ASTR4004/ASTR8004
Astronomical Computing
Assignment 3

Christoph Federrath, Tiantian Yuan, David Yong

due Tuesday, September 20, 2016, 09:15am

1 Image processing with IDL

See attachment 1 for details.

2 Statistics in IDL

See attachment 2 for details.

Please send your scripts/responses/produced figures for these two IDL projects to `mailto:christoph.federrath@anu.edu.au` *by the assignment deadline.*

Dealing with low S/N high-redshift Integral Field Spectroscopic (IFS) Data

After this exercise you would:

- Get more comfortable with IDL coding
- Learn how to manipulate 3D datacube
- Grasp the concept of co-adding data and characterising the errors of the co-added datacube

Background of the data you are going to deal with:

In 2011 we reported the metallicity gradient of a gravitationally lensed high-redshift spiral galaxy. The data you are going to use is from Yuan et al. (2011, ApJ, 732). The data were collected from the adaptive-optics aided Integral Field Spectrograph (IFS) instrument OSIRIS at KECK. Don't worry, we are not going to let you reduce the raw data. You are mostly going to experience some of the fun processes that we have to go through in order to dig out the signal from noise-dominated high-redshift data. This project will give you a good feeling of what the challenges are in analysing low S/N, high-redshift IFS data. It is also a good practice for your IDL coding skills.

The assignment datacube can be downloaded from here:

http://www.mso.anu.edu.au/~yuantt/IDLcourse/highz_data/

Steps:

1. You are given 18 3D datacubes of a lensed spiral galaxy that are obtained from 18 individual exposures in the observations. Read in one of these cubes in IDL (note that these are .fits files). Learn how to extract the spatial and wavelength information from the datacube.
2. For the original datacube, which dimension is the wavelength, which is RA and which is Dec? How did you come up with the answer?
3. Now transpose each datacube such that the first dimension is RA, the 2nd dimension is Dec, and the third dimension is the wavelength.
4. Please install ds9 or qfitsview if you want a more intuitive look at the IFS data. After you understand the IFS data, you should be able to generate a wavelength-collapsed (see hint *h1 below) 2D image for any of the 18 datacubes.
5. The redshift of the galaxy is about 1.489. Now generate a 2D H-alpha image from each datacube? Also please describe how you generated the H-alpha image.
6. Co-add the 18 datacubes and generate a final co-added 3D datacube that has the best S/N in the H-alpha emission line detection. (see hint *h2 for help)
7. (optional) Calculate the associated error cube from the above steps.

Please send us your IDL codes as well as a short description of how you did each step (you can add those descriptions in your IDL codes)

=============

Hints:

*h1: wavelength-collapsed: take the median in the wavelength direction

*h2: the datacube is dominated by sky noise, cosmic rays, and instrument background (imperfect flat-fielding). To get the best S/N in emission lines (in our case H-alpha), please fit a linear function to the continuum in the spectral direction for each spatial pixel, and subtract it out from each spaxel (pixel in wavelength direction). This will help you subtract the imperfect flat-fielding for the price of losing the first-order continuum of the science object (however, we don't care so much about the continuum of the science object anyway).

*h3: (optional) There may be some spatial shift among the 18 data cubes, i.e., the center of the spiral galaxy could have some offsets among the 18 datacubes. If you can find a way to determine the offset, the S/N of the co-adding can be improved by correcting for the offsets.

ASSIGNMENT 3, Attachment 2 – Astronomical Computing 2016

How many stars do we need to observe to distinguish between different model predictions of the metallicity distribution function?
Or more generally …
How many objects do we need to observe to distinguish between different probability distributions? (redshifts or other galaxy properties, black hole masses, exoplanets etc.)

Our clever and friendly neighbourhood theorists have models which predict the so-called "metallicity distribution function" (MDF), i.e., the number of stars at a given metallicity.

Here is one example from Salvadori (adsabs.harvard.edu/abs/2007MNRAS.381..647S)

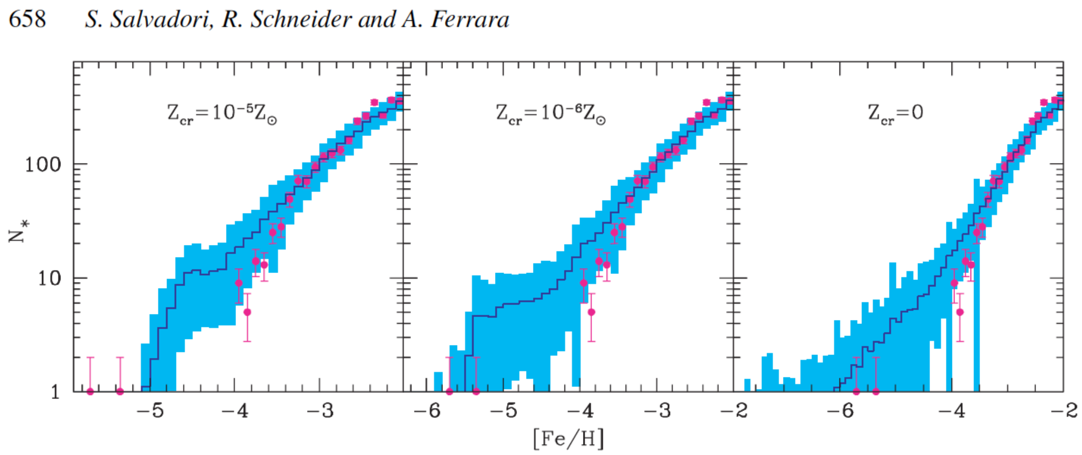
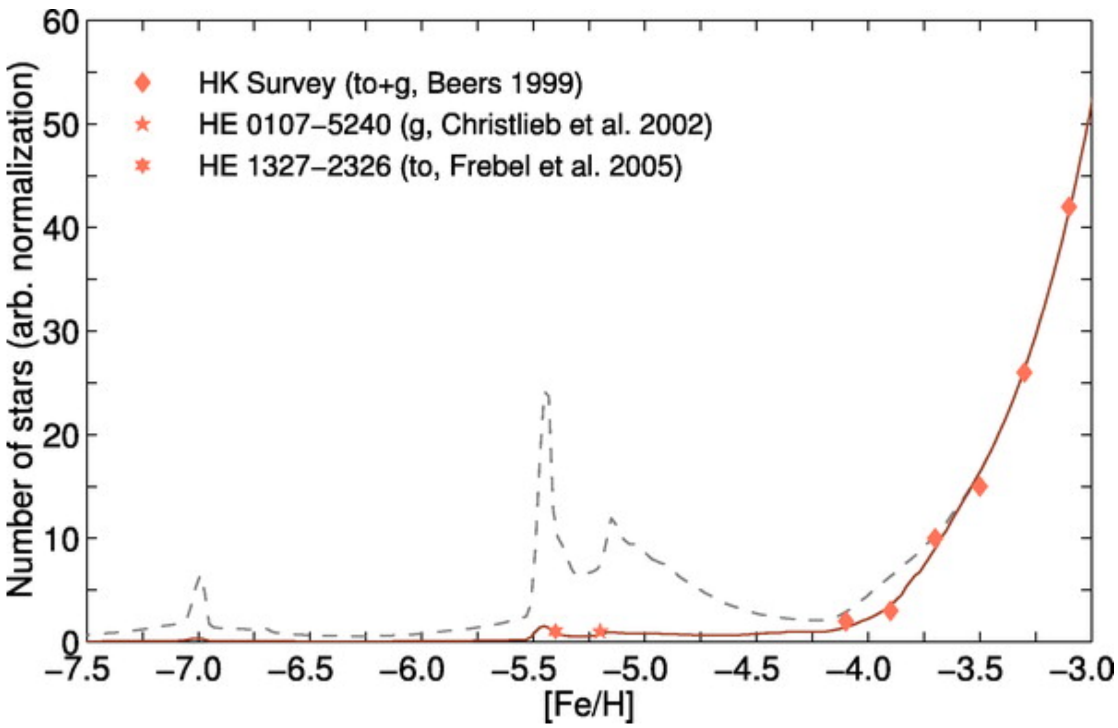


Figure 7. The same as the left-hand panel of Fig. 6 but for values of $Z_{cr} = 10^{-5}$, 10^{-6} and 0 $Z_\odot$.

Here is another example from Karlsson (adsabs.harvard.edu/abs/2006ApJ...641L..41K)



The input physics and assumptions in the models control the shape of the predicted metallicity distribution function. In particular, the so-called "critical metallicity" is a key quantity; it is the metallicity below which low-mass star formation is not possible and we would not expect to observe any long-lived low-mass stars today below that value (if you care about this, here is a review http://adsabs.harvard.edu/abs/2004ARA%26A..42...79B). Look again at the figure from Salvadori and notice how the critical metallicity affects the tail of the metallicity distribution function.

We would like to know how many stars do we need to observe in order to discriminate between these models. Indeed, quantifying the necessary sample size to test a hypothesis (discriminating between model predictions in this case) is a crucial ingredient for successful proposals (observing/funding/jobs). Recall that 6-10m class optical/IR facilities (e.g., Magellan, Keck, Gemini, Subaru, VLT) cost about AUD$1 per second(!) to operate and that the competition for telescope time is fierce.

Starting IDL

Here is a possibly useful introductory document for IDL:
https://www.atmos.colostate.edu/programming/IDL/idl_week1.pdf

You can search/find IDL routines here:
http://www.harrisgeospatial.com/docs/funclisting.html

In the dropbox link below, there is a **very** simple example (dy_example.pro)
To run this example, type ".r dy_example" from command line within IDL

The assignment

1. Here are the metallicity distribution function predictions in ASCII format.
https://www.dropbox.com/sh/jd0y3msb6bvjl6i/AACNAv9Y4-lbTrIRgQKggGAta?dl=0

For the Salvadori models, the columns are
column 1 : [Fe/H]
column 2 : N*([Fe/H])
column 3 : SigmaN*([Fe/H])
For the Karlsson model, the columns are [Fe/H] and #.

Here is a relatively easy way to read ASCII files in IDL:
http://idlastro.gsfc.nasa.gov/ftp/pro/misc/readcol.pro
And some information on IDL array definitions:
http://www.harrisgeospatial.com/docs/Creating_Arrays.html

2. Randomly draw 10 stars from each of these distributions in the metallicity regime [Fe/H] ≤ -4.0. Note that these are arbitrary distributions (i.e., neither Gaussian nor uniform). Look at this document if you don't know how to generate random numbers from an arbitrary distribution:
www.ece.virginia.edu/mv/edu/prob/stat/random-number-generation.pdf

Generating random numbers in IDL:
http://www.harrisgeospatial.com/docs/RANDOMN.html
http://www.harrisgeospatial.com/docs/RANDOMU.html
Are random numbers random?
http://www.idlcoyote.com/code_tips/randomnumbers.html

3. Compare any two distributions (e.g., Karlsson vs. Salvadori Z_crit=0) and quantify the likelihood that the data are drawn from the same distribution. Use your favourite statistical test; a well-known one is the Kolmogorov-Smirnov (KS) test:
http://idlastro.gsfc.nasa.gov/ftp/pro/math/kstwo.pro

What are the probabilities? Do this for all combinations of model distributions.

4. How reliable are those results? (If you are unsure, google "shot noise" or the Poisson distribution.) Repeat the above exercise 100,000 times and plot the distribution of the probabilities.

For what fraction of the realisations can you reject the null hypothesis at the 90% confidence level?

Writing loops in IDL:
http://www.harrisgeospatial.com/docs/FOR.html

Plotting in IDL:
http://www.harrisgeospatial.com/docs/PLOT_Procedure.html
http://idlastro.gsfc.nasa.gov/ftp/pro/plot/plothist.pro

5. Repeat the above steps but randomly drawing 100, 200, 300, 400, etc., stars from each distribution. Keep going until you can reject the null hypothesis at the 95% confidence level in 95% of realisations.

By finishing this assignment you will be able to tell how many stars we need to observe in the metallicity regime [Fe/H] ≤ -4 in order to discriminate between the Salvadori and Karlsson models, and, between the various Salvadori models with different critical metallicities.