Mem. S.A.It. Vol. 80, 429

Memorie della

Matching data and models in the Virtual Observatory

Ben Panter

Institute for Astronomy, SUPA, University of Edinburgh, Royal Observatory, Edinburgh EH9-3HJ, UK, e-mail: bdp@roe.ac.uk

Abstract. The MOPED* algorithm has been used to determine the star formation histories of over 300 000 galaxies from the Sloan Digital Sky Survey (SDSS) Data Release 3. In order to investigate the results from this immense number of galaxies they have been placed in the MyMPASDSS database hosted by the German Astrophysical Virtual Observatory (GAVO). This has enabled us to query simultaneously the MOPED, SDSS and Millennium databases and derive key results without reverting to lengthy file searches on a traditional workstation.

Key words. Galaxy: globular clusters – Galaxy: abundances – Cosmology: observations

1. Introduction

The quality of spectra of the observed light of unresolved stellar populations has reached sufficient accuracy that it is possible to make detailed studies of the physical properties of the stellar populations in these galaxies. An excellent example of this new generation of data-sets is given by the Sloan Digital Sky Survey (Gunn et al. 1998; Strauss et al. 2002; York et al. 2000) at low redshift, not only by the size of the spectroscopic sample (about 10^6 spectra) but by the quality and wavelength coverage of the spectra. At higher redshift the DEEP2 survey (Davis et al. 2003) is providing a similar database, albeit with a smaller wavelength coverage. Future spectroscopic surveys (e.g. Wide Field Multi Object Spectrograph, WFMOS) will yield larger samples at even deeper redshifts. Analysis of each galaxy in such samples is possible using the full spectrum (Cid Fernandes 2005; Heavens et al. 2004; Mathis et al. 2006; Ocvirk et al. 2006; Panter, Heavens & Jimenez 2003, 2004; Tojeiro et al. 2007), but with increasing numbers of galaxies the combination and interpretation of the results becomes difficult. In this proceeding we outline a handful of applications where we have used relational databases to efficiently store and access the data. Although the applications outlined below do not use the Virtual Observatory (VO) as it is envisioned, they outline database methods which are entirely appropriate to the VO approach of queries across distributed databases.

Send offprint requests to: B. Panter

2. The GAVO MySDSSMPA server

As described in Panter et al. (2007) and Panter et al. (2008), the MOPED algorithm has been used to extract star formation and metallicity histories of a magnitude limited sample of about 300, 000 galaxies drawn from the Third Data Release of the Sloan Digital Sky Survey (SDSS DR3; Abazajian et al. 2005).

Since the results for individual galaxies are not always robust, the analysis typically involves 'stacking' the results from some number of galaxies, generally more than a thousand at a time. For much of the early analysis of the MOPED galaxies, ever more complex selections were applied using programs generated in IDL and working with the entire database held in memory. For problems requiring only the MOPED recovered parameters this is possible on a standard workstation, but when analysis requires links to the main SDSS catalogue and more the problem becomes almost intractable, and tracing of errors very difficult indeed. For this reason the MOPED star formation histories and derived parameters (e.g. galaxy mass, supernovae rate, gas recycling fraction etc.) were placed in an SQL database on the GAVO MySDSSMPA server. Alongside this database sit databases containing the SDSS properties for every galaxy and the MPASDSS data archive, allowing concurrent queries which span all databases.

3. Calculating the metallicity over the sky

The SDSS-DR3 spectroscopic footprint covers 3732 sq. deg., roughly 10% of the sky. We use the metallicity history of the galaxies to create maps over this area of the enrichment history at different epochs. We use the HEALPix[1] algorithm to determine equal area patches on the sky, and calculate the mass weighted average metallicities for each patch and time bin as before. The entire process takes place in a Microsoft Transact-Simple Query Language (T-SQL) relational database, with galaxies labeled with their healpix id and then grouped by truncated versions of this label. Figure 1 shows the mass-weighted metallicity maps for our four highest redshift bins, smoothed with a boxcar filter of radius 2°. Over-plotted are the locations of the Brightest Cluster Galaxies from the SDSS C4 catalog (Miller et al. 2005), used to represent the distribution of cluster galaxies on the sky. It is clear by eye that in many regions the crosses follow the regions of higher metallicity.

Using the Virtual Observatory approach by posing the analysis as a database problem we are able to rapidly change the analysis as required, to increase or decrease resolution and even change from maps indicating metallicity to maps giving star formation, dust or some combination of these and other values.

4. Identifying modelling problems

The Virtual Observatory approach can also be applied to investigate the residuals found in spectral fitting. A detailed explanation of the method seen here is contained in Panter et al. (2007). By choosing two samples of spectra, one containing galaxies with a wide range of redshifts and the other with galaxies in a very narrow shell, and then stacking and subtracting their residuals, we are able to deduce where the models fail and identify missing residuals (fig. 2). This average deviation should not be confused with average goodness of fit however, as inspection of the relevant average χ^2 of the samples shows a slightly different story. The models have essentially infinite precision, so there is no penalty associated with rebinning. The converse is true for the spectra, as binning pixels while correctly propagating the error will reduce the standard deviation.

We use the database to determine different families of galaxies and then investigate the missing elements from the models. In particular the effect of alpha enhancement can be investigated by selecting early and late type galaxies and examining the differences between the residuals. The analysis can be rapidly reposed to investigate different models (see fig. 3).

5. The future of MOPED and similar approaches in the Virtual Observatory

The applications discussed above have described an interim database powered analysis method which requires all databases to be mounted on the same server. For the future a

[1] Details of the HEALPix package are available from `http://healpix.jpl.nasa.gov`

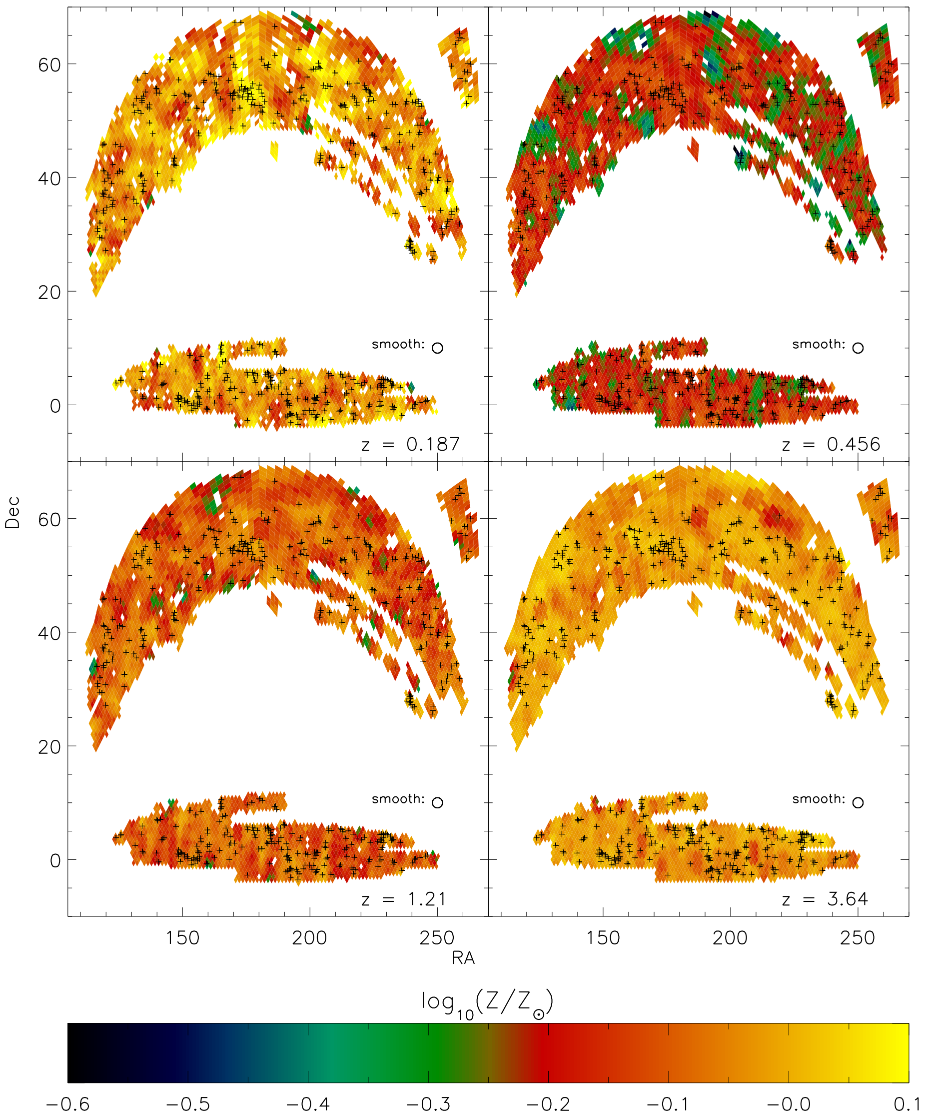

Fig. 1. HEALPix projection of the mass weighted gas metallicity for several look back bins averaged over galaxies in the redshift range $0.0 < z < 0.1$. Galaxies only contribute to a bin if $> 25\%$ of their star formation occurs in that bin. The overlaid crosses correspond to the bright cluster members in the C4 catalog. The metallicities are boxcar smoothed over a 2° radius. The size of the smoothing patch is shown in the bottom right of each panel

key aim would be to have queries run transparently across multiple databases hosted at different repositories. In this way extra information covering properties of galaxies identified

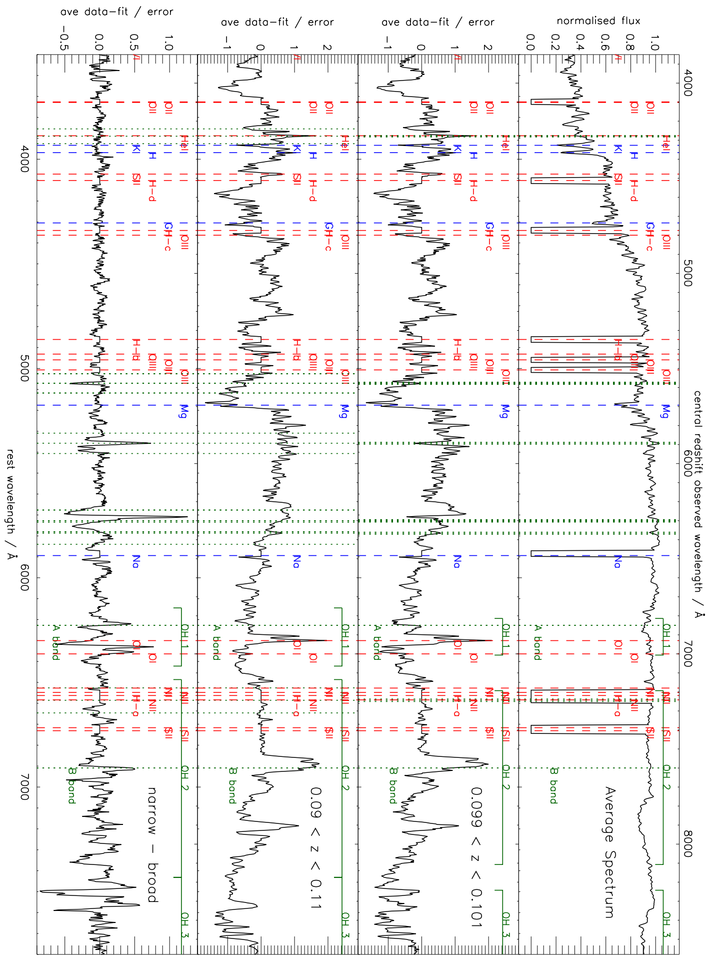

Fig. 2. The mean spectrum and the mean of the residual divided by the error for MOPED fits of DR3 galaxy spectra. The first panel shows the average spectrum for the galaxies in with high S/N and $z = 0.1 \pm 0.01$. The second and third panels show the residual at each pixel divided by the error at that pixel *averaged in the rest frame*. The first plot shows the average for 747 high S/N galaxies within the range $z = 0.1 \pm 0.001$, the second for 7406 high S/N galaxies within $z = 0.1 \pm 0.01$. The third plot shows the first residual set minus the second, and clearly separates areas contaminated from skylines (which are broadened in the second plot by the redshift range). Emission features are labelled in red, absorption blue and skylines green.

in the surveys, or even diffuse fields such as the galactic dust extinction or Cosmic Microwave Background (CMB) in that area of the sky can be folded into the analysis.

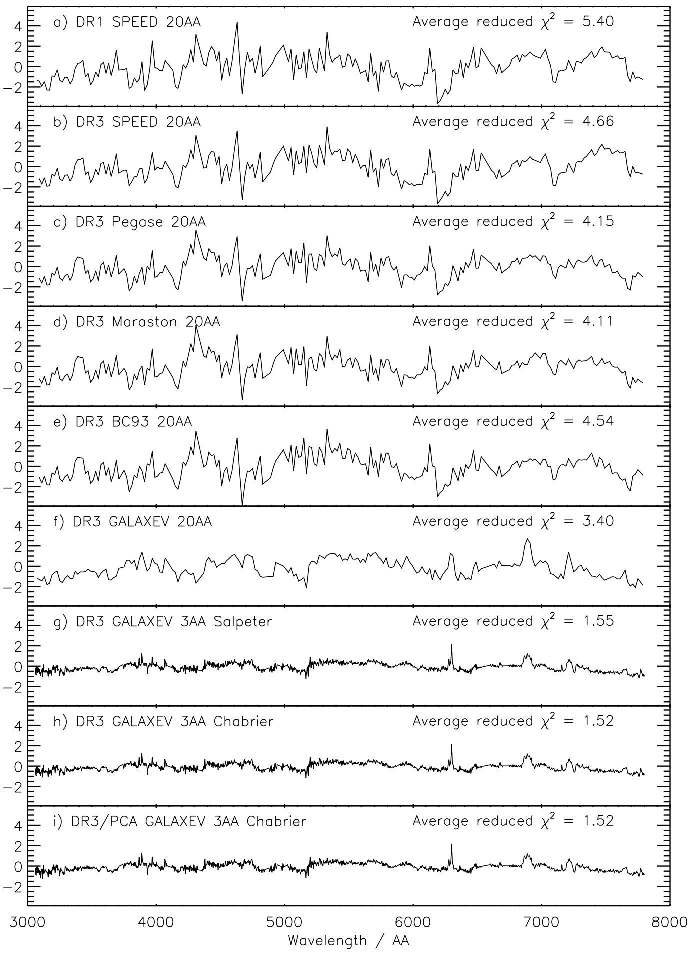

Fig. 3. Average residuals, following the method used to prepare figure 2 but instead using all the Main Galaxy Catalogue spectra in the two plates. The individual panels are labelled with the models and datasets which were used to generate them. From top to bottom, a) DR1 data, Jimenez et al. (2004) SPEED models; b) DR3 data, SPEED models; c) DR3 data, Fioc & Rocca-Volmerange (1997) PEGASE models; d) DR3 data, Maraston (2005) RHB models; e) Bruzual & Charlot (1993) models; f) DR3 data, Bruzual & Charlot (2003) GALAXEV models rebinned to 20Å; g) DR3 data, GALAXEV models at 3Å resolution using a Salpeter (1955) IMF; h) DR3 data, GALAXEV models at 3Å using a Chabrier (2003) IMF; i) DR3 data cleaned using the skyline extraction method of Wild & Hewett (2005). Residuals are averaged in the rest frame.

A further long term aim would be to develop the interoperability of codes such as MOPED, with the methods operating as VO services which can be run on any dataset offered, with various modelling choices. The idealist would require that the algorithm could be run on any galaxy spectra, using any models, with any parametrization - all packaged as a service open to any user. While this is a laudable aim, such freedom removes the interpretation and reality check offered by a skilled operator - just how much data can be reliably extracted from a spectrum, and can the results be trusted to be robust?

While I feel that the publication of data in the Virtual Observatory is essential, I remain to be convinced by the need to have every tool available to run on every dataset and model choice.

Acknowledgements. A portion of BDP's work was supported by the Alexander von Humboldt Foundation, the Federal Ministry of Education and Research, and the Programme for Investment in the Future (ZIP) of the German Government. We acknowledge the use of HEALPix (Górski et al. 2005) and software packages developed by David Fanning (Fanning Consulting), and the assistance of Gerard Lemson and GAVO for access to the Millennium/SDSS Database and advice on SQL. We thank Manuchehr Taghizadeh-Popp for assistance in the identification of SDSS galaxies in the Healpix tessellation scheme. Funding for the SDSS has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, NASA, the Japanese Monbukagakusho, and the Max Planck Society.

References

Abazajian et al. K., 2005, AJ, 129, 1755

Bruzual A. G., Charlot S., 1993, ApJ, 405, 538

Bruzual G., Charlot S., 2003, MNRAS, 344, 1000

Chabrier G., 2003, PASP, 115, 763

Cid Fernandes, R., Mateus, A., Sodré, L., Stasińska, G., & Gomes, J. M. 2005, MRAS, 358, 363

Davis et al. M., 2003, in Guhathakurta P., ed., Discoveries and Research Prospects from 6- to 10-Meter-ClassTelescopes II. Proceedings of the SPIE, Volume 4834, pp. 161-172 (2003). pp 161–172

Fioc M., Rocca-Volmerange B., 1997, A&A, 326, 950

Górski, K. M., Hivon, E., Banday, A. J., Wandelt, B. D., Hansen, F. K., Reinecke, M., & Bartelmann, M. 2005, ApJ, 622, 759

Gunn et al. J. E., 1998, AJ, 116, 3040

Heavens A., Panter B., Jimenez R., Dunlop J. S., 2004, Nature

Heavens A. F., Jimenez R., Lahav O.,2000, MNRAS, 317, 965

Jimenez R., MacDonald J., Dunlop J. S., Padoan P., Peacock J. A., 2004, MNRAS, 349, 240

Maraston C., 2005, MNRAS, 362, 799

Mathis H., Charlot S., Brinchmann J., 2006, MNRAS, 365, 385

Miller, C. J., et al. 2005, AJ, 130, 968

Ocvirk P., Pichon C., Lançon A., Thiébaut E., 2006, MNRAS, 365, 46

Panter B., Heavens A. F., Jimenez R., 2003, MNRAS, 343, 1145

Panter B., Heavens A. F., Jimenez R., 2004,MNRAS, 355, 764

Panter B., Thesis, 2005. Availiable at the Edinburgh Research Archive at http://hdl.handle.net/1842/774

Panter, B., Jimenez, R., Heavens, A. F., & Charlot, S. 2007, MNRAS, 378, 1550

Panter, B., Jimenez, R., Heavens, A. F., & Charlot, S. 2008, arXiv:0804.3091

Salpeter E. E., 1955, ApJ, 121, 161

Strauss et al. M. A., 2002, AJ, 124, 1810

Tojeiro, R., Heavens, A. F., Jimenez, R., & Panter, B. 2007, MNRAS, 381, 1252

York et al. D., 2000, AJ, 120, 1579

Wild V., Hewett P. C., 2005, MNRAS, 358, 1083