Galaxy Evolution with Machine Learning

(a very misleading title...)

Laerte Sodré Jr.

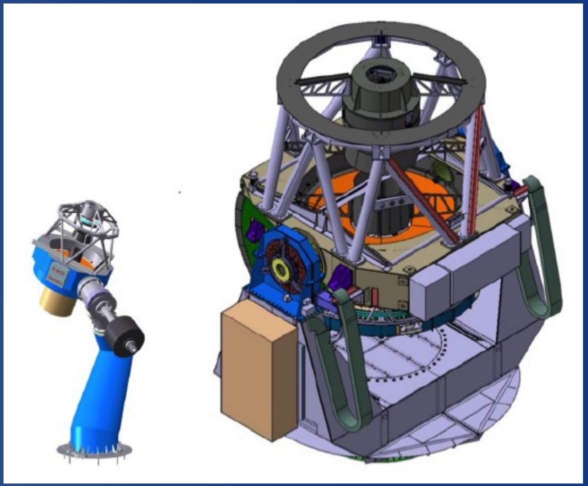
IAG – USP

FoF Meeting, Córdoba, March 29 – April 1, 2016

- **1. Motivation: *J-PAS, J-PLUS, S-PLUS, PFS...***
- **2. Estimation of emission lines for simulation of realistic spectra and survey design**
- **3. BEEP: Bayesian Estimation with Empirical Priors**
  - **Photometric redshifts for LSS**
  - **Spectral Synthesis for Galaxy Evolution**
  - **GALANTE survey for bright stars**

1. J-PAS

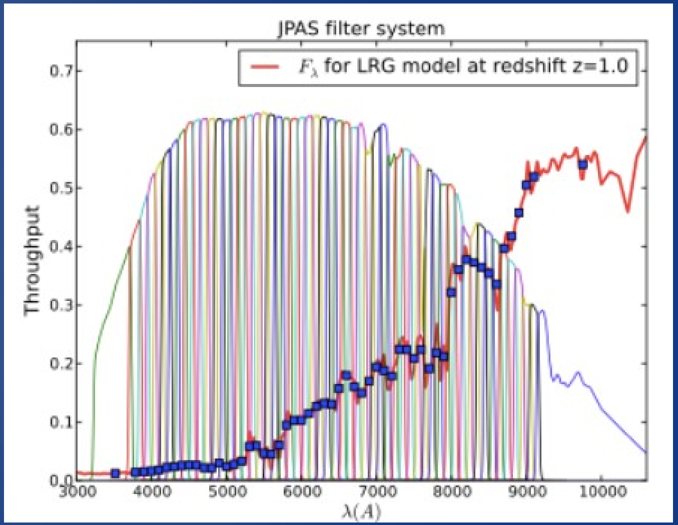
Javalambre Physics of the Accelerating Universe Astrophysical Survey

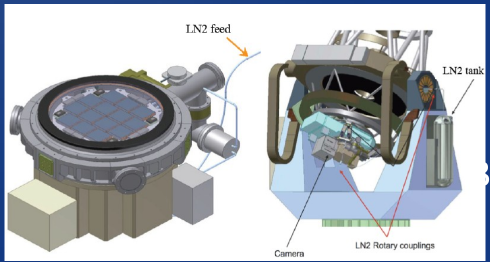


j-pas.org arXiv: 1403.5237

- Collaboration between Brazil and Spain
- Javalambre Astrophysical Observatory (JAO): two telescopes: 2.5m (FOV 3 deg diam) & 80cm (FOV 2 deg diam)

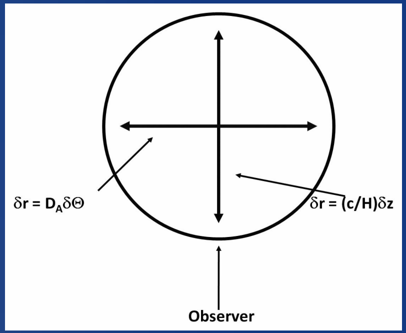
  managed by CEFCA- Centro de Estudios de Física del Cosmos de Aragón, in Teruel, Spain
- Photometric survey of ~8500 sq. deg. to i~22
- Photometric system with 59 filters (54 narrow band, 5 broad band)

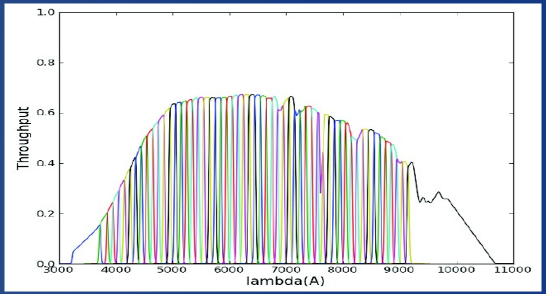




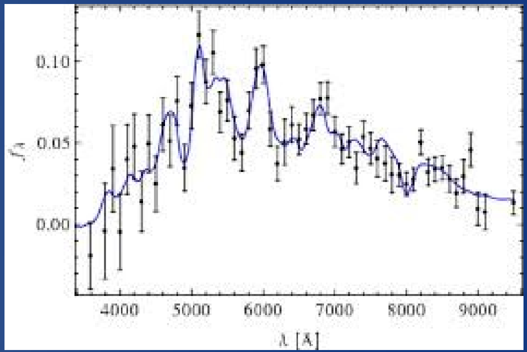
J-PAS Science

- **Main driver: the nature of Dark Energy through measurement of the BAO scale up to z~1**
- **Accuracy of ~0.003(1+z) up to z~1 for ~10 [8] LRGs: measurement of radial BAOs!**
- **DE science: BAOs, galaxy clusters, gravitational lensing, RSDs**
- ***Large scope of science: from asteroids to cosmology...***





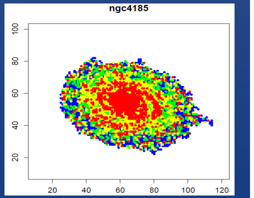
SN Ia

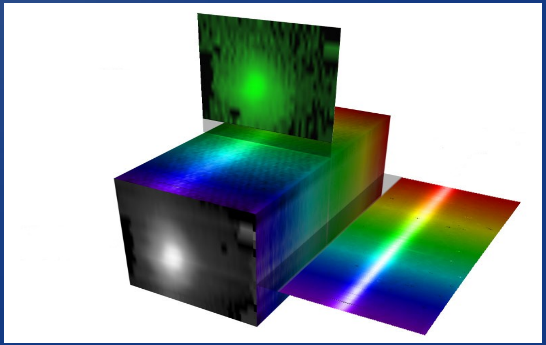


- **J-PAS photometry ~ low resolution spectrum (R~40-60) for each pixel in the sky up to 23 mag arcsec^{-2}!**

New window: "All-sky" IFU! Potential for new discoveries!

distribution of stellar population properties within galaxies





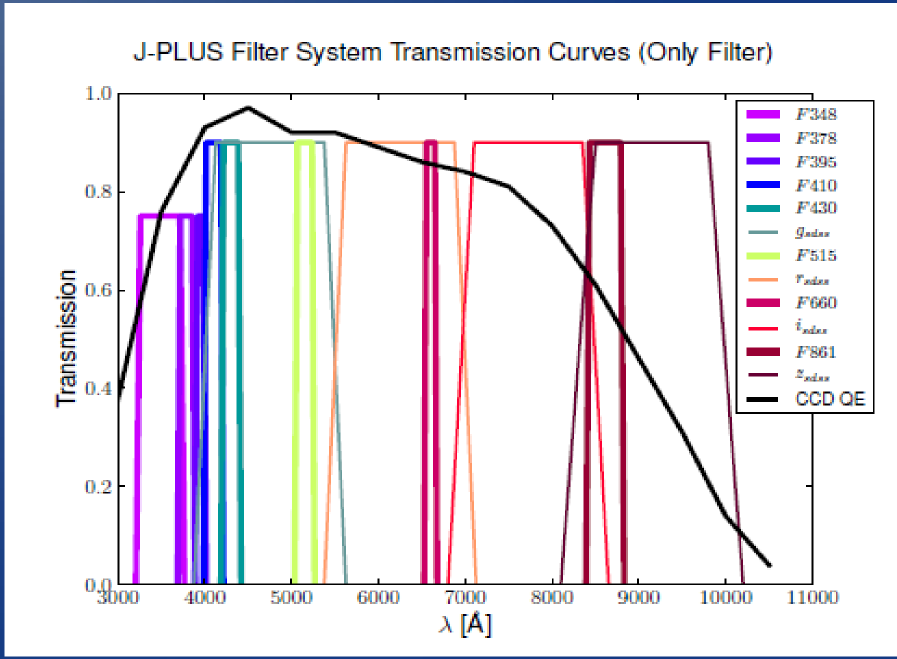
Extensions: J-PLUS (*T80-N*)

J-PLUS: Survey with T80-N (@JAO)

Motivation:

- **Photometric calibration for J-PAS**
- **Test of J-PAS scientific and technical management systems**

- **12 filters: SDSS griz + 8 narrow/intermediate band filters**
- **Survey area: 5000 sq. deg.**
- **~3 years, started Nov 2015**
- **Science: from asteroids to distant quasars**

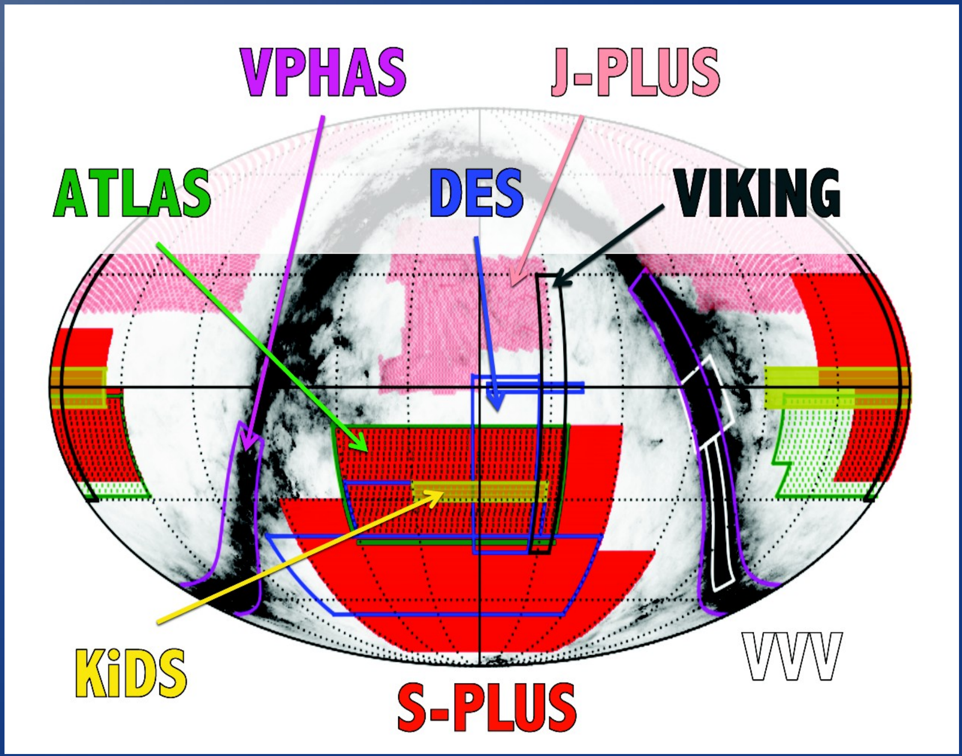


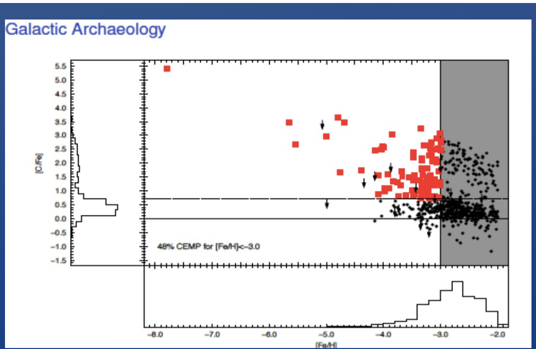
Extensions: S-PLUS *(T80-S)*

S-PLUS: Survey with T80-South (@CTIO)

Science with S-PLUS

- Stellar populations – which models fit the observations best? Extr
- Galaxy morphology, galaxy evolution Extr
- Groups, clusters Extr
- Planetary nebula and globular clusters (Galactic and extragalactic) Pet, Extr
- The halos of galaxies and intracluster light Pet, Extr
- Variable stars Low/High cadence, Galactic, Extr
- Spectral and morphological type classification Extr
- Halpha and OII maps - star formation in the nearby universe Extr
- The lowest-metallicity and carbon-enhanced stars in the Galaxy ultra-short
- Search for blue horizontal branch stars and blue stragglers Extr
- Asteroids Low/High cadence, Extr (if we pay attention to the order of filters)
- Supernovae High cadence, Low cadence
- Satellites in the Milky Way Extr
- Quasars Extr
- Quasar variability High cadence

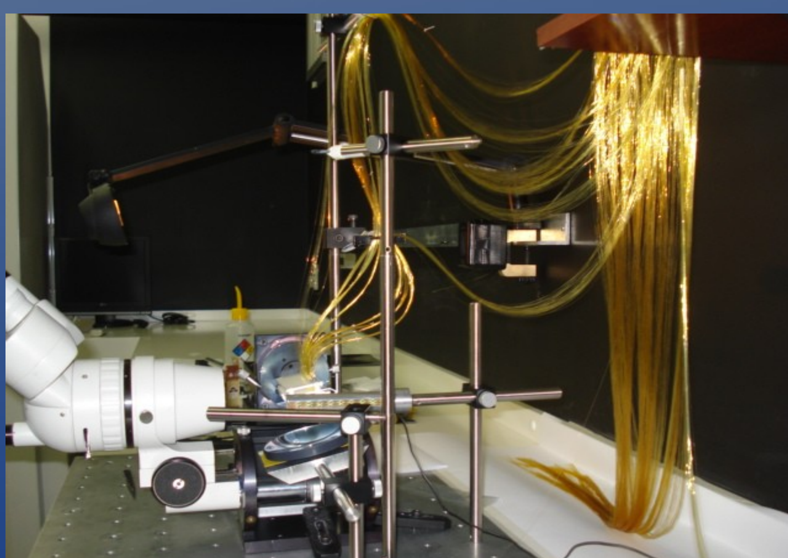


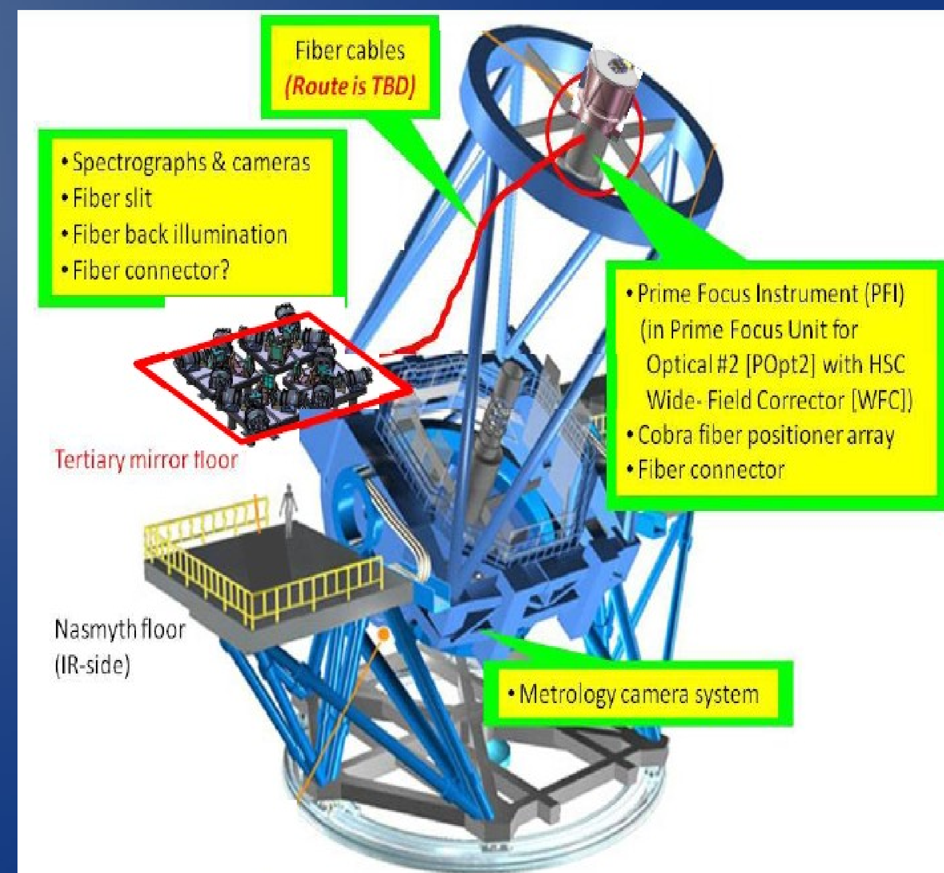


PFS/SuMIRe
Prime Focus Spectrograph for the Subaru Measurement of Images and Redshifts survey

- PI: Hitoshi Murayama Kavli IPMU (U. Tokyo)
- Survey epoch: 2019-2023
- Spectrograph for the Subaru Telescope: 2400 optical fibers within a FOV of 1.3 deg diameter
- Spectral coverage: 0.38 – 1.3 microns,

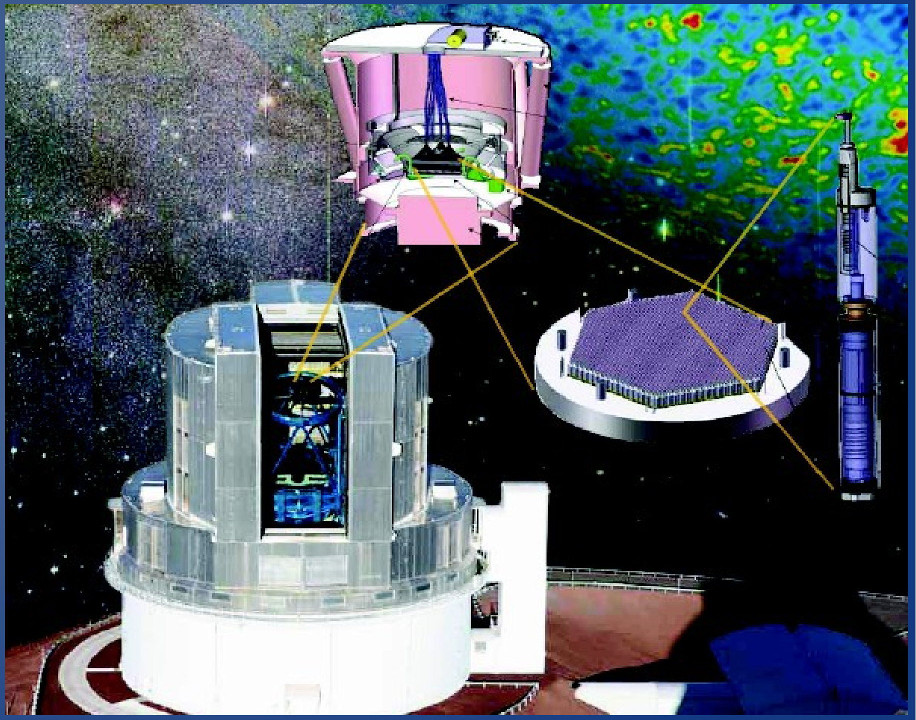
  R ~ 3000 – 5000
- Brazil (USP+LNA): optical fiber subsystem





PFS/SuMIRe
Prime Focus Spectrograph for the Subaru Measurement of Images and Redshifts survey

- **Science:**
- BAOs @ $0.8 < z < 2.4$ ($9.3\ h^{-3}\ Gpc^{3}$)
- Cosmological distances with accuracy of 3%; structure growth with 6%
- Local Cosmology: Milky Way & Andromeda history through the observation of $\sim 10^{6}$ stars
- Chemo-dynamical evolution and dark matter in Local Group dwarf galaxies
- Galaxy populations and structures @ $1<z<2$
- “Lyman break” & “Lyman alpha” galaxies @ $3<z<7$: glimpses on reionization

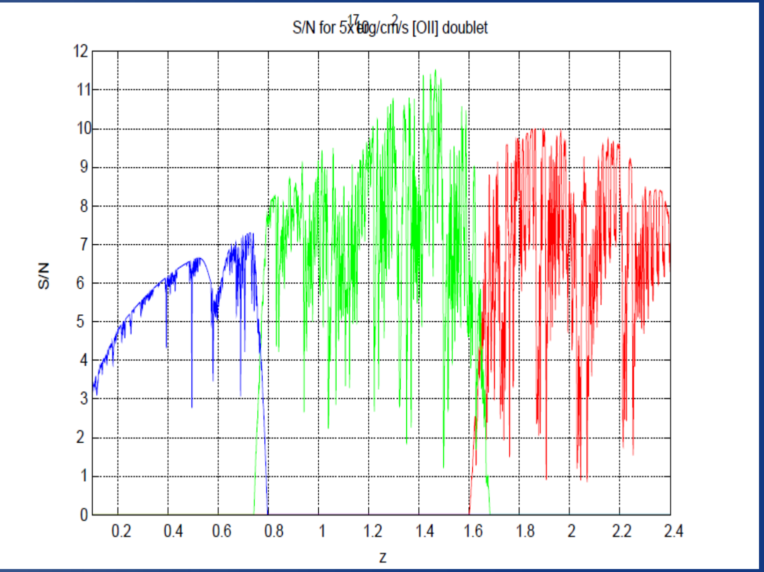


Takada et al., 2014 arXiv:1206.0737

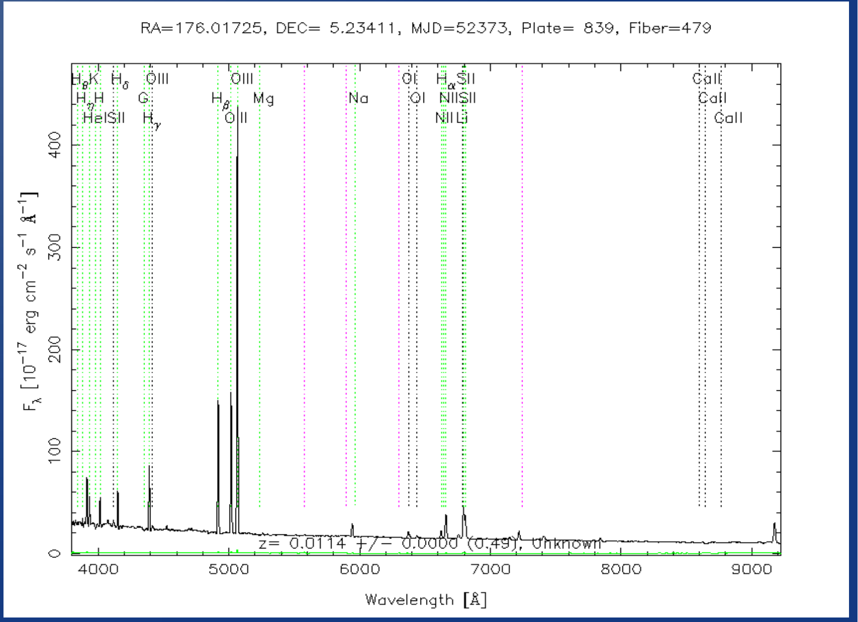
http://sumire.ipmu.jp/en/2652
https://www.youtube.com/watch?v=5mW3v2k8Ofo

2. A data-driven approach to emission line properties of galaxies

- **Motivation: *Survey design***
- **new surveys (e.g. JPAS, PFS) require realistic simulated spectra for defining their parameters and preparing analysis tools**

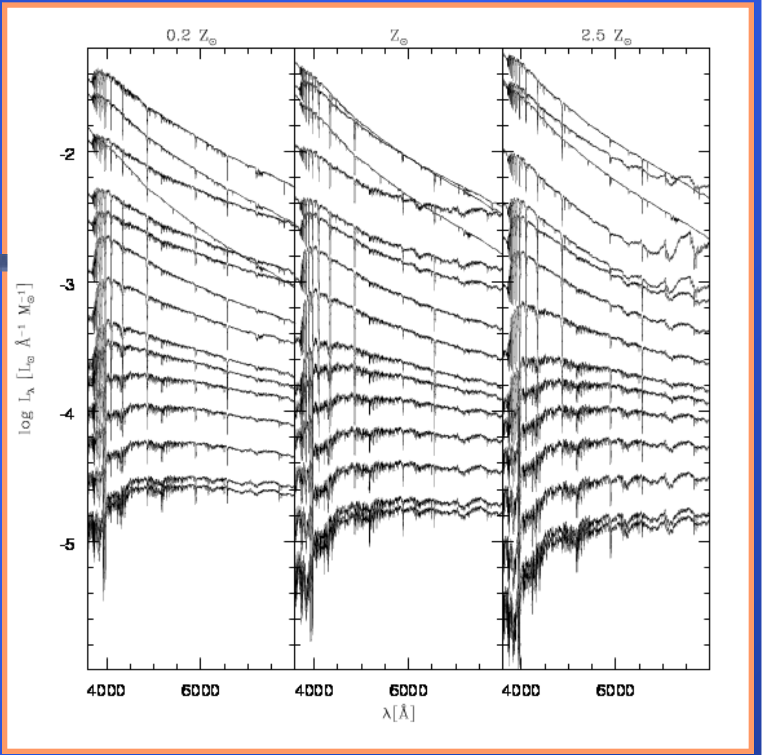


- **most spectral evolution models (e.g. Bruzual & Charlot 2003) simulate only the spectral continuum + absorption lines (stellar libraries)**



- **Here we propose an *empirical, data-driven approach,* to model emission lines given a simulated galaxy continuum + absorption lines spectrum**

Modeling Spectra with STARLIGHT



- STARLIGHT (Cid Fernandes et al., 2005) models a spectrum (continuum + absorption lines) using Bruzual & Charlot (2003) SSPs

- Each spectrum (regions without emission lines) is modeled by a sum of SSPs + extinction + line broadening

- Emission lines are not part of the synthesis; their parameters are obtained by the analysis of the observed minus fitted spectrum

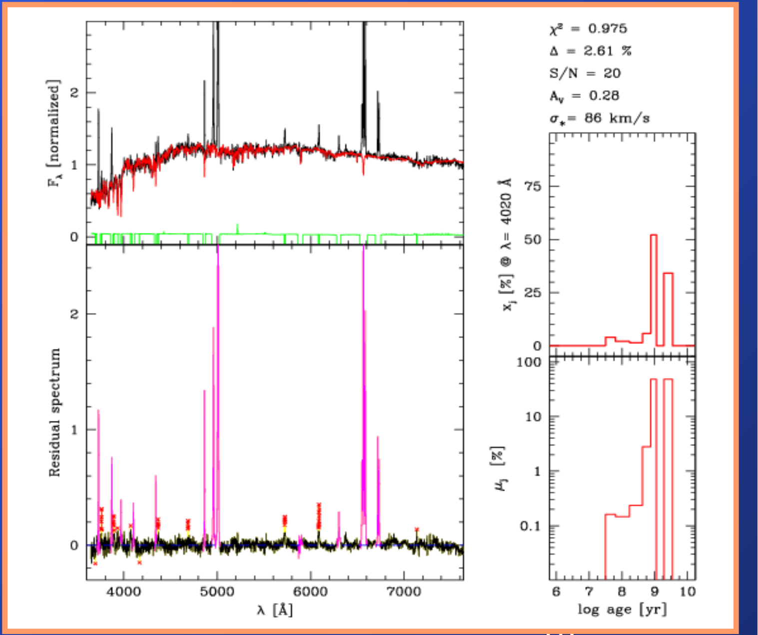
- Sample: SDSS/DR10 spectra with S/N>25

  BPT diagram:

  SF: 15445 AGN: 40748 P:16644

- What we want from STARLIGHT:

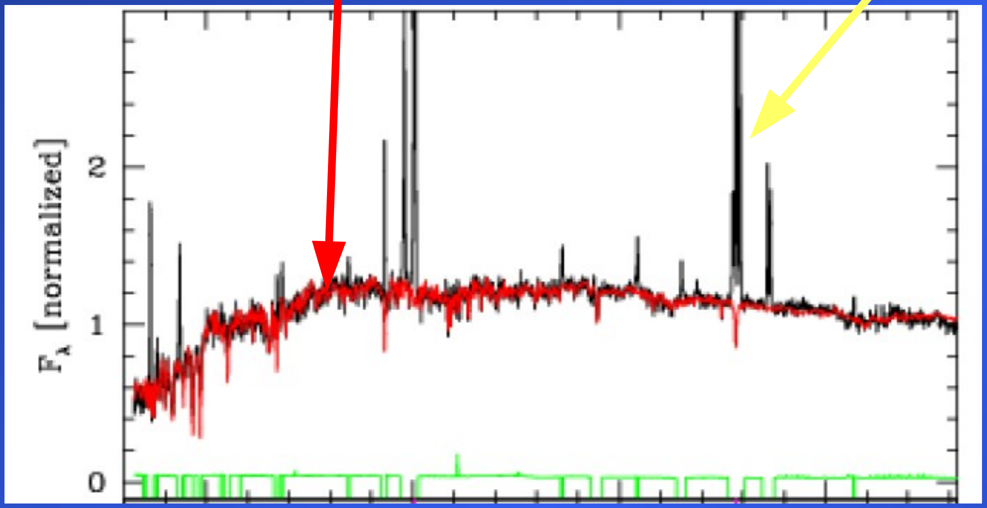
  model spectra + EWs of ELs



An Empirical Approach

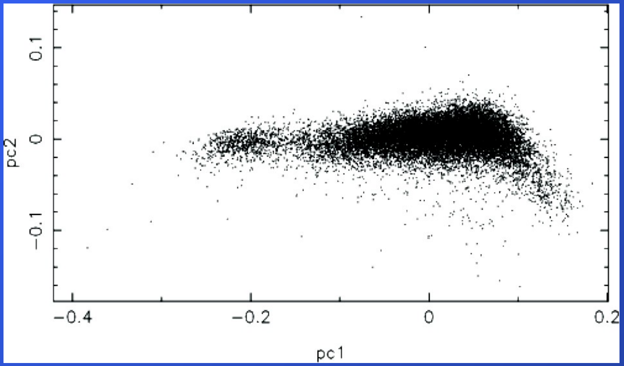
- *There is a connection between galaxy continuum and emission lines*
- Given f_λ, estimate emission line EWs

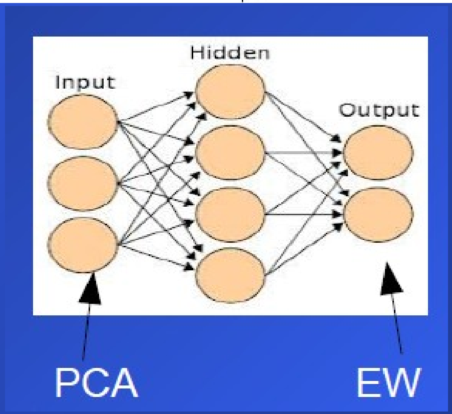
Model spectrum (red)

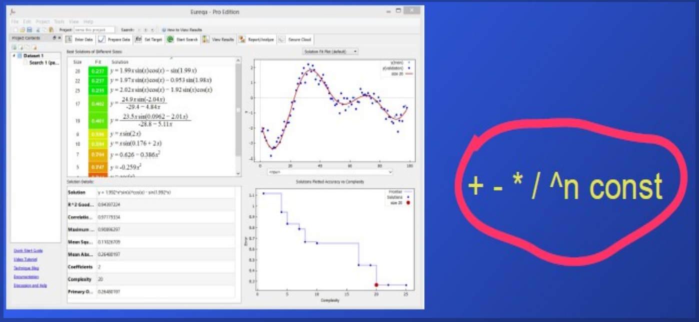


Here: *we use results of STARLIGHT modeling (Cid Fernandes et al. 2005): model spectra and EWs of ELs*

- some ML tools:
- PCA
- ANN
- symbolic regression with Eureqa...







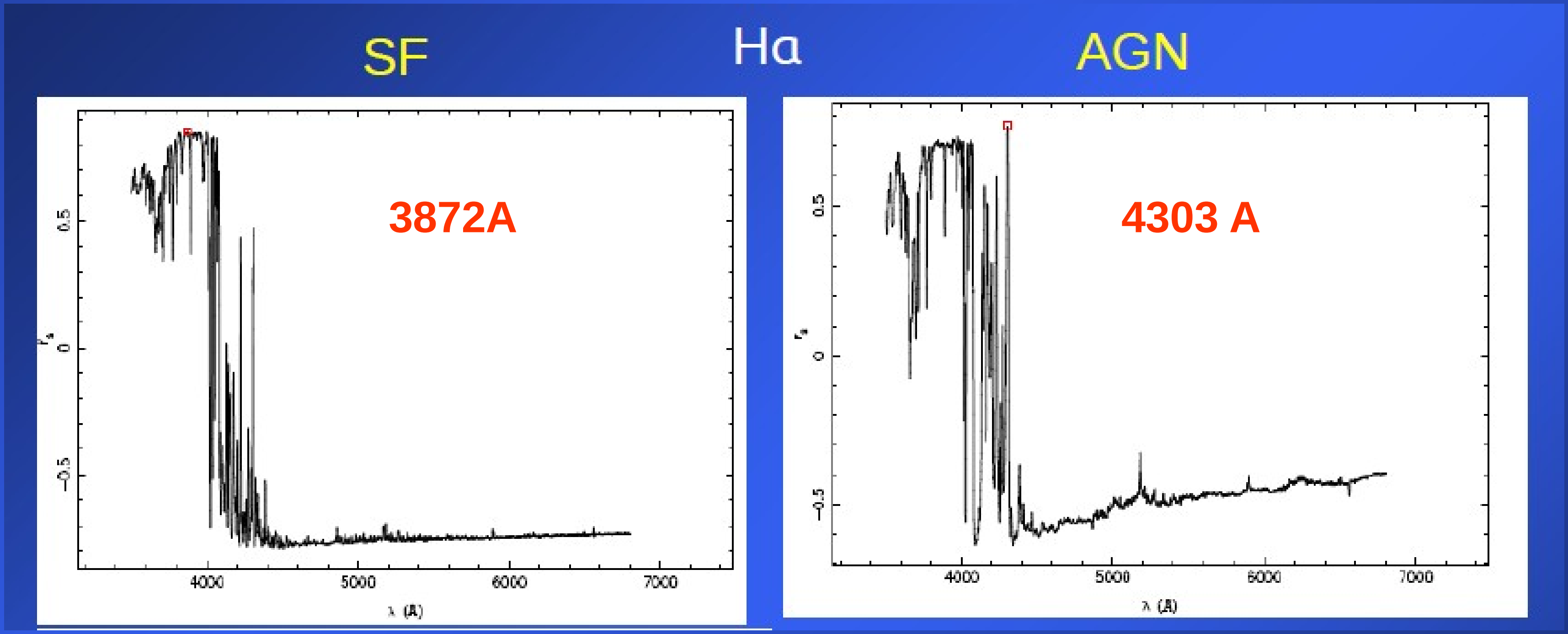
EW estimation

- We assume that, for each line, we may relate its EW to the flux of the model continuum at a certain ***pivot wavelength***:

$$EW = a + b\, f(\lambda_p) + c\, f(\lambda_p)^2$$

- We assume that the model spectra is normalized at λ=4020A

- λ_p: wavelength that maximizes the absolute value of the Spearman correlation coefficient between the line EW and $f(\lambda)$ for a sample



Typical uncertainties in EWs:

- **SF: 0.18 – 0.35 dex**
- **AGN: 0.25 – 0.41 dex**

families of lines tend to have similar pivot wavelengths!

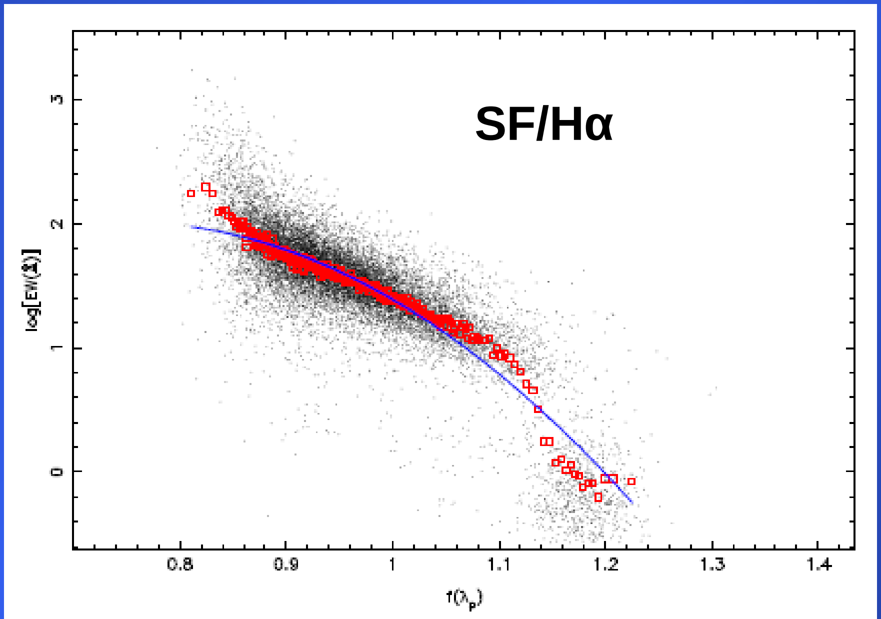


Table 1. Pivot wavelength for the SF sample.

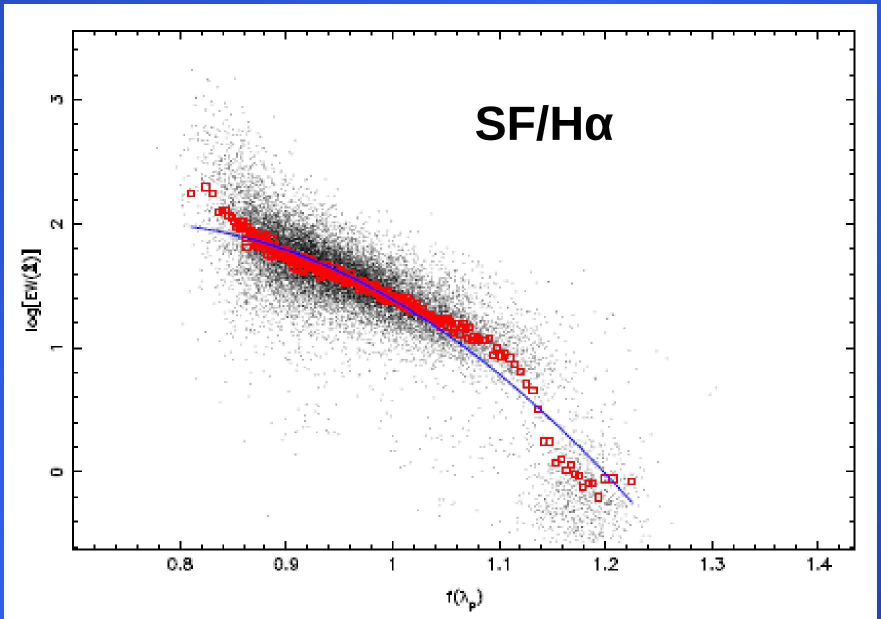
line	number of galaxies	r_s	λ_p (Å)
$[OII]3727$	13200	-0.846	4317
$[NeIII]3869$	13474	0.551	3863
$H\delta$	15055	0.739	3995
$H\gamma$	14960	0.785	3819
$[OIII]4363$	1790	0.410	4063
$H\beta$	15164	0.834	3819
$[OIII]4959$	14551	-0.857	4317
$[OIII]5007$	15135	-0.866	4317
$[OI]6300$	14966	0.778	4032
$[NII]6548$	14832	0.627	4062
$H\alpha$	15445	0.852	3872
$[NII]6584$	15445	0.666	4062
$[SII]6716$	14997	0.844	3871
$[SII]6731$	15023	0.840	3871
$[HeI]4471$	10395	0.318	3583

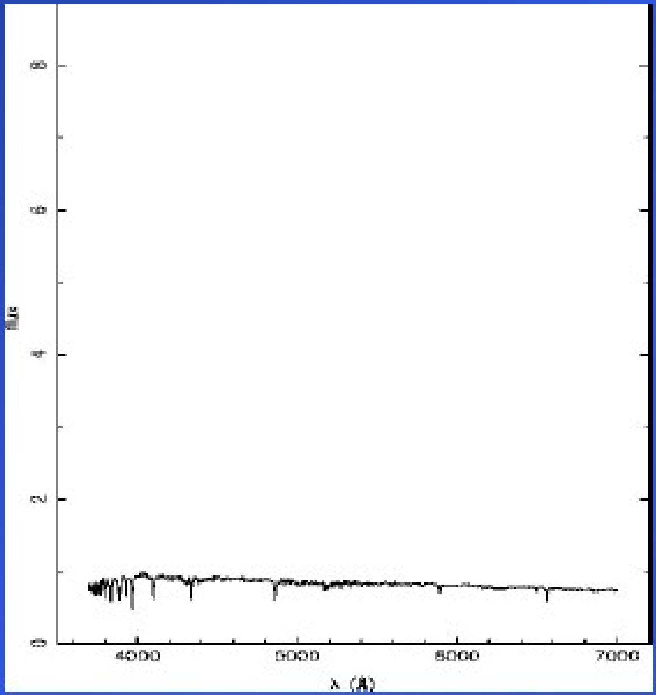
Table 2. Pivot wavelength for the AGN sample.

line	number of galaxies	r_s	λ_p (Å)
$[OII]3727$	37841	0.547	4303
$[NeIII]3869$	22177	0.356	4073
$H\delta$	30239	0.585	3996
$H\gamma$	25436	0.734	3995
$[OIII]4363$	6029	0.345	4071
$H\beta$	34227	0.704	3972
$[OIII]4959$	31271	0.515	3985
$[OIII]5007$	37679	0.524	3975
$[OI]6300$	35678	0.365	4005
$[NII]6548$	36925	0.754	4303
$H\alpha$	40748	0.768	4303
$[NII]6584$	40748	0.753	4303
$[SII]6716$	35887	0.695	4303
$[SII]6731$	35746	0.682	4303
$[HeI]4471$	12866	0.440	3883

For each line:

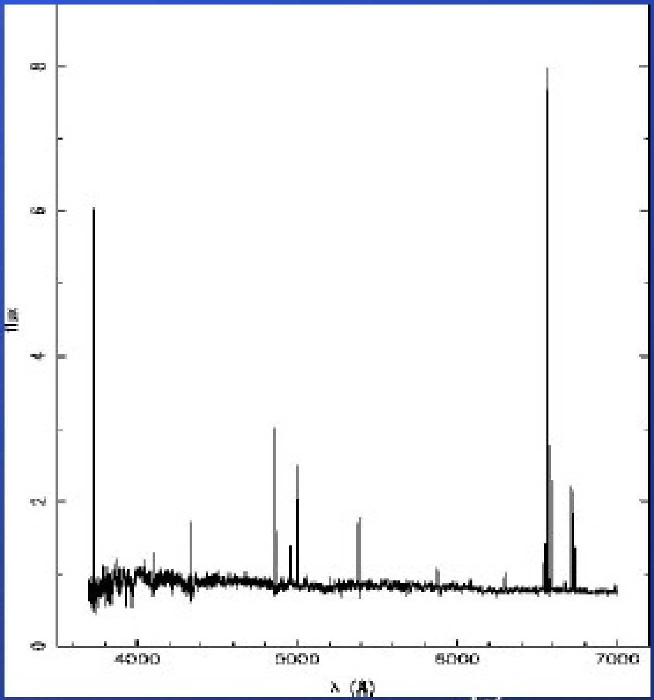
$$EW = a + b\, f(\lambda_p) + c\, f(\lambda_p)^2$$





model

lines+noise



“observation”

3. Bayes Theorem for parameter estimation

We want to estimate a parameter x from data D and a set of assumptions/models H

- **Bayes theorem:**

$$P(x|D,H) \propto P(D|x,H)\, P(x|H)$$

posterior **likelihood** **prior**

- **photometric redshifts, spectral synthesis:**

two basic approaches:

templates

training set

actually, *two different priors* *(and each method is as good as its prior)*

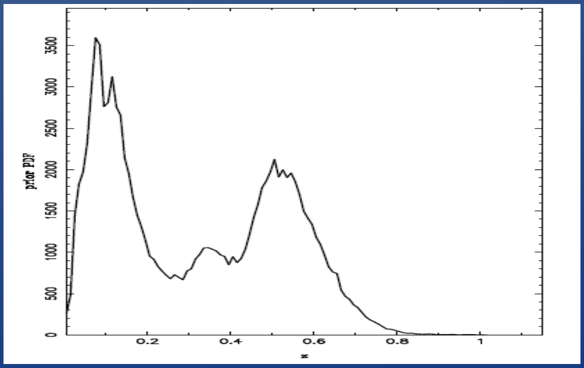
BEEP:
Bayesian Estimation with Empirical Priors

- **Wolf 2009**
- Training set TS for photometric redshift estimation:

  N galaxies with M magnitudes and known redshift D':$\{x_{ij}, z_i\}$
- Target galaxy y with magnitudes y_j
- Prior: redshift distribution of the TS

$$P(x|D,H) \propto P(D|x,H)\, P(x|H)$$

as far as we sample the prior with the TS → posterior = likelihood



BEEP:
Bayesian Estimation with Empirical Priors

- Training set (TS) for *photometric redshift* estimation:

  $D:\{x_{ij},z_i\}$

- Target galaxy y with magnitudes y_j
- Likelihood that z is equal to the redshift of the i-th galaxy in TS:

  $L(z_i) \propto \exp(-E_i)$ $\quad E_i=1/2 \sum(y_j-x_{ij})^2/\sigma_j^2$

  normalization: $\sum L(z_i) = 1$ $(i=1\ldots N)$

- posterior $P(z|y,D)$: to obtain a PDF for z we compute the likelihood for all galaxies in D

BEEP:
Bayesian Estimation with Empirical Priors

$P(z|y,D) \propto P(y|z,D)\, P(z)$

- TS: $\{x_{ij}, z_i\}$
- target galaxy photometry: $\{y_j\}$
- Likelihood $L(z_i) \propto \exp(-E_i)$

$E_i = 1/2 \sum (y_j - x_{ij})^2 / \sigma_j^{\,2}$ normalization: $\sum L(z_i) = 1$ $(i=1...N)$

PDF $P(z|y,D')$: Histogram

we represent $P(z|y,D)$ by an histogram with bins $\{z_h\}$

$P(z_h|y,D') \propto \sum L(z_i)$ for all galaxies in TS with z_i within bin z_h

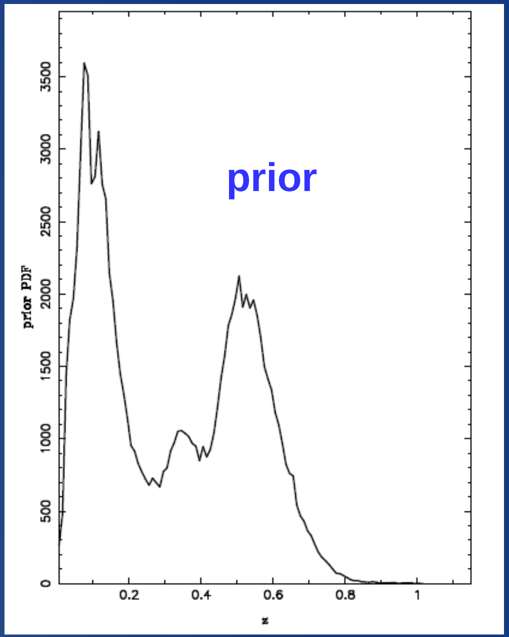
BEEP: photometric redshifts for SDSS DR10

D: ugriz dered magnitudes + zspec

- training set: 100,000

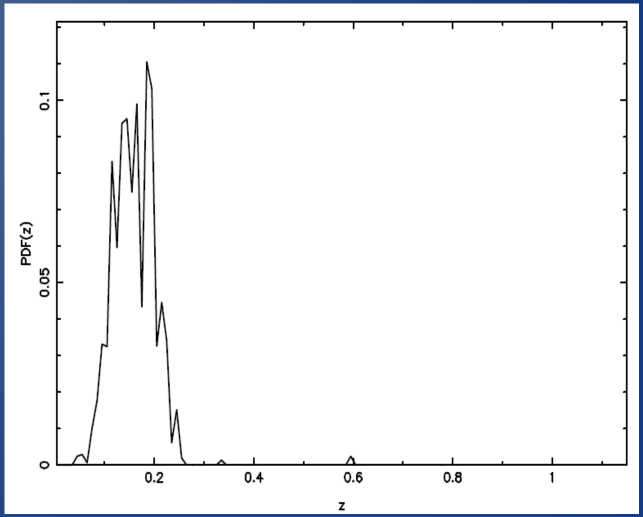
  test set: 10,000

  PDF: histograms with 105 Δz=0.01 bins



redshift estimates:

- most probable redshift: PDF peak
- nearest-neighbour redshift: clone
- mean redshift: $\langle z \rangle = \Sigma L_i z_i$

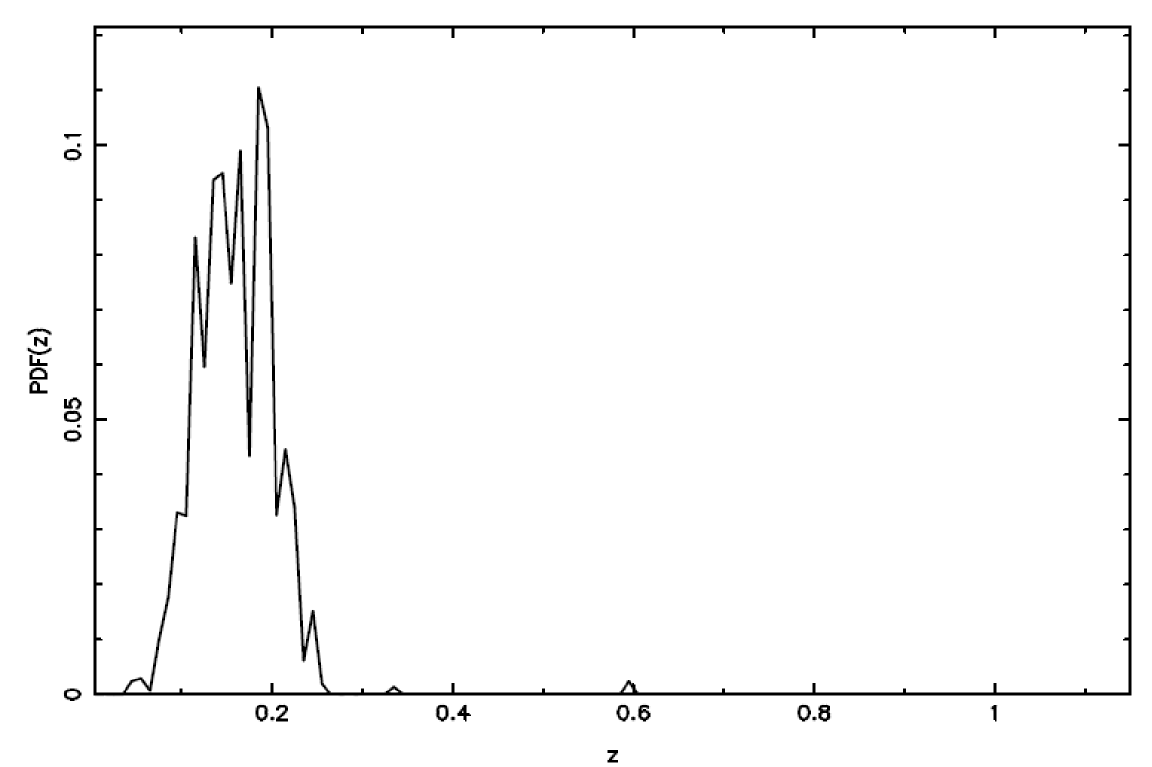


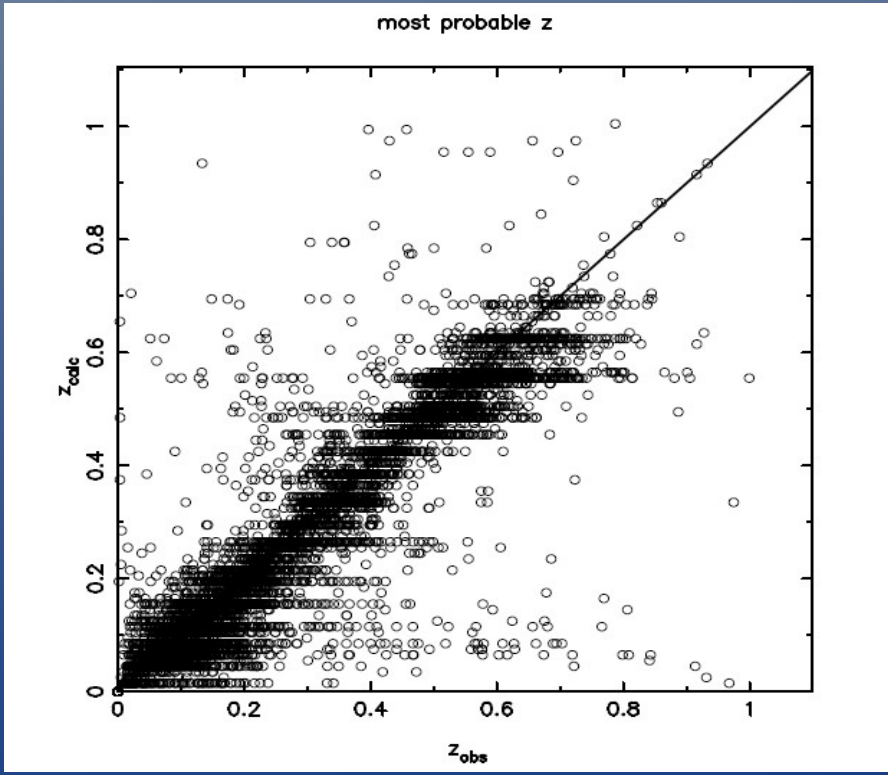
- **quality of the estimation:**

  redshift dispersion $\sigma_i = [\Sigma L_i (z_i - \langle z \rangle)^2]^{1/2} / (1 + \langle z \rangle)$

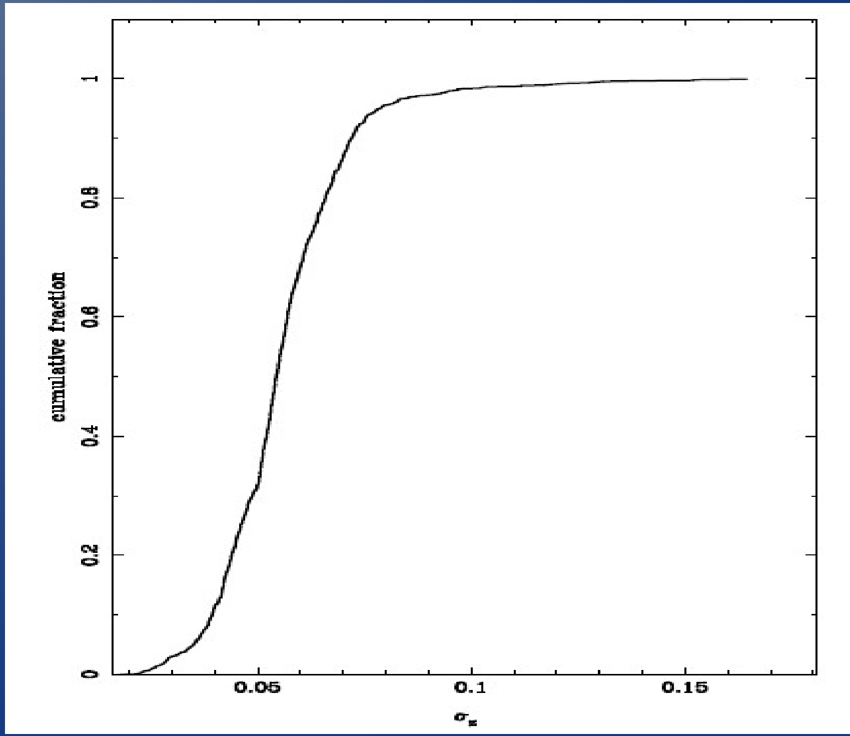
an example:

- maximum probability...0.185
- mean redshift...............0.160
- median redshift............0.157
- clone redshift..............0.184
- z_{spec}............................... 0.195
- quality: σ_z = 0.038





σ=0.056



BEEP: galaxy spectral synthesis

- Challenge within J-PAS (still going on...)

  aim: compare & improve SED fitting codes

- Input:
  - 98 synthetic spectra (different SFHs, extinctions, S/N,...) + zphot
  - J-PAS & J-PLUS photometry
  - Filter response curves, extinction law
  - Set of 2079 SSPs (MILES) with different ages and metallicities

- Output:

  mean ages, metallicity, extinction...

BEEP: galaxy spectral synthesis

- Training Set: 200K simulated spectra in J-PAS & J-PLUS bands
- Model spectrum:
  - star formation rate:
  $$\Psi = \Psi_0 \times \exp(-t/\tau)$$
  - mean (mass-weighted) stellar age:
  $$< t >= \tau \frac{(1+x_1) - (1+x_2)\exp(-\Delta x)}{1 - \exp(-\Delta x)}$$

  where $x = t/\tau$, $x_1 = t_1/\tau$, $x_2 = t_2/\tau$, $\Delta x = x_2 - x_1$

  - $P(< t >)$ uniform between t_1 and t_2
  - parameters: number of SSPs, z, A_v, $< t >$

BEEP: galaxy spectral synthesis

Parameter	J-PLUS	J-PAS
Av	0.63	0.53
log $<t_l>$	0.47	0.39
log $<t_M>$	0.29	0.29
log $<Z_l>$	0.50	0.50
log $<Z_M>$	0.38	0.38

PRELIMINARY!

Competitive results!

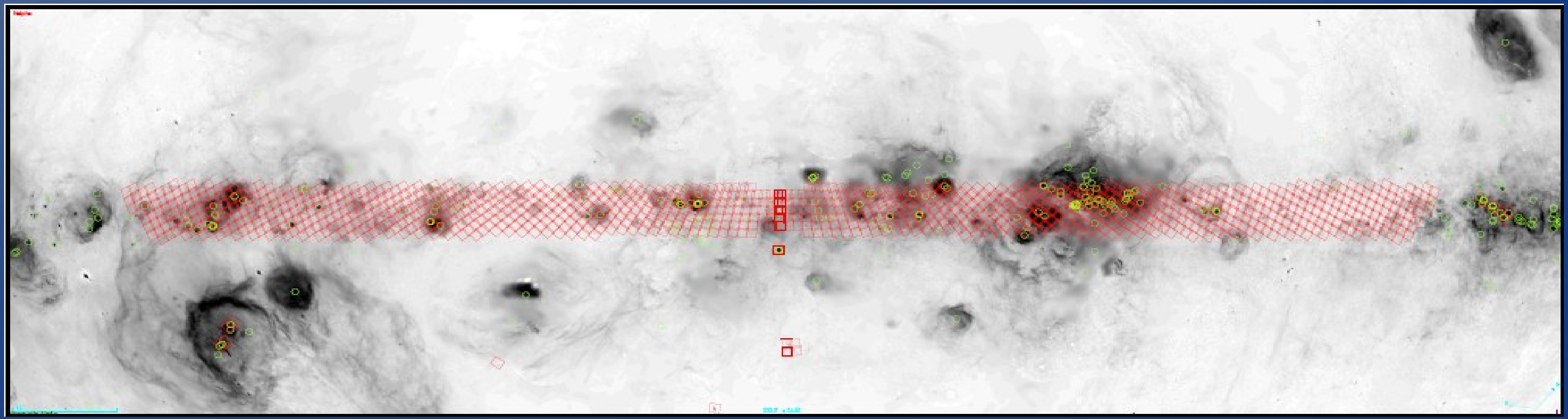
BEEP allows to use empirical and theoretical training sets

BEEP: GALANTE SURVEY

- Bright-star northern Galactic plane survey with T80N@JAO: AB mag=18, S/N=70

  estimation of stellar parameters: effective temperature, gravity, metallicity, extinction

- with Paula Coelho, Walter dos Santos Jr.:
- BEEP with theoretical (e.g., Coelho 2914)/observed (SDSS/SEGUE)/mixed training sets
- combination of different methods with data fusion



summary

- **J-PAS, J-PLUS, S-PLUS:**

  current surveys need up-to-date tools to maximize the extraction of information from data

- **Estimation of emission lines from model spectra**

  realistic spectra for survey design

- **BEEP: Bayesian Estimation with Empirical Priors**

  an useful tool for photometric redshifts and spectral synthesis