Observational Astrophysics
19. Data Processing

Rodolfo Smiljanic
Autumn/Winter 2021/2022

Nicolaus Copernicus Astronomical Center
Polish Academy of Sciences
ul. Bartycka 18
00-716 Warsaw, PL

E-mail: rsmiljanic@camk.edu.pl
Office: 115

`http://users.camk.edu.pl/rsmiljanic`

1 Introduction

After using the telescope and instrument of choice to record an image or spectrum with the detector, you will find yourself with a data file that needs to be analyzed. Before doing science with the data, you will need to perform a few processing steps to remove certain electronic effects from the data. Your goal is that only the signal from your source remains. This processing step is called “data reduction”.

The initial basic steps of data reduction are the same for photometry or spectroscopy. These steps involve the calibration frames called “bias”, “dark”, and “flat field”. A few additional steps might be in order, depending if some problems appear in your data, such as bad pixels or fringes. Further steps will then be needed to measure and calibrate the magnitudes, in case of photometry. For spectroscopy, one will at least still need to trace the spectrum on the CCD, extract it, and assign to it a wavelength scale. You might also need to apply a Doppler correction, perform flux calibration, and remove telluric lines (depending on your science interests). Sky background correction is also needed, although in general this step is performed together with the measurement of the instrumental magnitude (for photometry) and the extraction of the spectrum (for spectroscopy). Things can also get a bit more complicated if your data is complex, like with IFU data cubes or echelle spectra.

Here, I will summarize the basic steps of data reduction that are common to photometric and spectroscopic data. We will also look at the definition of signal-to-noise ratio (S/N). For the other data reduction steps, I will refer to additional reading material that you can check depending on your interests.

2 Read these texts

The basic equations and concepts of signal-to-noise ratio can be read in Section 1.5 of “The design and construction of large optical telescopes” (Bely 2003)[1]. A somewhat extended discussion can then be found in the notes on signal-to-noise ratio by Mike Bolte, Lick Observatory, from his course on “Modern Observational Techniques” in `https://www.ucolick.org/~bolte/AY257/s_n.pdf`

---

[1] `https://link.springer.com/book/10.1007/b97612`

Then we need to read about characterization and calibration of images obtained with CCDs. For that, have a look at Chapter 9 of the book “Electronic Imaging in Astronomy” by (McLean 2008)[2].

A quick introduction about the FITS format (and visualization tools) might also be in order. For that, please read the Chapter on “Data and Data Archives” from the book “Software Systems for Astronomy” (Conrad 2014)[3].

3 Summary of concepts

- The signal recorded by the detector is a combination of signal from your source, from the sky background, from the detector itself, and from the electronics used with the detector. All background sources can be, in principle, characterized separately and then be subtracted from your data. What can not be corrected though are the fluctuations of all these background sources, the “background noise”.

- For the data obtained with a CCD, the signal is a counting of electrons. If the events releasing electrons are happening at a constant rate and the events within an interval of time are independent from those of the next interval of time, then these events can be described by Poisson statistics.

- The interesting property that follows from this is that the standard deviation, σ, is the square root of the number of events (S) in the signal: $\sigma = \sqrt{S}$.

- The readout noise of a CCD has two components. One appears during the conversion of the analog signal to a digital number. This conversion is not perfect but will produce a statistical distribution around a mean value (i.e., even if the same pixel, with the same charge, could be readout several times, each time a slightly different value would be obtained). The second component is the addition of spurious electrons in the process, creating further fluctuation in the value that is read. The readout noise is given in units of electrons per pixel.

- The speed of the readout influences the level of the readout noise. The flow of the current through the amplifier can cause thermal variations within the amplifier, creating unwanted electrons. Slower readout introduces less noise, but increases the overhead time. So the decision of which readout speed to use depends if one is observing faint or bright targets (i.e., if the signal is weak or strong in comparison to this source of noise) and how much time one can spare for the readout. Typical readout options in modern astronomical CCDs will produce readout noise between <1-5 electrons per pixel (rms).

- Dark current is a name for the electrons generated within the semiconductor (Si for CCDs) because of its own thermal properties. The thermal generation of electrons is a function of temperature, and to reduce this signal is why CCDs for astronomical use are cooled down to very low temperatures. The problem is of course that electrons generated in this way can not be separated from the electrons generated from photons coming from your source. The noise given by the dark current then provides a noise floor of the CCD. The dark current in modern liquid-nitrogen cooled CCDs is of at most a few electrons per pixel per hour. Dark

---

[2] https://link.springer.com/book/10.1007%2F978-3-540-76583-7

[3] https://link.springer.com/book/10.1007/978-1-4614-7058-8

noise is no longer an important problem for CCDs, but may still be significant for detectors used in the infrared.

- To express the reality of the signal we compare the level of that signal (S) with the level of the noise (N) through the "signal-to-noise ration" (S/N or SNR). This gives the strength of the signal in units of its standard deviation.
- The typical, simple version of the S/N can be expressed by an equation with this form:

$$S/N = \frac{N_*}{\sqrt{N_* + n_{\rm pix}(N_S + N_D + N_R^2)}}, \tag{1}$$

  where N_* is the total number of electrons from the source within the detector area where the measurement is being done, N_S is the total number of electrons received from the sky background per pixel, N_D is the total number of electrons received from the dark current per pixel, N_R is the readout noise per pixel, and $n_{\rm pix}$ is the number of pixels.
- You might come across different ways of writing the S/N equation (for example, instead of the total number of electrons one can write the rate per second times the exposure time). In addition, other factors might be included explicitly, such as errors introduced in the estimation of the background (see Merline & Howell 1995)[4], issues with digitization noise, and factors from additional processing of the data.
- For bright sources, the S/N equation reduces to S/N $= \sqrt{N_*} = \sqrt{R_* t}$ (where t is the exposure time and R_* the rate of photons from the source).
- If the source and the background are faint, the noise from the detector might dominate. This situation is said to be "detector-noise limited". The S/N equation reduces to:

$$S/N = \frac{N_*}{\sqrt{n_{\rm pix}(N_D + N_R^2)}}, \tag{2}$$

  In this case the S/N is proportional to the number of photons collected from the source and the S/N will improve linearly with time. This is a typical case for high-resolution spectroscopy, as the background noise per spectral element decreases with increasing resolution.
- There is also a "background limited" case, when the target is weak and sky background strong, dominating the noise. The S/N equation reduces to:

$$S/N = \frac{N_*}{\sqrt{n_{\rm pix} N_S}} = \frac{R_* t}{\sqrt{n_{\rm pix} R_S t}} \propto \sqrt{t} \tag{3}$$

  where R$_S$ is the rate of photons from the sky background. Here the S/N scales with the square root of the exposure time. The only chances to get outof this situation are either to use a larger telescope, improved image quality (e.g., with adaptive optics), or try to avoid the strong background (e.g., strong airglow lines).
- The bias (or zero) frame is used to measure the signal value produced by the empty, unexposed pixels of a CCD. Even if that value is zero electrons, when this is readout, a fluctuation will be introduced. To avoid negative numbers (and thus avoid using one bit to register the signal of the number), the CCD is setup with a positive offset from zero for each accumulated image. This offset is the bias level.

---

[4] `https://ui.adsabs.harvard.edu/abs/1995ExA.....6..163M/abstract`

- The bias level can be measured with the overscan region. These are not real pixels from the CCD, but are generated sending additional clock cycles to the CCD output electronics. The correction then consists in determining a mean (or median) value of the counts within the overscan region, and subtracting this "pedestal level" from the other frames.

- Alternatively, one can use the whole bias frame for that. The bias frame will provide information on any two-dimensional structure within the bias level. The subtraction is then done pixel by pixel. For that, one ideally constructs an average (or median) bias frame from $\sim$10 (or more) individual bias images. Averaging will also take care of other effects that might affect pixels in a single image (e.g., cosmic ray hits, readout noise). Most modern CCDs show little to none bias structure, so the overscan correction might be all you need.

- Variations in the zero level of a CCD can happen with time. Traditionally, one would take several bias images before and after an observing run. Nowadays, it might be enough to take bias frames every few days.

- Dark frames are taken with the shutter closed (i.e., the CCD is not being exposed to light) but for a time period that is equal to the time used to observe the source. Dark frames are used to take into account the signal introduced by the thermal properties of the detector. Not always the dark increases linearly with time, so just scaling a shorter exposure to a longer time might not be adequate. Nowadays, the dark current of modern liquid-nitrogen cooled CCDs is very small (a few electrons per hour per pixel), so dark correction is usually ignored.

- If dark correction is needed, the usual approach is again to take several images and average them in some form. Note that the bias level will also be included inside the dark, so a separated bias correction can in principle be skipped.

- Flat field frames are used to correct pixel-to-pixel sensitivity variations in the CCD (which are of the order of less than a few percent). It also helps to correct non-uniform illumination of the detector (e.g., caused by dust in the optical elements). Flat field correction is hardly ever done perfectly. The general idea is to uniformly illuminate the CCD to obtain its response function. In practice, however, obtaining uniform illumination to a high level is hard. In addition, the pixel-to-pixel variations are wavelength dependent. Thus, the illuminating source should have a spectral distribution similar to that of what was observed.

- For the flat field frames, you will want to get high levels of signal (without saturating the CCD). A S/N of 1000 will allow you to correct variations of the 0.1% level. As before, the typical approach is to take several flat field frames and average them. Frames are needed for each filter, wavelength region, or instrumental setup used for the science observations.

- The flat field correction consists in dividing the science frame by the (normalized) flat field. One is usually looking for a uniform background of light that matches the color of the dark background sky and that illuminates the CCD in the same way as the background sky. There a few different variations of the flat field frame:

    1. Dome flats: a big white screen is mounted somewhere inside the dome. Lamps are used to illuminate that screen that is then observed by the telescope. A variety of lamps with different color temperatures might be available. Dome flats might not provide the best correction. The lamps might be poor matches for the background and, because the screen is not at infinity, the illumination of the focal plane might be different from that

obtained with the sky. The advantage is that the images can be taken at any time, even during the day with a close darkened dome.

2. Twilight flats: these are taken during dusk or dawn, by pointing the telescope to a blank area of the sky. The illumination is essentially the same as with observing the dark sky. Also, the spectrum of the twilight is similar to the dark sky (zodiacal light). However, the brightness of the twilight is not uniform. There is gradient with stronger light towards to where the Sun is. This might not be a big issue for observations of a small field of view, but might be significant for observations taken with large mosaics. In addition, the amount of time available during twilight might be enough for taking just a few frames with counts that are maybe not as high as possible with dome flats.
3. Sky flats: these are taken during the night, observing a patch of dark sky. They offer the perfect colour match to the background and illuminate the CCD in exactly the same way. You will however need to use night time to take these calibration images and they need long exposures to build up signal to noise. Several images are needed and they should be processed with some type of spatial filtering to remove the contribution of objects in the area.
4. Lamp flats: these are used for spectroscopy. A high-intensity lamp is used to illuminate the slit, and then the spectrograph, in a similar manner to that of astronomical objects.

- Fringes can appear in your image because of interference of light reflected within the CCD. Fringing is not a problem in most of the wavelength range used in observations with CCD. However, fringes might appear for observations in the red part of the optical spectrum; observations done using narrow-band filters; and when observing strong narrow emission lines. If fringing is caused by atmospheric OH emission, it can be quite variable in short time scales. One problem for the removal of fringes is that fringing caused by night sky emission lines will not appear in dome or twilight flats. In this case, sky frames can help (but they are time consuming to obtain). One alternative is the use of neon lamps, with emission lines, for obtaining flat fields which will show fringing effects (see Howell 2012)[5]. Modern CCDs with anti-reflection coating have reduced the problem, but not completely eliminated it.

- Individual pixels in a CCD might show some cosmetic problems, something that was very common in older CCDs. Some pixels might show very low efficiency (bad pixels), others might generate too high dark current (hot pixels). Some electrodes might fail, rendering a part of a column useless (because it can not be readout anymore). A bad pixel mask can be generated out of dark frames, for example, to interpolate over these problems. If you know that the CCD you are using suffers from such problems, it might be important to take a series of dithered images of your science source, to avoid that the same part of the source is always falling in a bad region of the CCD.

- Cosmic rays that hit the CCD can create large number of electrons, because of the high energy they transfer to the material. These can be identified as bright peaks in your image/spectrum. The longer the exposure, the higher the chances and the number of cosmic ray hits one might get. It is usually recommended to take at least three science images of your target, so that such hits can be identified and averaged out from the final science frame. However, nowadays there are algorithms to identify cosmic ray hits and remove them by interpolation that work extremely well, even if you only have one frame for your science.

---

[5] https://ui.adsabs.harvard.edu/abs/2012PASP..124..263H/abstract

4 Additional reading

If spectroscopy is your technique of interest, then I recommend that you also read Part 3 of the text on “Astronomical Spectroscopy” by (Massey & Hanson 2013)[6].

Several sections of the book “Electronic Imaging in Astronomy” (McLean 2008)[7] might be interesting, if you want to know more about CCDs and their use in astronomy.

References

Bely, P. 2003, The design and construction of large optical telescopes (Springer)

Conrad, A. R. 2014, Software Systems for Astronomy (Springer)

Howell, S. B. 2012, PASP, 124, 263

Massey, P. & Hanson, M. M. 2013, Planets, Stars and Stellar Systems, 2, 35

McLean, I. S. 2008, Electronic Imaging in Astronomy

Merline, W. J. & Howell, S. B. 1995, Experimental Astronomy, 6, 163

---

[6] `https://link.springer.com/referenceworkentry/10.1007%2F978-94-007-5618-2_2`

[7] `https://link.springer.com/book/10.1007%2F978-3-540-76583-7`