Data Analysis Project 1

Due Friday, 06 Apr, at the beginning of class. Turn in the Python/IDL/Matlab code for all of your work.

This is the first of three data analysis projects. Projects will be graded on a scale of 0 to 3, and you cannot drop a project grade. However, you may resubmit your project to improve your grade.

There will be a progress check on Friday, 3 Apr. I will ask one or more of you to report on your methdology and results through section 2.1 (determining readout noise).

Reading assignment: Read Mclean, Ch. 3.2 (Telescope design), Ch. 7.2 (Basic Principles of CCD's), Ch. 8.5 (Noise sources in CCD's), anc Ch. 10.3 (Principles of Image Analysis and Processing).

Data Reduction and Noise Analysis of Visible Wavelength CCD Data

1 CCD Image Basic Data Reduction

For this homework, you will perform data reduction and noise analysis tasks on CCD images of the Ring Nebula (M57). The CCD images were taken by John Bally, using a C-14 telescope at his Breckenridge observatory. Since there are a routine set of data reduction tasks performed on CCD data, there exists specific software, IRAF, to perform these routines. However, the goal of this homework is not to learn IRAF, but to really understand what is being done "under the hood," and understand CCD detector performance, so you will perform the usual IRAF data reduction tasks, but using Python/IDL/Matlab and writing your own code. The flexibility of Python/IDL/Matlab will also allow you to examine the data in any way you wish, including characterizing the image noise.

Download the CCD data, which are 13 FITS files, from the `Homework Assignments/data` folder.

The data files consist of 3 bias frames, 3 dark frames, 3 flat frames, and 4 astronomical images of M57 using various color filters.

The first part of the homework consists of performing standard data reduction tasks on the CCD data. You will perform the tasks described in the IRAF tutorial http://joshwalawender.github.io/IRAFtutorial/index.html in Section V, "Basic Reductions for Imaging Data." Read this section of the tutorial for a more detailed explanation of the reduction tasks and why they are performed.

1) Read the FITS header info into Python/IDL/Matlab using the appropriate Python/IDL/Matlab procedure (for IDL users this is located in the Goddard astro library). Identify the filter and integration time for each of the 13 files.

2) Read the FITS image data into Python/IDL/Matlab. What data type are the image array pixel values, and what range to they have? Have a look at the images, in particular the raw science frames. (FYI: not all FITS readers and FITS display programs have the same orientation convention.)

3) Follow the tutorial, step B, to combine the 3 dark images to form a dark_master image, using the median value for each pixel. Do the dark images all have the same exposure time, and is this exposure time the same for the raw M57 frames?

4) Perform a similar combination of the 3 bias images to form a bias_master image.

5) Now, combine the 3 flat images to form a flat_master image. To do this:

- Subtract the bias_master from each flat image (Why? And why not use the dark_master instead?)
- Normalize each flat image by its mode. (Again, why? And how do you find the mode of the pixel

values using Python/IDL/Matlab?)

- Combine the 3 images by taking the median value from the 3 images.
- Again, normalize the combined image by its mode (and again, why?) to form a flat_master image.

6) Form a reduced image of each of the four M57 frames by subtracting the dark_master, and dividing by the flat_master. Plot these images, with appropriate intensity contrast to see details of the nebula structure and background stars.

2 Noise Analysis

Calculate all noise contributions in units of DN, unless otherwise stated.

2.1 Readout Noise

First, characterize the readout noise (RON) by examining the 3 bias frames.

1) Find the mean and standard deviation of the pixel values in the first bias frame. Why are both of these non-zero?

2) Plot an image of the first bias frame, and histogram the pixel values (pick a subset of the images if your computer can't handle 3 megapixels). You will notice that the distribution looks mostly Gaussian, but there is a long tail, with some very high outliers (due to warm pixels on the chip). Some of these points in the tail are systematic, not statistical, and we do not want to include them in the noise analysis. To get around this, histogram the difference image between two bias frames. Why does this distribution have less of a tail now? Find the standard deviation of the differenced bias frame, and from these, infer the standard deviation of a single frame.

3) Estimate the readout noise contribution to DN uncertainty, if a three-frame bias average is subtracted from the astronomical image. (You can assume an unweighted mean average instead of a median average.)

2.2 System Gain

Recall that the system gain G is the ratio between detected photons (electrons) and DN, usually given as "electrons per DN." Knowing the gain is important to understand the system efficiency, and in characterizing Poisson noise due to photons sky noise and celestial sources. There are three ways to determine system gain.

- Obtain from the FITS header (keyword E_GAIN).
- Get the value from the chip manufacturer's data sheet.
- Calculate gain from the flat exposures, assuming these are Poisson noise dominated. The gain G in electrons per DN can then be calculated from the mean and standard deviation of a subsection of the flat, using the relations
$$\sigma_e = G\sigma_{\mathrm{DN}} = \sqrt{N}$$
where σ_e is the standard deviation in units of electrons, σ_{DN} is the standard deviation in units of DN, and the mean electron count N is given by
$$N = G\mu_{\mathrm{DN}},$$
where μ_{DN} is the mean DN count, all assuming Poisson statistics. Then the gain G is simply
$$G = \frac{N/G}{N/G^2} = \frac{\mu_{\mathrm{DN}}}{\sigma_{\mathrm{DN}}^2}.$$
Make sure you use a large subsection of the image to reduce uncertainty.

Use the last method to calculate the gain the flat exposures. To do this,

1) Subtract two flat frames to create a difference image, in order to eliminate anomalous hot pixel values.

2) Estimate the variance of a single flat by taking the sample variance in a clean subsection of the difference image. Remember to multiply the differenced image sample variance by the appropriate factor to estimate the variance of a single flat frame, σ^2_{DN}. Verify that you can neglect the effects of readout noise (RON), by comparing σ_{RON} to the standard deviation in your image subsection.

3) Find the sample mean for the same subsection of one of the flat frames, to estimate μ_{DN}.

4) Estimate the gain G using the method outlined above. Estimate your uncertainty for G by repeating the procedure for several different subsections of the frames, and/or selecting different flat frames with which to do the analysis.

5) Compare your result for G to the gain given in the FITS file header. Do they agree, within the uncertainty?

2.3 Characterizing Dark Current

There is also dark current, which is a pixel count offset that increases linearly with time due to thermal noise in the detector. There is uncertainty associated with dark current which contributes to the overall noise budget.

1) Examine one of the dark frames by displaying it and creating a histogram of pixel values, and compare it with one of the bias frames you analysed. Calculate the mean and standard deviation of the dark frame (or a subsection of it), and compare with the values from the bias frame (or same subsection of the bias frame). Plot the histogram in a way that shows hot pixels, and quantify the numbers of pixels in different parts of the distribution.

2) Now subtract two dark frames and calculate the mean and standard deviation. Why is the standard deviation so much lower? What are the contributions to this noise from the readout noise and dark current? To calculate this, assume tha the readout and dark current contributions are independent and each have a Gaussian distribution.

3) Estimate the dark current contribution to DN uncertainty, if a three-frame average is subtracted from the astronomical image. (You can assume an unweighted mean average instead of a median average.) Remember, there is a dark current contribution both from the astronomical image and from the three-frame dark master that you subtract from the astronomical image.

2.4 Sky Noise

Airglow at visible wavelengths also contributes to image noise, by simply contributing a photon background, with count statisics that follow a Poisson distribution.

1) Locate a dark subsection of your reduced image. Histogram the data values from the dark section. If there are large outliers, reject them. Then, find the mean and standard deviation of the data in the dark subsection.

2) Subtract the readout noise and dark current contributions to the standard deviation. (You already subtracted the readout and dark current contributions to the mean in your data reduction process.) Is there residual noise? Assuming this residual noise is due to sky noise, calculate the sky noise that you would expect, given the mean sky flux value in your dark subsection. Is the sky noise what you would expect, given Poisson statistics?

3 Photometry Uncertainty

Lastly, estimate the uncertainty on photometry of stellar flux in the image. Take, for example, the bright star just the the right of M57 in the image (pixel coordinates 480, 1440). Estimate the total statistical uncertainty in the peak amplitude (in units of DN), taking into account contributions from:

- readout noise,
- dark current,
- sky noise,
- photon count statistics from the source itself.

What is the dominant source of noise? What other contributions do you expect to uncertainties in photometry?

Compare your uncertainty estimate to the marginal uncertainty in amplitude parameter A returned by your Gaussian beam fitting routine on this star. Do they agree? Why or why not?