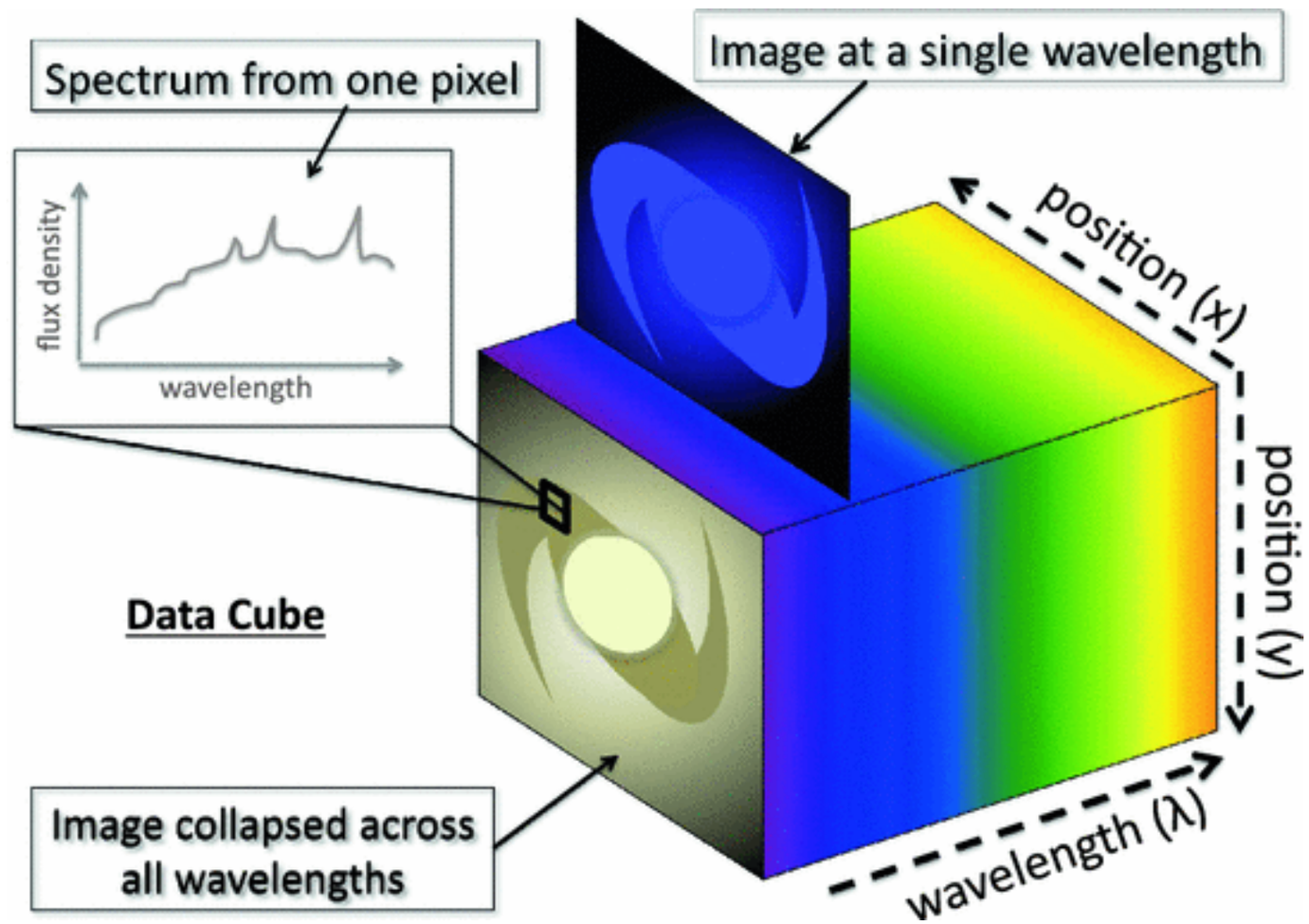
J. Falcón-Barroso

ML meeting - Ringberg Castle (9 -13 December 2019)

Learning from the fossil-record

- Galaxy spectra holds fundamental info about:
  - The way the stars move (stellar kinematics)
  - Their chemical composition (stellar populations)
- Both tracers have retain dynamical and population memory of galaxy evolution
- Extraction of those parameters is poised with degeneracies that are not necessarily easy to control

The Data


Spectrum from one pixel
flux density
wavelength
Image at a single wavelength
position (x)
position (y)
wavelength (λ)
Data Cube
Image collapsed across
all wavelengths

How do we retrieve that info?

Populations

weights

convolution

$$G_{model}(\lambda) = \sum_{1}^{K} [w_k \cdot T_k(\lambda)] \star B$$

Galaxy model spectrum

Library of K template spectra

LOSVD

Kinematics

Residuals

$$r_n = \frac{G_{model}(\lambda_n) - G(\lambda_n)}{\Delta G(\lambda_n)}$$

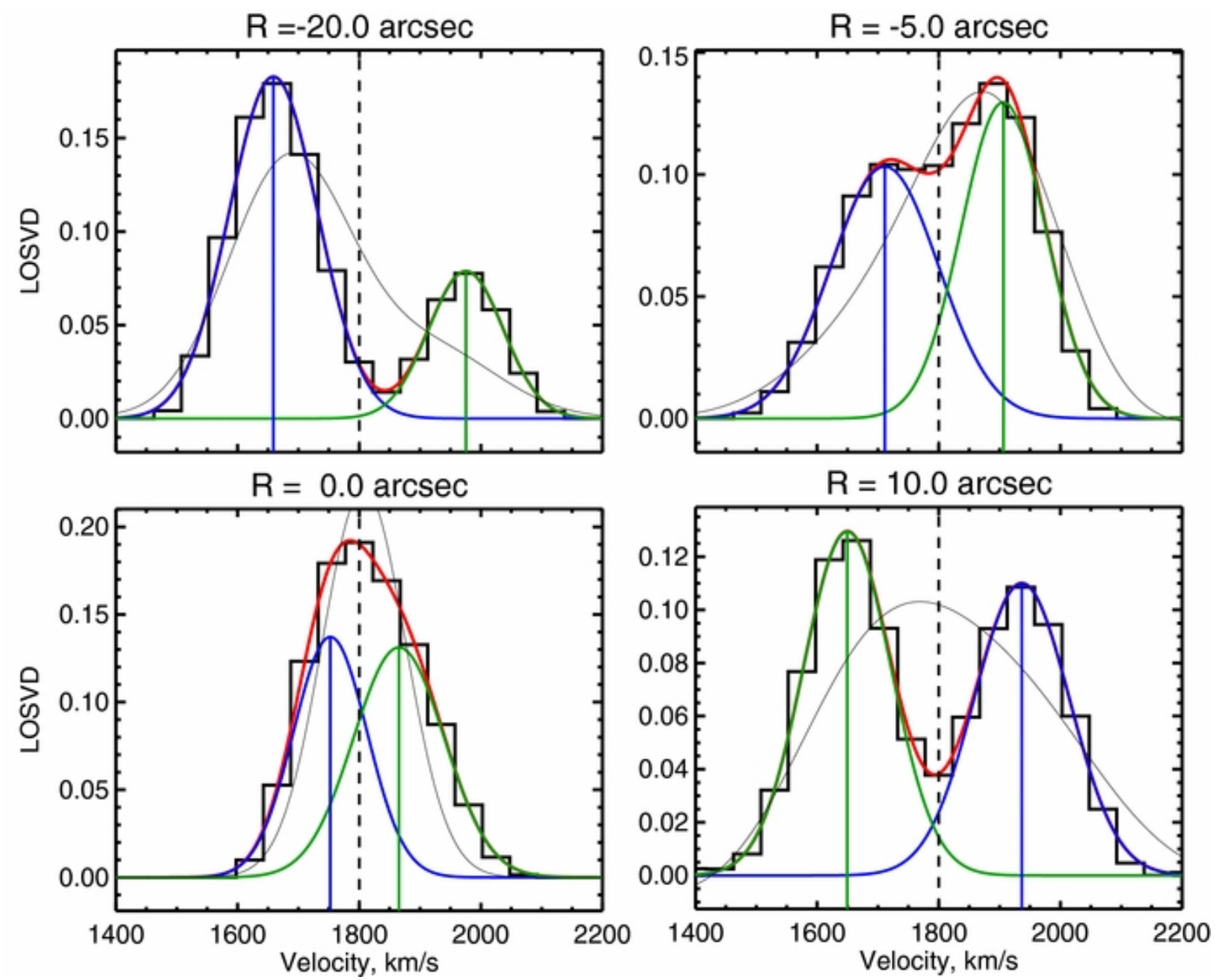
Minimisation

$$\chi^2 = \sum_{1}^{N} r_n^2$$

The LOSVD

The LOSVD

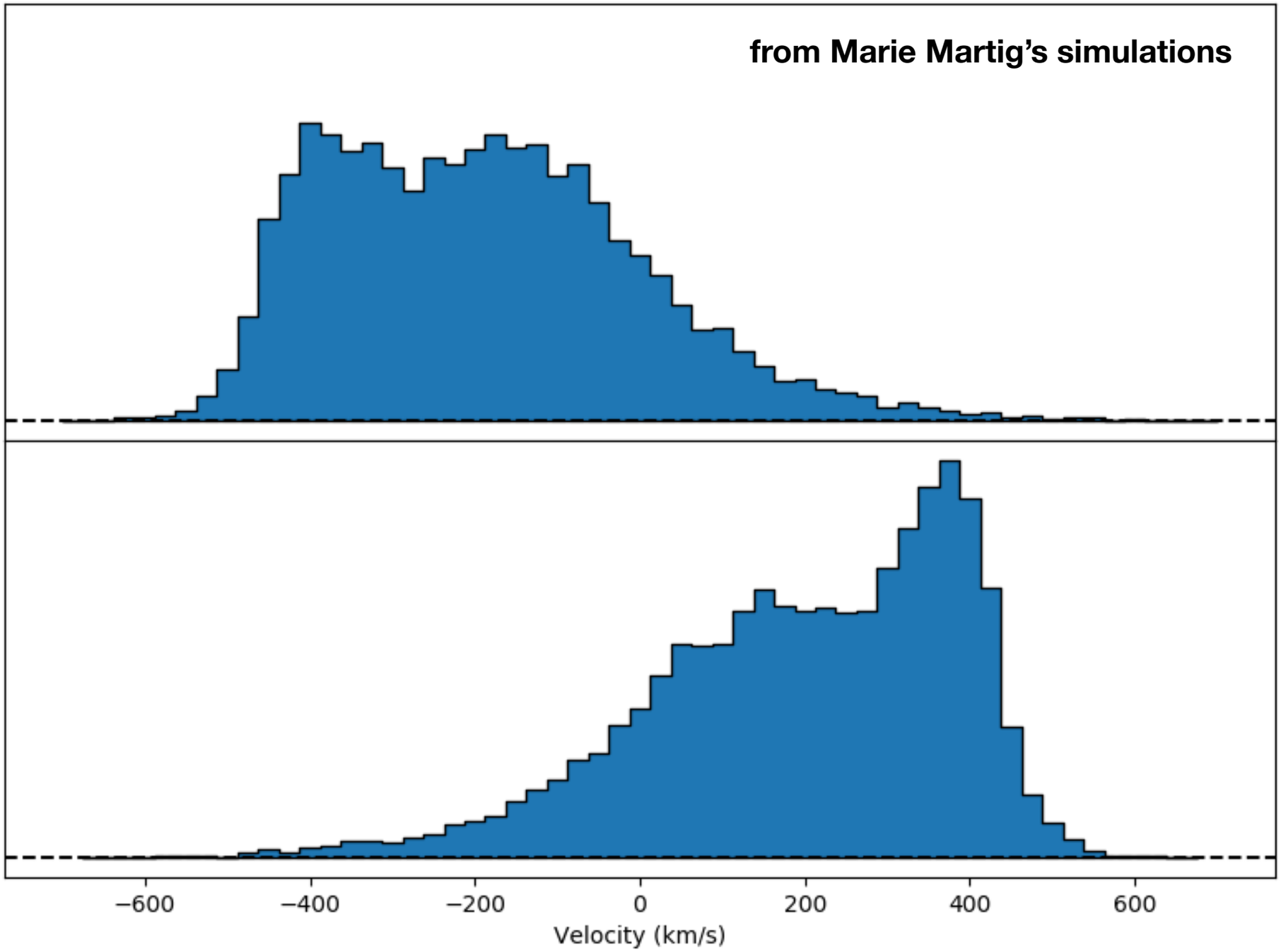
- Observations can show complex LOSVDs…



Katkov et al. (2013)

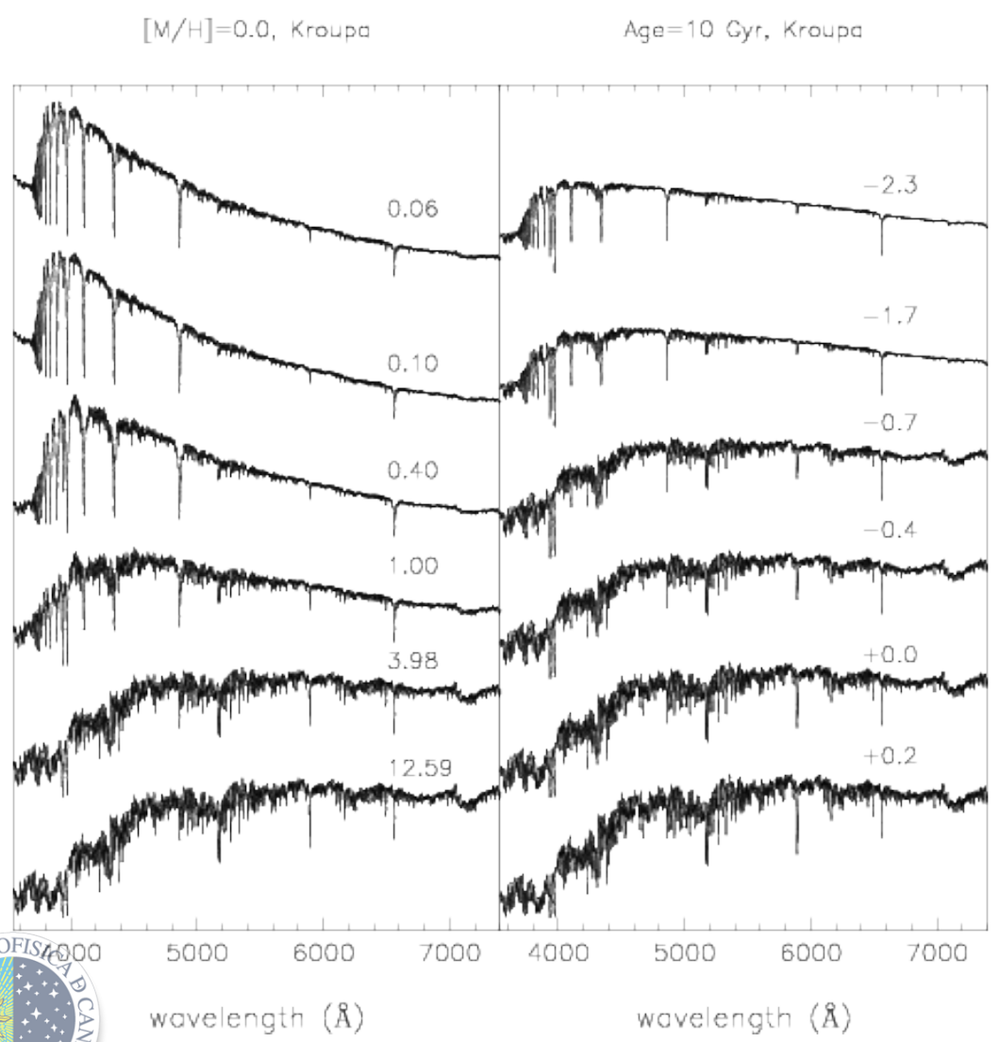
The LOSVD

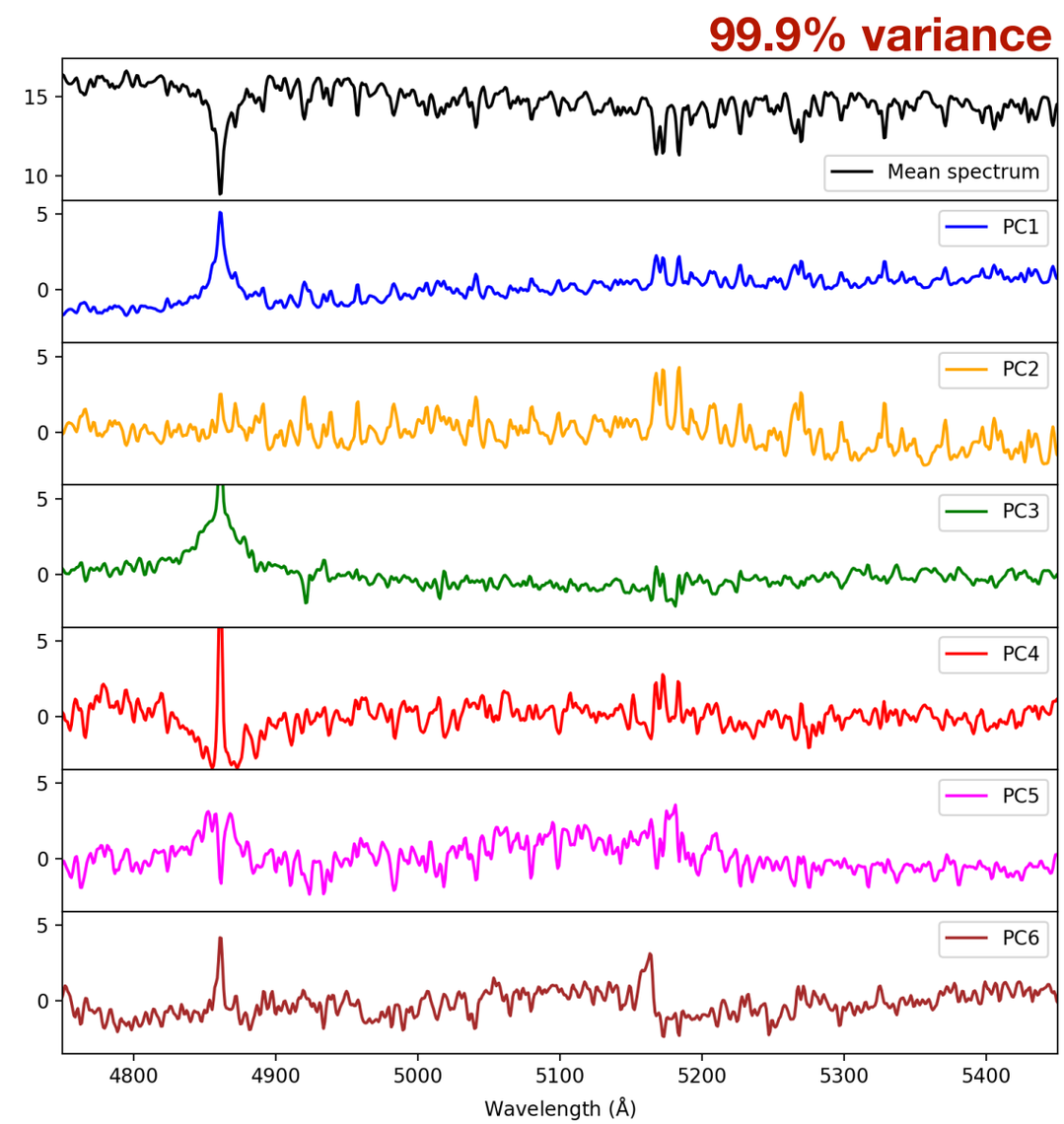
- And numerical simulations too!



LOSVD recovery

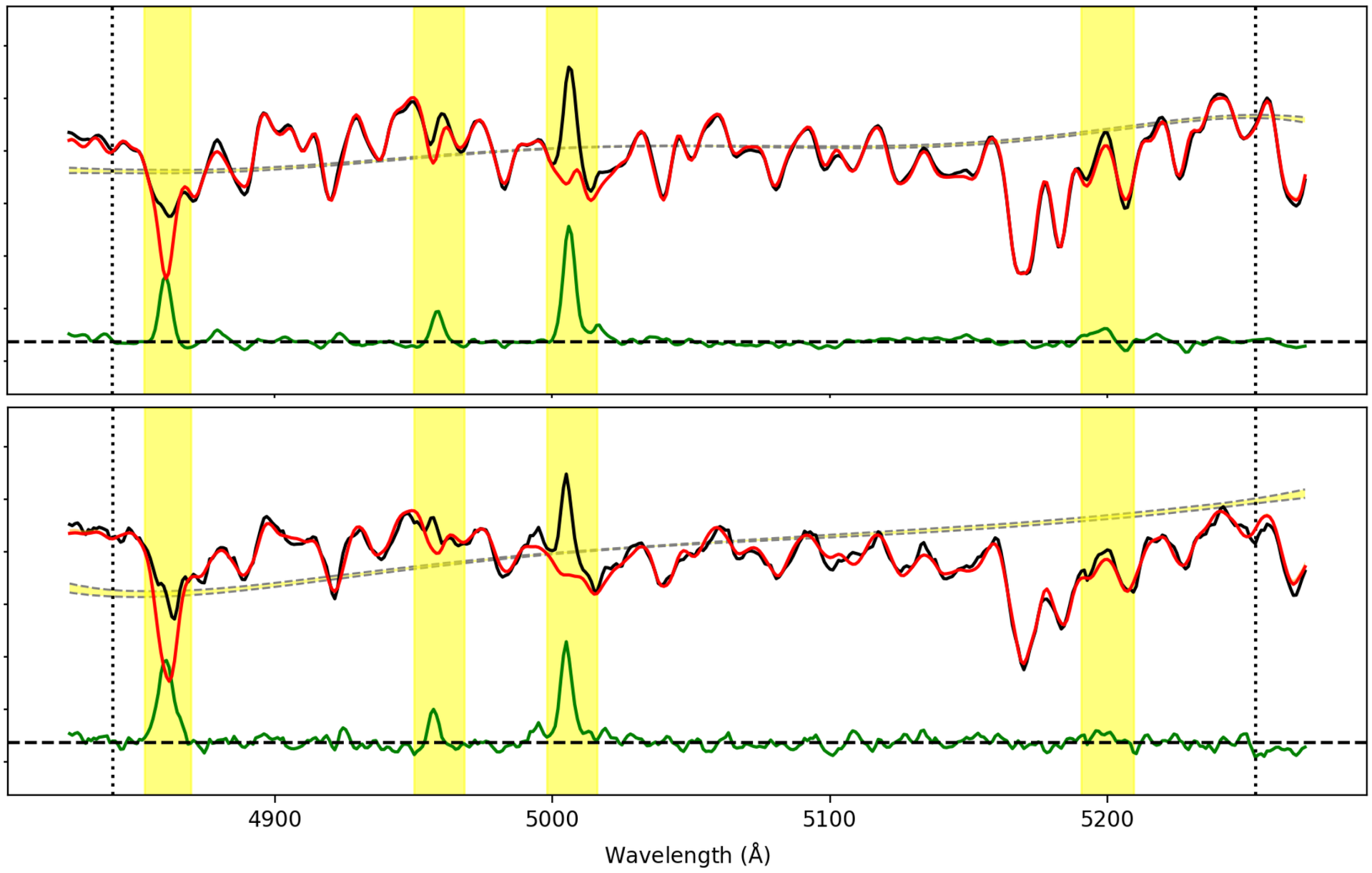
- Difficult to characterise the uncertainties in the recovered LOSVD
- Bayesian frameworks deal with this, but still slow because number of templates can be large
- Typical template libraries contain ~1000 spectra with 5000 pixels
- This is extremely time consuming for any sampler
- Dimensionality reduction of the input template library helps (e.g. PCA)



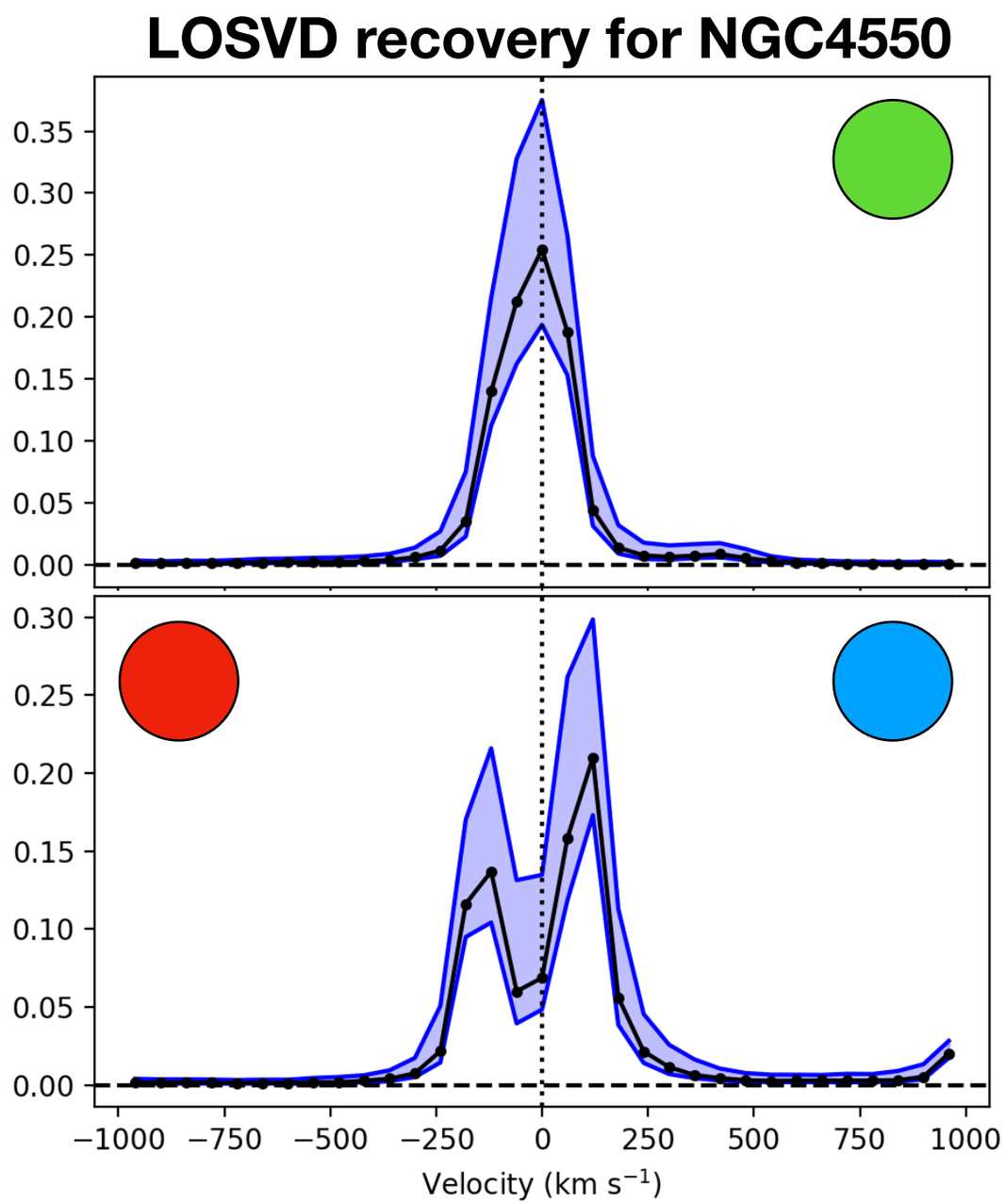


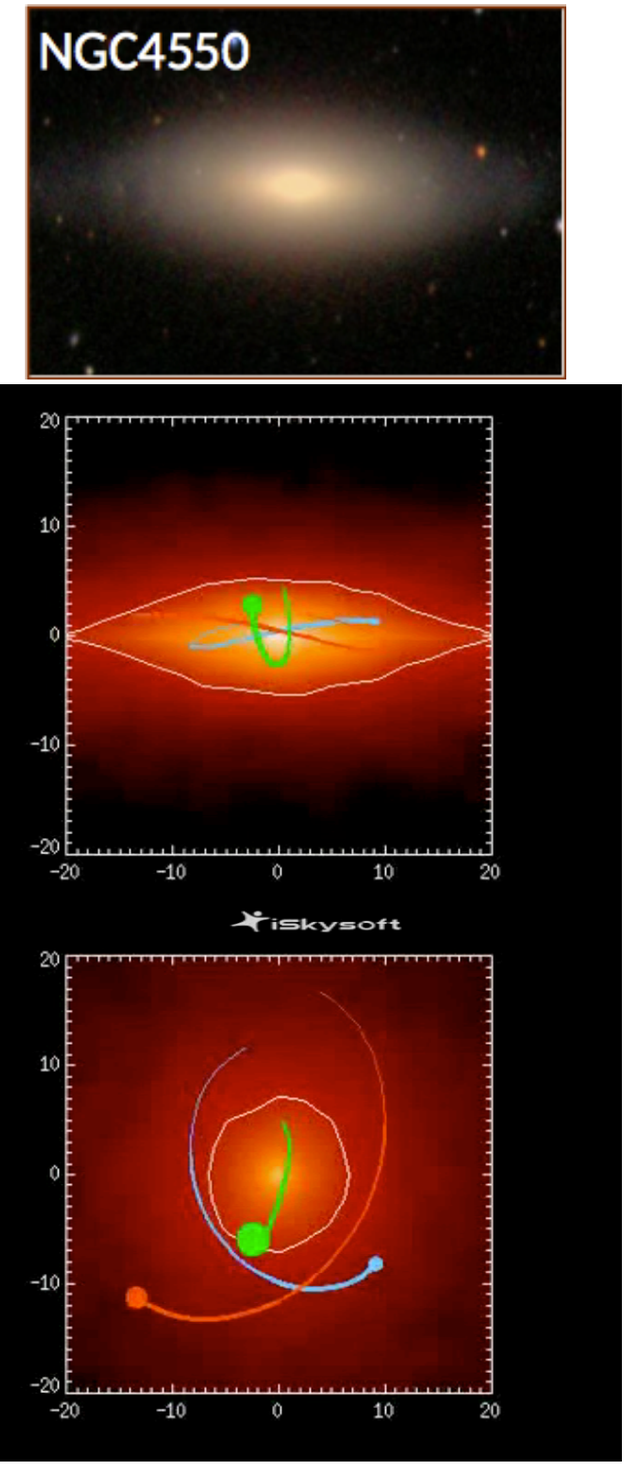
BAYES-LOSVD

- A Bayesian non-parametric recovery of the LOSVD using Stan http://mc-stan.org/



BAYES-LOSVD





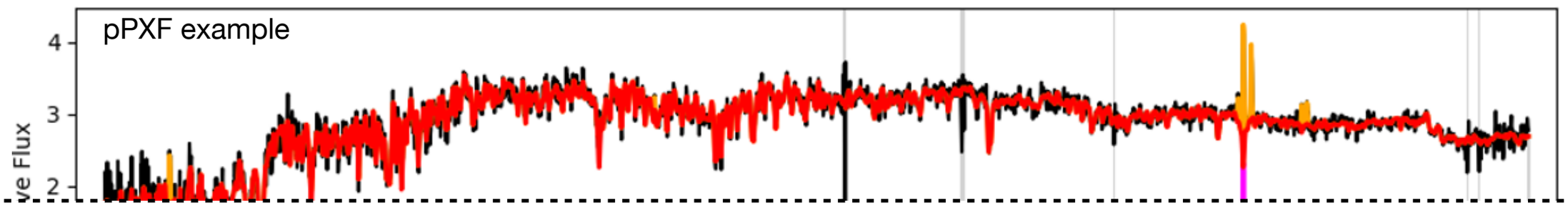
(Credit R. van den Bosch)

Stellar Populations recovery

Stellar Populations recovery

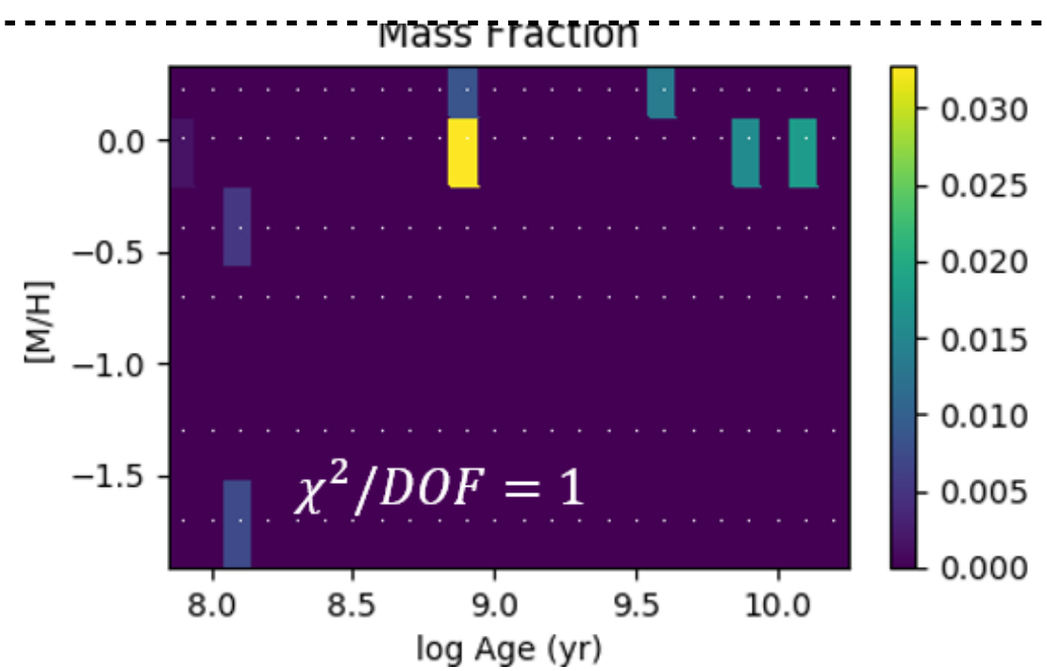
Regularisation

- The weights associated to each template tells you something about the star formation history and chemical evolution of a galaxy
- NNLS estimators are relatively quick. They can handle 1000s of templates.
- Regularisation is typically involved to obtain more physically meaningful solutions



Can we avoid regularisation?
Can we let the data tell us how much regularisation is allowed?

No regularised

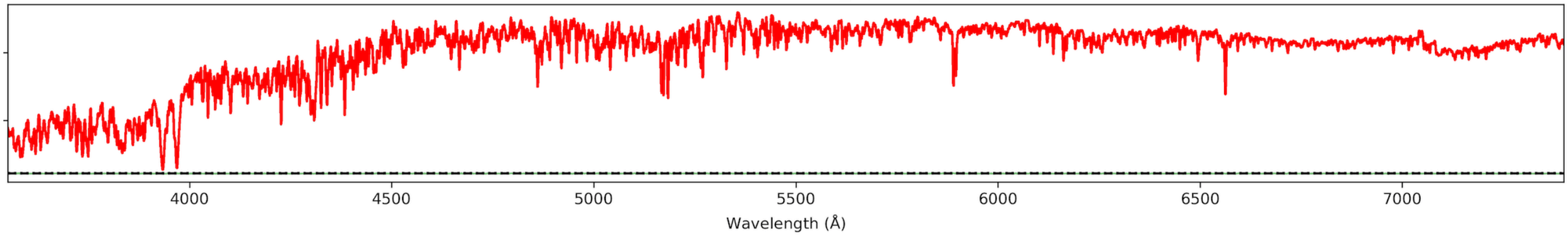


Regularised

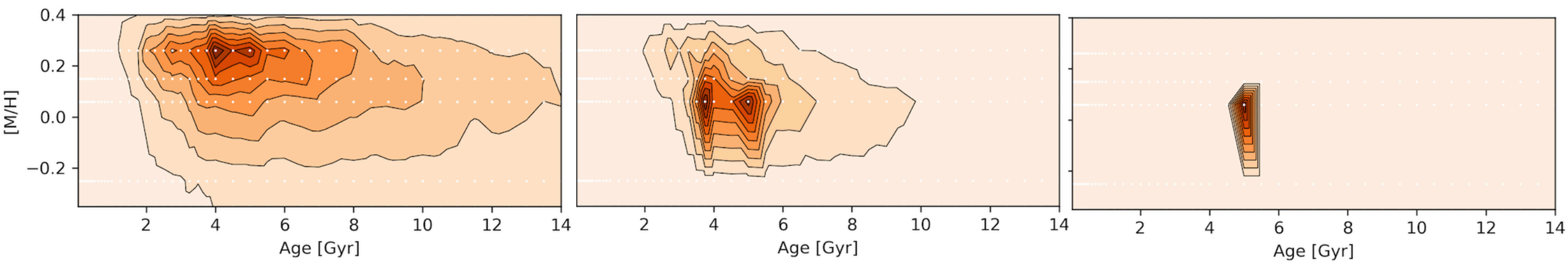
Stellar Populations recovery

Robustness of the solution

Age = 5 Gyr [M/H] = +0.06 dex



SNR = 50 **SNR = 100** **SNR = 1000**



-
-

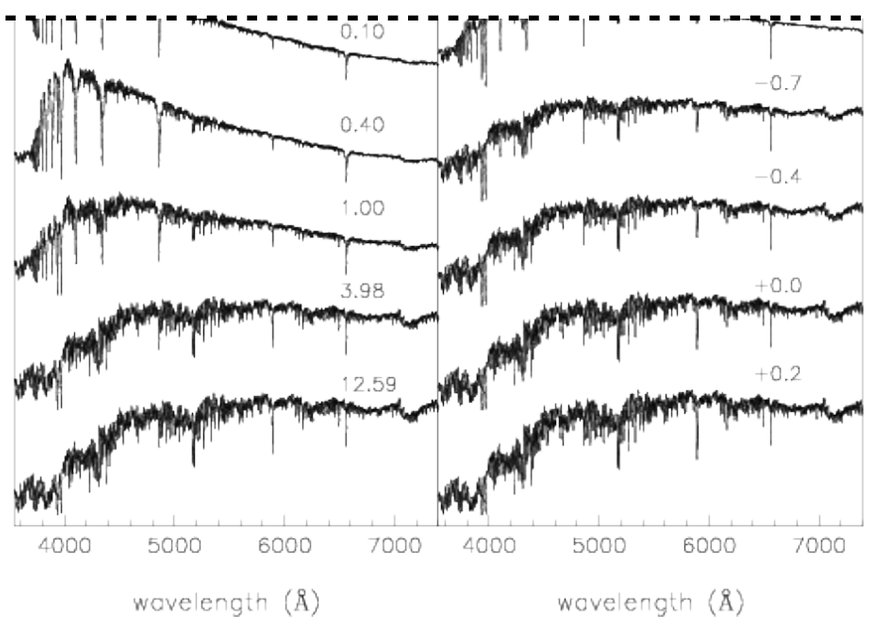
How can we get more robust solutions for intermediate SNR levels?

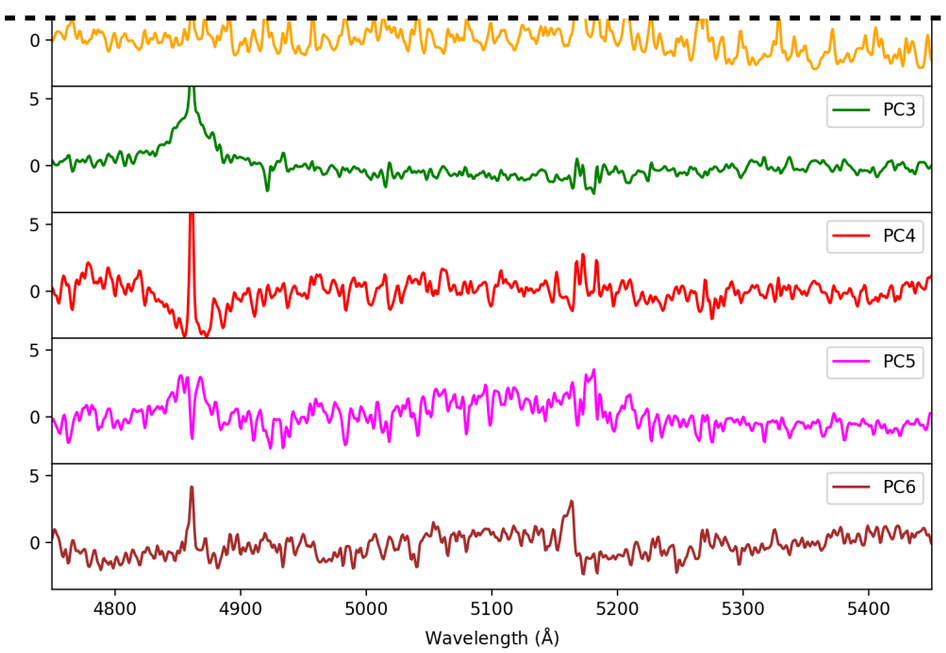
Stellar Populations recovery

Dimensionality reduction on templates

- Using full library of templates is extremely time consuming
- PCA, ICA, NNMF, …. reduce dimensionality, so it could help
- Inversion produces non-physical solutions (i.e. negative weights)
- Can we perform the fitting in some latent (lower dimensional) space that allows to reverse the process? ***(see Prashin's talk tomorrow)***

Can dimensionality reduction help speeding up computation without a strong penalty on robustness of the solution?





Summary

- Life of galaxies is encoded in their spectra
- Kinematics and populations hold the key information
- Bayesian frameworks give robust accounts of degeneracies in the recovery of both LOSVD and stellar populations, but not very fast
- Stellar populations more difficult to unveil: regularisation, robustness, speed
- Eventually we want to work with IFU data and fit all spectra simultaneously, but we need to get the correct solutions for single spectra first
- How can Machine Learning help handling these issues?