EZ AND GOSSIP, TWO NEW VO COMPLIANT TOOLS FOR SPECTRAL ANALYSIS

P. Franzetti[1], B. Garill[1], M. Fumana[1], L. Paioro[1], M. Scodeggio[1], Paltani S.[2], and Scaramella R.[3]

[1]INAF, IASF Milano "G. Occhialini", via Bassini 15, I-20133 Milano
[2]Integral Science Data Center, Ch. d'Écologia 16, CH-1290 Versoix (Geneve)
[3]Osservatorio di Roma, via Frascati 33, I-00040 Monteporzio Catone (Roma)

ABSTRACT

We present EZ and GOSSIP, two new VO compliant tools dedicated to spectral analysis.

EZ is a tool to perform automatic redshift measurement; GOSSIP is a tool created to perform the SED fitting procedure in a simple, user friendly and efficient way.

These two tools have been developed by the PANDORA Group at INAF-IASF (Milano); EZ has been developed in collaboration with Osservatorio Monte Porzio (Roma) and Integral Science Data Center (Geneve).

EZ is released to the astronomical community; GOSSIP is currently in beta-testing.

Key words: Virtual Observatory.

1. EZ

EZ (Easy-Z, i.e. easy redshift) is a tool to perform automatic redshift measurement, optimized to be used on huge amounts of data as the ones provided by modern spectroscopic surveys. In this tool all the expertise that more than 50 people have accumulated in measuring the redshift for various tens of thousands galaxies within the VVDS (Le Févre et al. , 2005) and the zCOSMOS (Lilly et al. , 2006) survey has been merged.

EZ has been developed starting from the KBRED and VIZ software packages, originally implemented in IDL by R. Scaramella, which have been extensively used and tested on the data provided by the two surveys.

The main EZ features are the following:

1. It uses a set of templates and a complex combination of algorithms to measure the redshift (see the following section).
2. It can be used interactively or in batch mode.
3. It has been written using the PYTHON language for the graphical part and the C language for the high performance computational part.
4. It has been coded using a modular architecture so that it can be easily upgraded and extended.

1.1. Algorithms

The main algorithm used by EZ (the so called *standard solve*) gets as input values a spectrum, a set of spectral templates and a redshift range. Redshift computation can be divided into four logical steps:

1. compute a cross-correlation between spectrum and redshifted template (spectrum and template are both continuum subtracted).
2. select the redshift values with the highest correlation peaks.
3. for these redshifts, compute a mean-square error between the spectrum and the redshifted template.
4. take the redshift with the lowest mean-square error as final redshift measurement.

However, the core of EZ is its *decisional tree*, which has a higher level of complexity with respect to *standard solve*.

The *decisional tree* balances and exploits the different informations derived by the spectrum, trying to mimic in the best way all the manual decisions that an astronomer takes during the process of redshift measurement. The algorithm tries to recognize emission lines within spectrum; for each line found, it estimates its flux and classifies if either as strong or weak; then searches for redshift values matching at least two emission lines. If no emission line is found, EZ executes a *standard solve* method over the whole redshift range using selected templates.

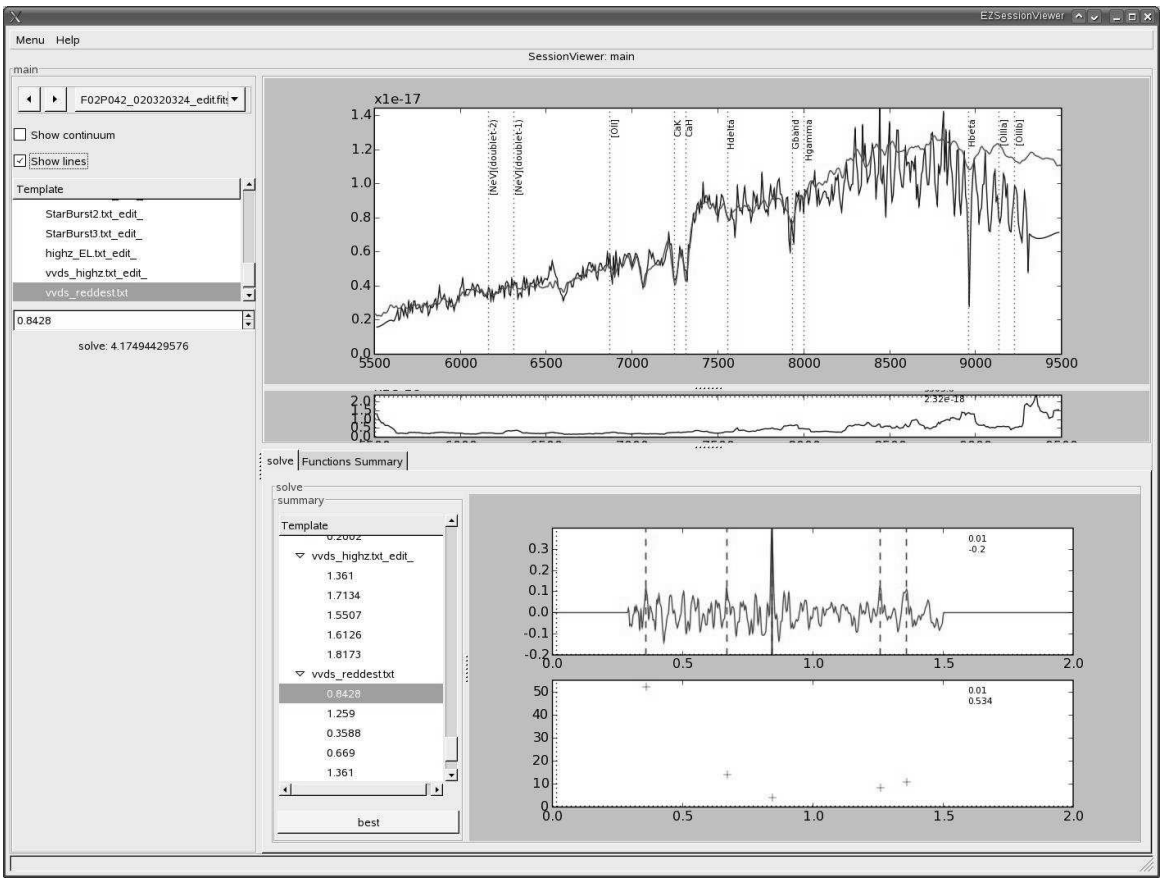

Figure 1. The EZ main graphical interface that allows the user to deeply analyze the details of the measurement. The spectrum and the best template at the best redshift are plotted; most common emission and absorption lines are also plotted. The lower panels contain all the details about the measurement procedure; it is possible to analyze in detail the correlation and the chi square functions. Also other solution with a lower level of confidence are proposed

1.2. Performance

EZ tests have been performed using data coming from the VVDS and the zCOSMOS survey.

The VVDS survey has been carried out with a low resolution grism ($\sim 7.14\dot{A}/pixel$). For our tests we used about 8500 spectra. Overall the data quality is good. The zCOSMOS Bright survey currently contains about 1000 spectra (Lilly et al. , 2006), obtained with a spectral resolution grism of $\sim 2.5\dot{A}/pixel$, in excellent weather conditions.

In both surveys, the redshifts have been computed with a semi-automatic pipeline (Scodeggio et al. , 2005) and then manually checked by astronomers.

Each redshift has been flagged with one of the following redshift quality flag:

- **4**: more than 95% of reliance;
- **3**: around 90% of reliance;
- **2**: around 75% of reliance;
- **1**: around 50% of reliance;
- **9**: redshift based on a single emission line.

During benchmarks, EZ has been used in full automatic mode. The manually measured redshifts have been assumed as the reference values, and EZ results have been compared to such values.

In table 1 we report EZ results on these two data sets. Both for the VVDS and the zCOSMOS sample, the success rate is excellent (90%) on spectra with a secure flag (flag 3 or 4). Considering the whole sample, we note a considerable difference in success rate between the VVDS (73%) and the zCOSMOS sample (>80%). Such discrepancy is mainly due to the overall better quality of zCOSMOS data in terms of observational conditions.

1.3. EZ and the Virtual Observatory

Within EZ, a *standard* interface to the VO has been implemented, similar to the one provided by many other

Flag		All	4	3	9	2	1
N. of spectra	VVDS	8467	2226	2228	321	2494	1195
	zCosmos	1158	616	278	78	89	81
Correct z (%)	VVDS	**73.0**	**95.9**	**89.9**	**56.4**	**63.1**	**23.8**
	zCosmos	**82.4**	**91.1**	**85.2**	**74.3**	**66.3**	**45.7**

Table 1. EZ performance on VVDS data and zCOSMOS Bright data.

tools. Using this interface registries can be queried, and data downloaded to be processed by EZ.

The query of an SSA service in its simple implementation provides either a single spectrum or all the spectra found in a sky region. Such approach is not particularly useful in many astronomical studies: often the selection of a sample is based on parameters which are not directly related to the spectrum itself and cannot be queried through an SSA service. The following example illustrates a typical use case which has been implemented within EZ.

- The user queries a public data set cone service to get all the available data in a sky region and loads them within a tool like TOPCAT.
- He analysis the data to find a sub-sample in which he is really interested.
- TOPCAT sends this sub-sample to EZ, using the PLASTIC protocol.
- EZ queries the VO to find the spectra for the objects in the selected sub-sample and downloads them

This simple work-flow is an example on how combine services and protocols in an efficient way to speed up the scientific analysis of the huge amounts of data which are currently available. This is a crucial point to be addressed for the widespread acceptance and usage of the VO within the astronomical community.

1.4. Future developments

Since the beginning of the software design, we have set the base for a flexible and modular software. In order to improve EZ performance, in the next future we will consolidate and expand EZ computational core, introducing new computational algorithms and studying more sophisticated decisional trees.

Furthermore, to keep updated EZ VO-compliant features, we are following with close attention the development of the VO technology.

2. GOSSIP

GOSSIP (Galaxy Observed-Simulated SED Interactive Program) is a tool created to perform SED fitting, i.e. the fitting of the electro-magnetic emission of an object (the SED, spectral energy distribution) versus synthetic models to find the simulated one that best reproduces the observed data. GOSSIP has been developed to perform this task in a simple, user friendly and efficient way.

The main GOSSIP features are the following:

1. It is able to build-up the observed SED of an object (or a large sample of objects) combining magnitudes in different bands from different instruments and also the optical spectrum.
2. It can load sets of synthetic models. It can load models defined by the user, but it is also able to load directly the output data files produced by the most used synthesis population codes (PEGASE and Bruzual & Charlot).
3. It performs the fitting procedure in batch mode on a single CPU or on a BEOWULF cluster (porting to the computational GRID is on-going).
4. It performs post-processing operations (like computation of absolute magnitudes and Probability Distribution Functions for the models parameters).
5. It is able to send the fitting results to other tools (like, for example, TOPCAT or VISIVO for visualization) using the PLASTIC protocol.
6. As EZ, it has been written using the PYTHON language for the graphical part and the C language for the high performance computational part.

2.1. GOSSIP and the Virtual Observatory

We are currently developing the *standard* interface based on the one we have implemented in EZ. This interface will be able to download single spectra from a SSA service to be fitted against synthetic models. However the narrower the wavelength coverage is, the worse the SED fitting procedure works; using a single optical spectrum the fit is so poorly constrained that it is very often quite useless.

In many cases much more information than the simple optical spectrum does indeed exist in the VO, like aperture magnitudes in several optical and NIR filters, but this information is currently not usable for a correct SED fitting. In fact a detailed information about the aperture at

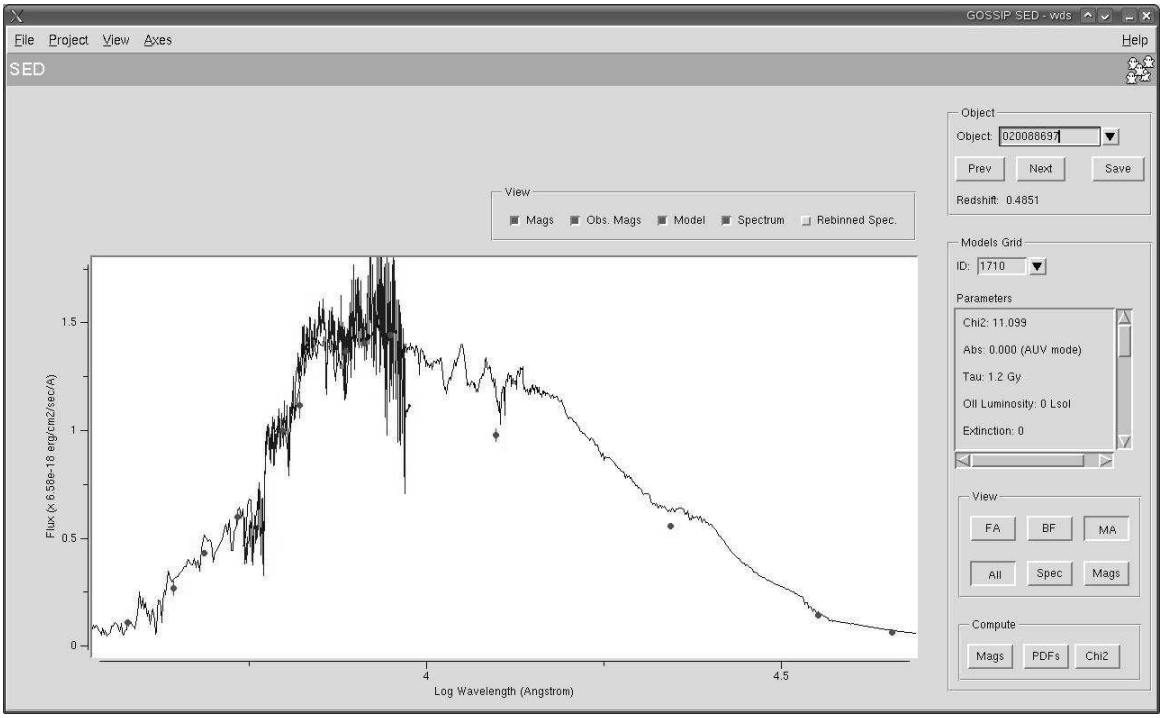

Figure 2. The main SED visualization window of GOSSIP. The SED built by the program composed by the spectrum and the magnitude points is shown. This VVDS SED is composed by data from ESO WFI and SOFI data, VVDS CFHT-12K data, CFHTLS data, UKIDSS data and SPITZER data. The best-fitting PEGASE model is over-imposed to the data and its main parameters are summarized in the right panel

which each data point has been obtained is required to build up a reliable SED (see for details the talk by A. Boselli).

Another complementary approach could be to have the possibility to download in one-shot a complete observed SED provided by some VO service. This approach would be more efficient as the SED would be built by the data provider itself who is expected to process data points providing them in a single common aperture. In any case, both these approaches require a detailed characterization of the magnitudes; in particular the exact shape of the transmission curve associated to each magnitude is absolutely fundamental for the computation of the synthetic magnitudes from the models to be fitted against the data.

While aperture characterization is already provided in the spectral data model (see McDowell et al. , 2007), a specification of the filter response curve is not yet included in it; moreover the SED data model is still under definition and, at the moment, it is not possible to obtain ready made SED. So both the above use cases are not yet achievable within the current VO implementation. GOSSIP is being developed following with close attention the development of the VO technology to be ready, when these issues will be tackled and solved, to exploit all the VO potentialities.

3. CONCLUSIONS

We have developed EZ and GOSSIP, two new VO compliant tools dedicated to spectral analysis. EZ is a tool to perform automatic redshift measurement which is able to reach a global measurement success rate of 80% on good quality redshift sample. It is currently distributed on the PANDORA Group web pages (*http://cosmos.iasf-milano.inaf.it/pandora/EZ.html*). GOSSIP is a tool created to perform SED fitting in a simple, user friendly and efficient way. It is currently in beta-testing and it is foreseen to be distributed by the end of the year.

The features developed within these programs combined with their smart VO interfaces which are able to exploit in efficient way the available VO resources, make them very useful tools for spectral analysis of the large data samples currently available to the astronomical community.

REFERENCES

McDowell, J. et al. 2007, http://www.ivoa.net/Documents/latest/SpectrumDM.html

Lilly et al. 2006, in preparation

Le Févre et al. 2005, A&A, 439,845

Scodeggio et al. 2005, PASP, 117, 1284