AMAZED : Algorithms for Massive Automated Z(redshift) Estimation and Determination

V. Le Brun for the AMAZED Team

P.-Y. Chabaud, M. Gray, C. Surace, D. Vibert, A. Allaoui, M. Sarkis, A. Schmitt (CeSAM)

Contributions from S. Arnouts, M. Bethermin, J.G. Cuby and the Euclid and PFS science working groups

A versatile fully automated redshift measurement library

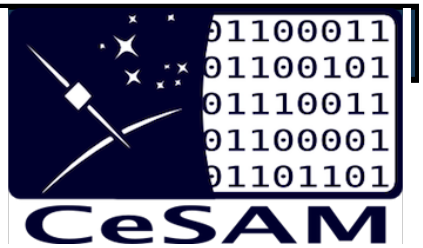
- Human participation up to now (VVDS, VIPERS, DEEP2, even SDSS)
  - No more than 10^5 spectra
  - Validation of software results :

  Redmonster (Hutchinson et al. 2016), EasyLife (Garilli et al. 20XX)

- Impossible for surveys in preparation (Euclid, PFS, Roman, …) :
  - Several millions of objects
  - High observing rate
  - We don’t want to do it anymore

Building on the heritage of VIMOS surveys

A versatile fully automated redshift measurement library



- Adaptable to any instrumental configuration (UV/visible/IR/?)
- Takes into account instrumental effects : LSF, (co)variance
- Fully automated bayesian inference redshift measurement (and error)
- Bayesian object classification (Galaxy, quasars, stars)
- Quantification of reliability and ultimately quality of fit (detection of monsters)
- Core of Euclid, PFS and Roman pipelines

The methods (Galaxies and QSO)

- A least-square fit of model parameters weighted by signal variance
  - Fixed redshift grid
  - Logarithmic sampling for FFT
  - Go back to Tonry & Davis (1979) for a fundamental description

- Redshift Probability Distribution Function calculation for each model
- Combination of all zPDFs into one
- The N best redshifts are identified from the PDF peaks
- Measurements of spectral features at the best redshifts solutions

The galaxy model

- Continuum (including stellar absorption lines)
  - *21 templates built from BC03 models (Tremonti+ 2003)*
  - *The amplitude is fitted*
- Emission lines
  - *Predefined list of lines*
  - *Relative ratios predefined (and updated) from 13 VVDS stacked spectra*
  - *Redshift and width are free parameters*
- Interstellar absorption lines
  - *Velocity shift is fixed (-150 km/s), but could be free*
  - *Relative ratios predefined from 13 VVDS/Steidel stacked spectra*
- Intergalactic Medium absorption
  - *Tabulated from Meiksin (2002) : 7 curves at 11 redshifts between 2 and 7*
- Interstellar extinction
  - *Tabulated from Calzetti (2000)*

The method – The model

- Continuum (including stellar absorption lines)
  - *21 templates built from BC03 models*
  - *The amplitude is fitted*
- Emission lines
  - *Predefined list of line*
  - *Relative ratios*
  - *Position a*
- Interste
  - *del stacked spectra*
- Inte
  - *curves at 11 redshifts between 2 and 7*
  - *ecti (2000)*

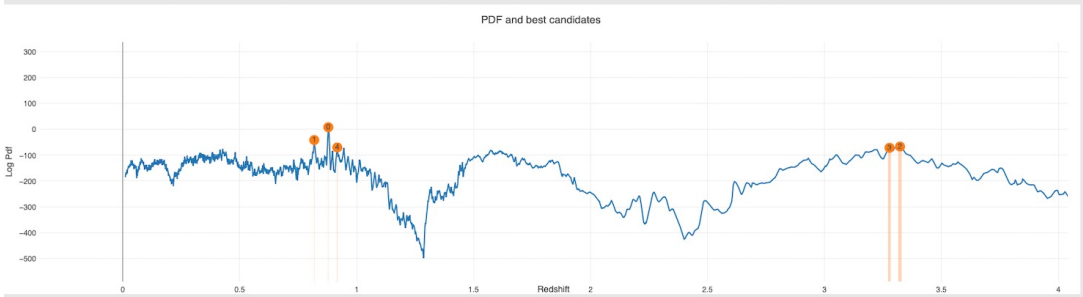
Bring your own if you don't like ours :
- All model components are handled as configuration files
- The user can decide to use them or not

The method – Redshift Determination

- PDF from each model are combined
  - *Marginalization (over all model parameters), final PDF delivered.*
- The best redshift is taken at the maximum of integrated probability
  - *Error on redshift estimated via Gaussian fit*
  - *Integral value under the PDF peak as Reliability level*
  - *To be improved with ML/DL Techniques*
- Secondary redshift values at following peaks
- Code is able to integrate priors
  - *Strong lines : greater probability for "Main Strong lines" (Ha, OII, OIII)*
  - *Hα : greater probability to be an Ha line*
  - *N(z) : an a-priori redshift distribution of Ha emitters*

The method – Redshift Determination

Previous | Next | Results List | Undo zoom



0 Redshif

R..	Reds..	Redshif..	Redshift..	Template..	LinesRatioName	Conti..	Con..	ContinuumA..	VelocityE..	VelocityA
0	0.87826..	0.0002542..	0.9952371..	ssp_1.4Gyr_z..	tpl_NEW-Sbc-extended_TF_catalog....	0.3	-1	1.307702733e-17	230	450
1	0.81802..	0.0003250..	4.8010451..	t5e9_12gyr_z..	tpl_NEW-Sbc-extended_TF_catalog....	0.6	-1	4.520180513e-17	370	450
2	3.32257..	0.0024472..	5.7055066..	ssp_290Myr_..	tpl_COMBINE-ave-Lya-abs-AND-Sc..	0.9	-1	0	10	450
3	3.27784..	0.0021489..	3.0255436..	ssp_290Myr_..	tpl_NEW-Im-extended-blue_TF_cata..	0.9	-1	0	10	150
4	0.91659..	0.0004208..	2.2199397..	cst_6gyr_z02....	tpl_COMBINE-ave-BX-highblue-AN..	0.8	-1	1.019394829e-16	370	450
-1	0.879									
-2	0									

Type of lines: Emission | Absorption | All

Strength of lines: Strong | Weak | All

...	...	Name	Typ	Flux
91...	48...	HbetaA	A	-5.04e-18
81...	43...	HgammaA	A	-1.89e-18
80...	43...	GBand	A	-2.80e-18
77...	41...	HdeltaA	A	-1.45e-18
74...	39...	CaII_H	A	-1.05e-18
73...	39...	CaII_K	A	-1.23e-18
73...	38...	H8A	A	-1.29e-18
72...	38...	H9A	A	-1.19e-18
71...	37...	H10A	A	-6.63e-19
70...	37...	H11A	A	-4.87e-19
67...	35...	FeI	A	-2.00e-19
91...	48...	Hbeta	E	0.00e+0
81...	43...	Hgamma	E	0.00e+0
77...	41...	Hdelta	E	0.00e+0
74...	39...	Hepsilon	E	0.00e+0

The usage – Integration to a pipeline or standalone use

- Interaction with the input and output data : the python client
  - *Generic one with a given data model for public version*
  - *We can provide one adapted to the datamodel used in a pipeline*
  - *Both in input and output*
- Handles parallelization

Performances on simulated data

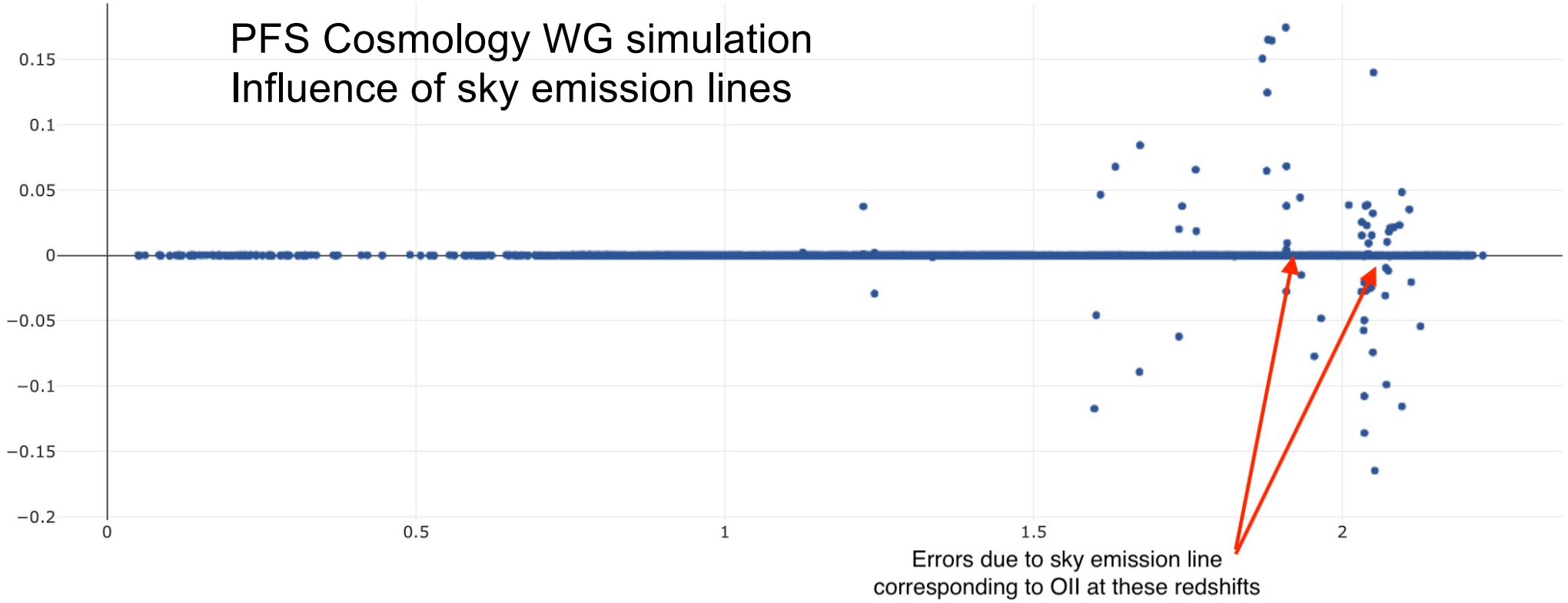
- Efforts from the PFS and Euclid Science Working Groups

- Within requirements so far but…
  - Not all instrumental effects included yet
  - Simulated continuum against simulated continuum templates

  (because building coherent observed templates over 500-18000 A is tough)

Others already said it, but early full E2E simulations are fundamental

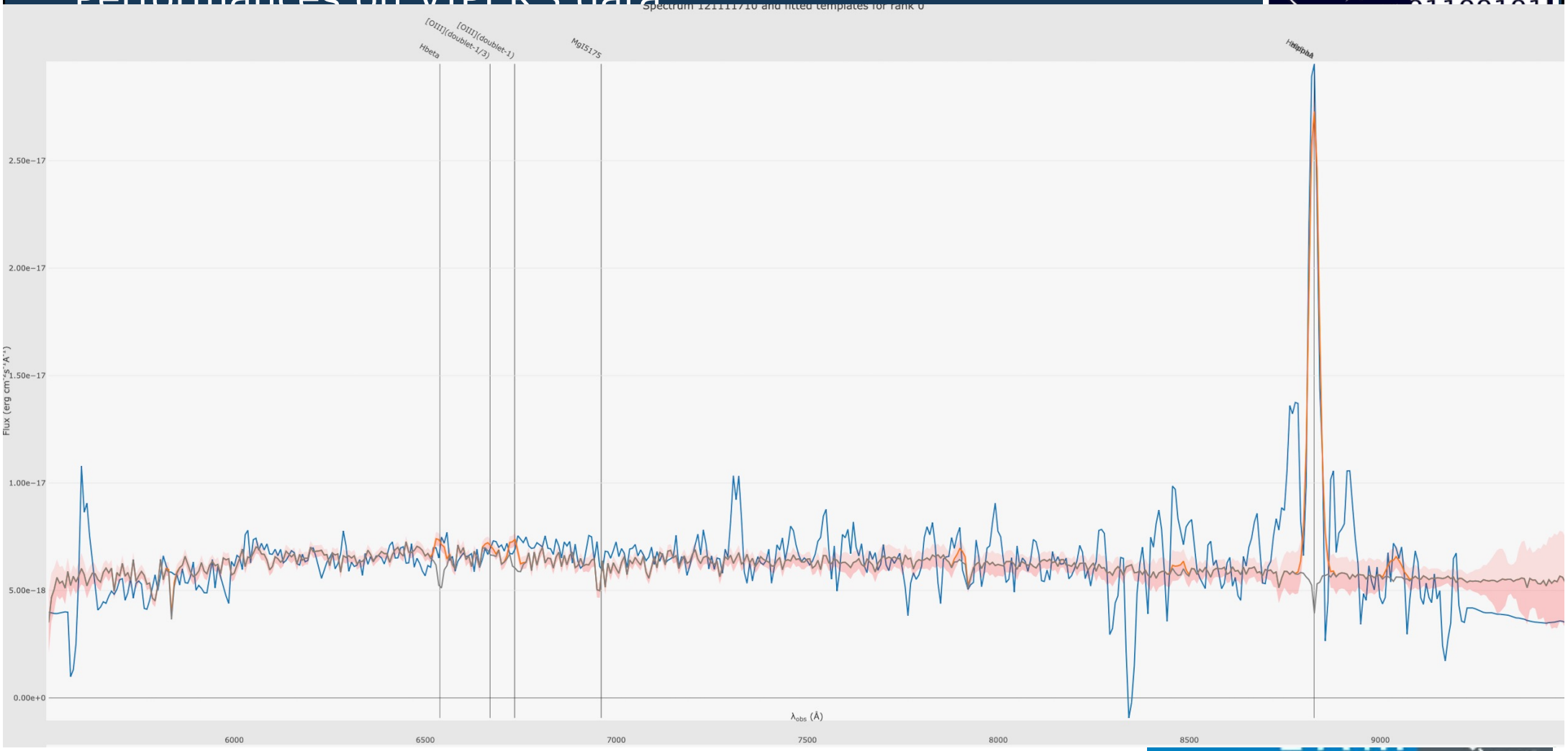
Performances on simulated data



Performances on VIPERS data

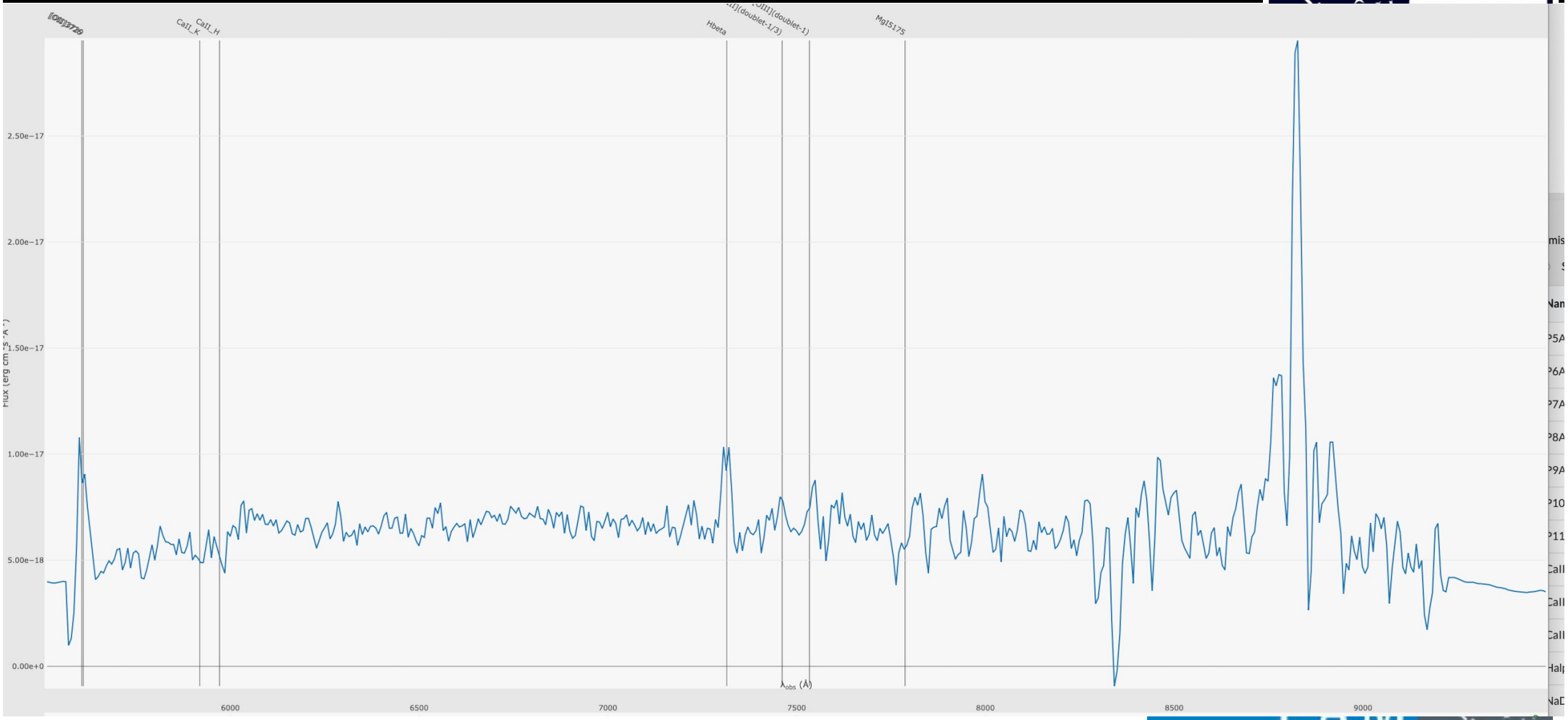
- Results on the flag 4 spectra : 19658 galaxy spectra
- Caveats
  - Variance in VIPERS spectra is ‘not perfectly’ evaluated
  - There can be superposition of spectra
  - Flags (sometime strongly) depend on the personality/experience
  - Human measurement based on features not available
    2D spectrum
    Feature on the edge of spectrum

Performances on VIPERS data



LABORATOIRE D'ASTROPHYSIQUE DE MARSEILLE

Performances on VIPERS data

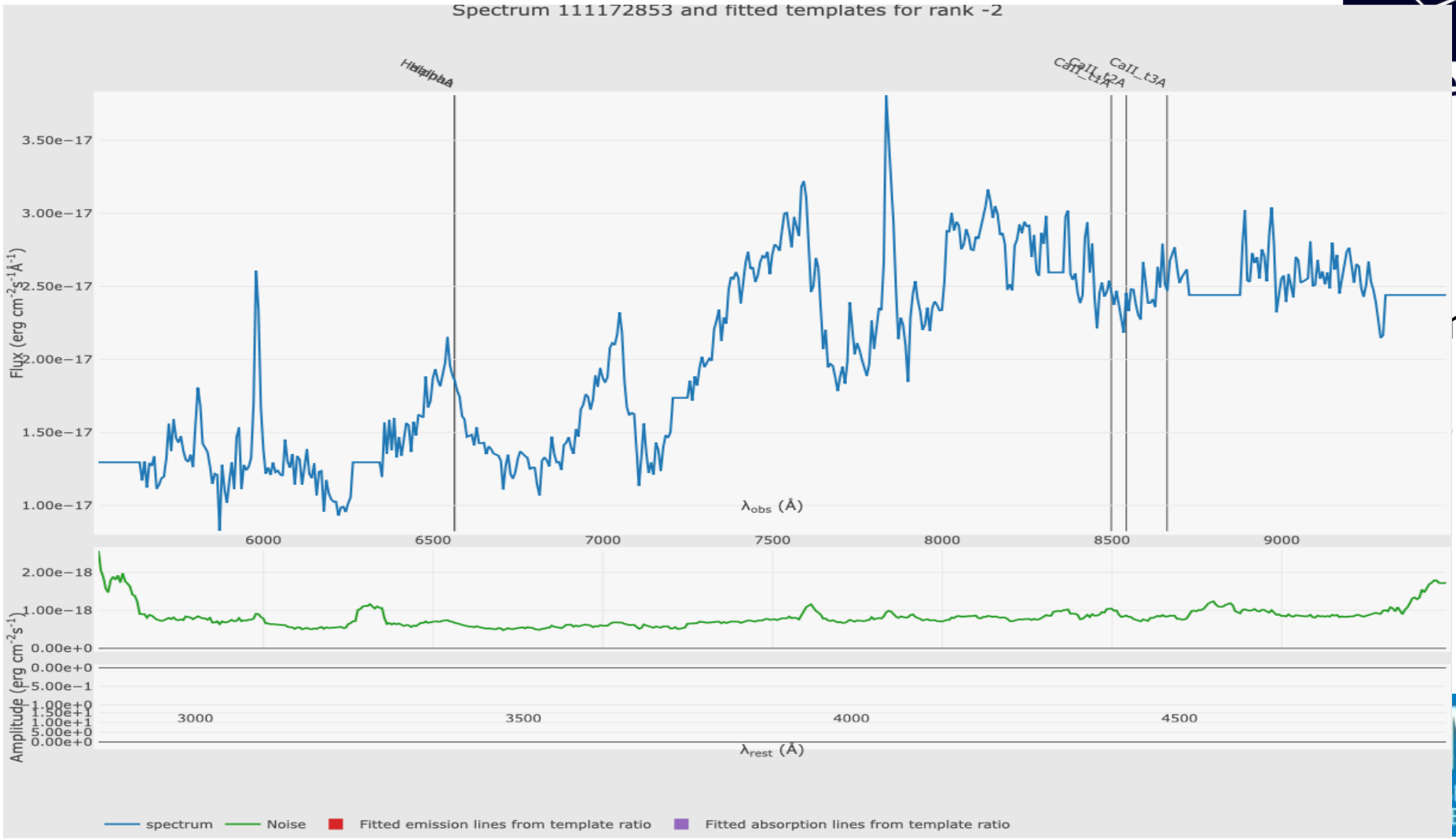


LABORATOIRE D'ASTROPHYSIQUE DE MARSEILLE

Performances on VIPERS data

- Results on the flag 4 spectra : 19658 galaxy spectra
- Caveats
  - Variance in VIPERS spectra is ‘not perfectly’ evaluated
  - There can be superposition of spectra
  - Flags (sometime strongly) depend on the personality/experience
  - Human measurement based on features not available
    2D spectrum
    Feature on the edge of spectrum

Performances on VIPERS data



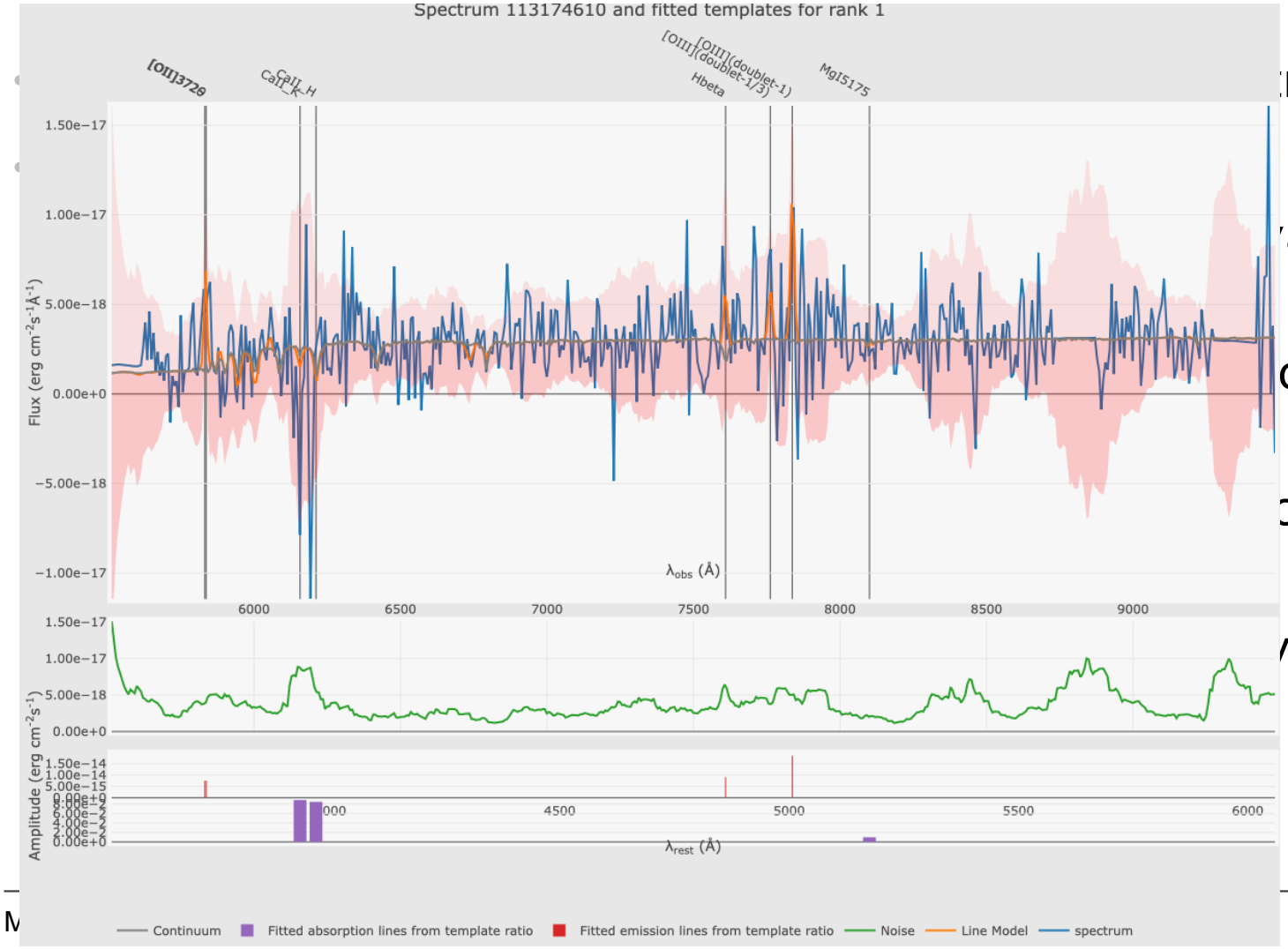
nain

ence

Performances on VIPERS data

- Results on the flag 4 spectra : 19658 galaxy spectra
- Caveats
  - Variance in VIPERS spectra is ‘not perfectly’ evaluated
  - Not a simulation, even with double human check, errors remain
  - Flags (sometime strongly) depend on the personality/experience
  - Human measurement based on features not available
    2D spectrum
    Feature on the edge of spectrum

Performances on VIPERS data



ra

valuated

ck, errors remain

onality/experience

vailable

M

Performances on VIPERS data

- 96.3% success with |dz/(1+z)| < 0.002 (*300 km/s at z=1*)

- 0.4<z<0.53 galaxies represent 20% of sample but 35% of errors
  - Lack of adequate template ?

(Red+ templates were built during VVDS)

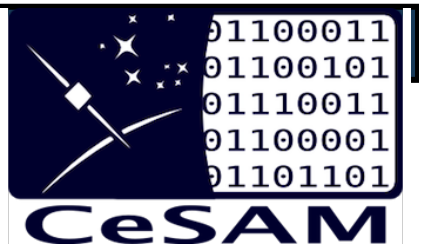
- Merit (measures confidence,)
  - Cutting merit >0.99 excludes 7.5% of the sample but 1/3 of errors
  - Dynamics to be refined

Next steps

- Inprove reliability
- Validate object classifier
- Provide homogeneous measurements on public data available at
  - https://cesam.lam.fr/aspic

- First public release in 2022
- Open to collaborate on your favorite project

Next Marseille Cosmology Conference Edition



From galaxies to cosmology with deep spectroscopic surveys

A tribute to Olivier Le Fèvre

25-29 oct. 2021 Marseille (France

https://specsurveys2021.sciencesconf.org