

Notes

Marco Codato

April 28, 2022

1 Introduction

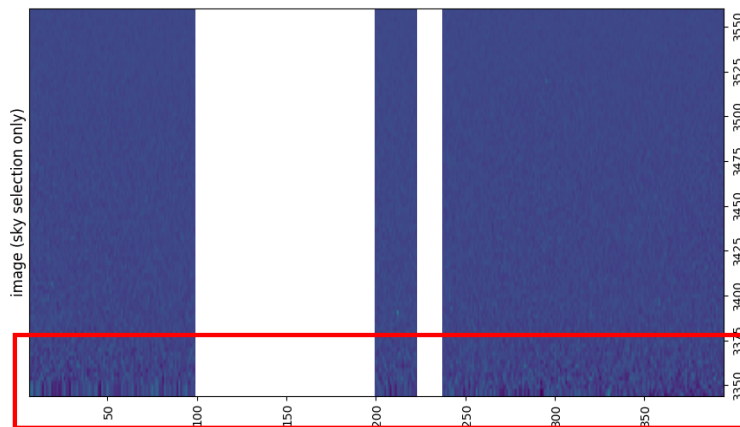
Moved to the GitHub repository `codatomrc/tesi`.

2 Background spectrum extraction

2.1 Automatic detection

The script I am writing works as follows.

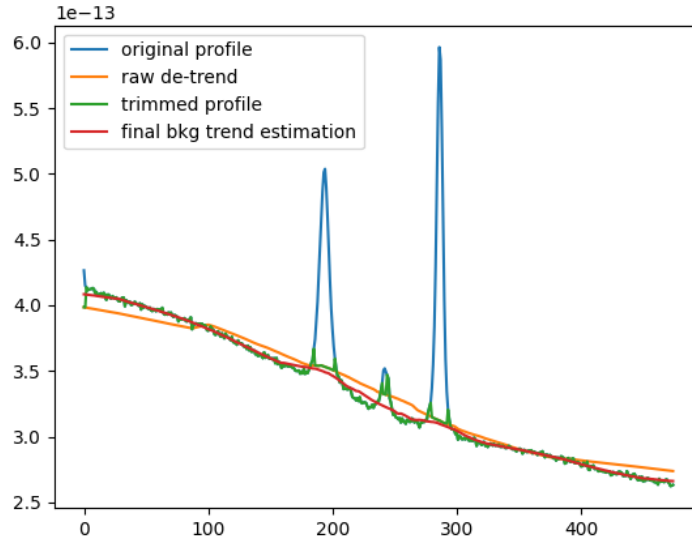
- Automatically explores a directory and find all the raw files, marked as `filename.fc.fits`. All the files are then processed one by one.
- For each frame relevant information is extracted from the header (hdr). In particular the script requires
 - information about the wavelength of each pixel (initial wavelength, increment of each pixel and width of the detector and horizontal binning factor),
 - slit width and length (thus the detector height and vertical binning factor, aside the CCD scale),
 - year of observation
- For each frame a preliminary integration over all the wavelengths is performed by summing the data in the file along the dispersion direction. This is an efficient way of detecting astronomical sources, that are spatially limited, against the background, supposed uniform along the slit.
- I preliminarily decided to cut blue wavelengths at 3500 Å. This value was bull-eyed looking at bkg spectra and realizing that this is the threshold where the noise becomes dominant.
- Cosmic rays and excessive noise are removed. In particular in the bluer part of the spectrum, due to the low sensitivity of the detector, after the flux calibration the noise is extremely high, seriously affecting the quality of the measurements. For example see the image below I removed these features by scanning



each column of the frame looking for bright sharp peaks. Some good settings to remove both cosmic rays and noise is to find peaks very thin, with a width less than 3 px and an height of more than 5 times the average level of the bkg (i.e. the value estimated from the luminosity profile, divided by the number of pixels along the horizontal direction) bisognerebbe stimare la dimensione apparente dei raggi cosmici e confrontarla con la scala sul CCD per essere sicuri che funzioni con tutte le immagini.

The peaks were identified with the function `scipy.signal.find_peaks` while their FWHM was estimated with `scipy.signal.peak_widths`. To be sure all the cosmic ray trace or the noise fluctuation was contained in the detected width I added a further 1 px on both directions along the columns.

- The columns that after the cleaning are depleted of more than the 25% their pixels are discarded as they are very likely to be noisy columns.
- On the cleaned frame I take again luminosity profile along the slit (i.e. integration on the wavelengths) to search for the astronomical sources. Some of the frames present a systematic trend in the bkg which makes the source detection not trivial. I decided to proceed as follows:
 1. Estimate the general trend of the luminosity profile. I used a biweighted detrending algorithm with a windowing of 200 px and the fine-tuning parameter `cval=10`, using the function `wotam.flatten`. Such large window size is still not enough to cancel out the effect of the brightest features.
 2. Trim the original light profile by removing all the data that emerges from the bkg global trend. I used as threshold the detrended profile, vertically shifted by a factor equal to the 5% of the average bkg level. A good estimator for the avg bkg level turned to be the median of the original light profile.
 3. Perform a new detrend on the trimmed data. In this case I used a much narrower windows size, or 50 px since it was not necessary to dilute the bright peaks in correspondence of the sources. This can be considered as an estimator for the global bkg trend along the slit.



- The source peaks are detected automatically with `scipy.signal.find_peaks`. Peaks were identified after removing the contribution of the systematic bkg trend detected in the previous step. I selected only those features with a width larger than 3 px (to be sure not to include unremoved cosmic rays) and an height greater than the 5% of the bkg (median) level.
- I computed the width of the sources with `scipy.signal.peak_widths`. In particular I considered the FWHM which proved to be more solid than the total width. Since many kinds of astronomical sources have a central core and bright wings, I decided to remove a that span $2.5 \times \text{FWHM}$ wrt the center of the peak to take into account such wings. In some cases the luminosity profile of the sources is not gaussian and the tails are brighter than the gaussian prevision. In these cases I considered as related to a source also all the pixels around the peak that are brighter than what expected from the detrended profile.
- At the end the script produces:
 - A plot of the luminosity profile along the slit (the one from the original frame, and after the cosmic ray removal) plus the masked source regions.
 - The bkg spectrum from the selected rows only, integrated along the spatial direction (plus the same for the full cleaned frame as a comparison).
 - A new file where cosmic rays, the UV noise and the astronomical sources are removed. The information in the hdr are preserved and the data when the file was produced is added.

A quantitative approach. I tried to sketch a qualitative model where I assumed gaussian profiles of the sources. The source ended when its flux was comparable to the amplitude of the bkg noise. The final distance Δ from the center of the source was

$$\Delta = \frac{1}{2\sqrt{2\ln 2}} \text{FWHM} \sqrt{\ln(S_{\text{max}}^2/B)}$$

where S_{max} was the maximum counts of the source, while B the average level of bkg around the object. Since many extended objects cannot be reproduced by gaussian profiles this estimation may be misleading in many cases. Since the empirical procedure discussed above seems quite solid with current data, implementing a rigorous analytical approach that accounts for different luminosity profiles of the sources, is definitively unnecessary.

2.2 Comments on the first results

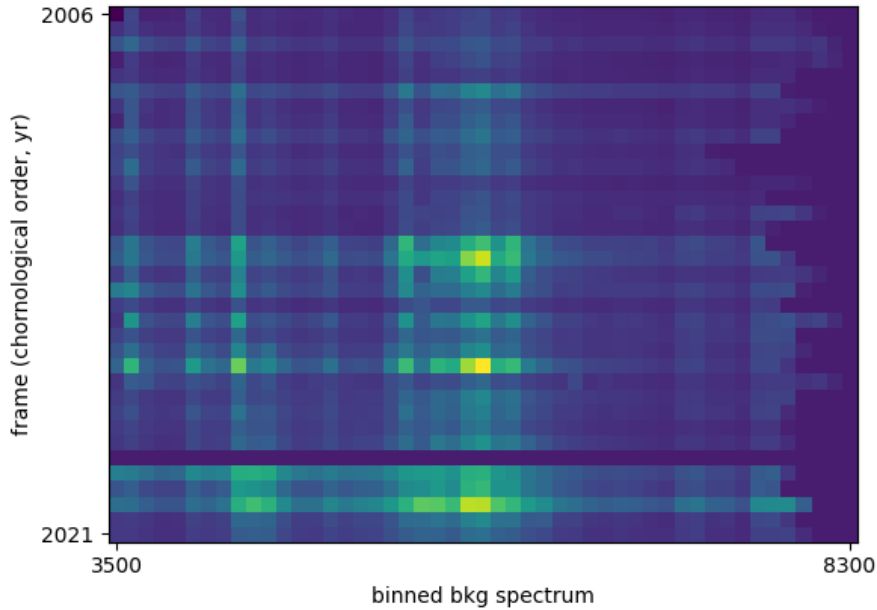
Frame limitations(?) I wonder whether the whole CCD area is suitable for collecting data or it would be better/necessary to neglect some specific regions, both in the spatial and dispersion directions.

What about the lines “CCDSEC” and “BIASSEC” in the header of the frames?

3 Bkg analysis

Once the bkg is extraced from each frame and saved into a new one, I can developpe a script that reads the new filtered files and process them automatically.

Bkg spectra are chacacterized by few very bright emission lines plus one/two continuum region(s). A very first analisys may consist on lower the resolution of the spectra and compare them with each others. In particular I decided to bin them in 100 Å bin size and crhonologically order them.



4 Appendix: Python codes

4.1 Bkg extraction code

```
import numpy as np
import matplotlib.pyplot as plt
from astropy.io import fits
from scipy.signal import find_peaks, peak_widths
from datetime import datetime
import glob
import os
from wotan import flatten
from scipy.optimize import curve_fit

#####
#OPTIONS
save_plots = False
save_FITS = False
plot_profile = True
plot_spec = False
show_ima = False

#PARAMS
peak_height = 0.05 #height above the bkg level
data_col_frac = .75 #minimum fraction of valid pixels in a column
width_mult = 2 # interval to exclude around a source, wrt center in FWHM units
cr_width = 2.5 # cr trace spatial width
cr_prominence = 5 #threshold height wrt average column level
cr_pad = 1 # number of px to exclude around cr, in a fixed column
LAMBDA_lim = 3500 #A, limit blue wavelength
#####

#browse all the *.fc.fits files in a directory and its subdirectories
main_path = './Asiago_nightsky/2020/'
#main_path = './'
file_ls = glob.glob(main_path+'**/*.fc.fits', recursive= True)
names = [os.path.basename(x) for x in file_ls]

#initialize the .log file
f = open("bkg_extr.log", "w")
f.write('Running bkg_extr.py at '+
        datetime.now().strftime("%H:%M:%S, %Y-%m-%d")+'\n')
f.close()

#append to the log
warnings_count = 0
def log_append(message):
    f = open("bkg_extr.log", "a")
    f.write(message+'\n')
    f.close()

#####
#####

#process all the files found
for name,file in zip(names,file_ls):

    print('processing file '+name+'\n', end='\r')

    #open a FITS file
    hdul = fits.open(file)
    hdr = hdul[0].header

    #extract wavelength information from the header
    NAXIS1, NAXIS2 = hdr['NAXIS1'], hdr['NAXIS2']
    LAMBDA0, DELTA = hdr['CRVAL1'], hdr['CDELTA1']

    #generate the lambdas array
    if hdr['CTYPE1'] != 'LINEAR':
        log_append('WARNING: no linear wavelength calibration')
    LAMBDA = np.arange(LAMBDA0, LAMBDA0+NAXIS1*DELTA, DELTA)
    if len(LAMBDA) == NAXIS1+1:
        LAMBDA = LAMBDA[:-1]

    #remove extreme blue wavelengths
    LAMBDA_start_id = len(LAMBDA)-len(LAMBDA[ LAMBDA>LAMBDA_lim])
```

```

LAMBDA = LAMBDA[LAMBDA_start_id:]

year = hdr['DATE-OBS'][:4]

#aperture information from the hdr
SLIT = hdr['SLIT'] #microns
try:
    BINX, BINY = hdr['BINX'], hdr['BINY']
    TELSCALE = hdr['TELSALE'] #arcsec/mm
    CCDSCALE = hdr['CCDSALE'] #arcsec/px
except KeyError:
    BINX, BINY = hdr['HBIN'], hdr['VBIN']

    log_append(' WARNING: no scale info in the hdr (using defaults)')

    TELSCALE = 10.70 #arcsec/mm #TO BE CHECKED!!!
    CCDSCALE = 0.60 #arcsec/px #TO BE CHECKED!!!

SLIT_angular = SLIT/1000 * TELSCALE #slit size in arcsec
SLIT_px = SLIT_angular / CCDSCALE / BINX #slit size in px

#####
#bkg level estimation
raw_data = hdul[0].data[:,LAMBDA_start_id:]
raw_integr = np.sum(raw_data, axis = 1)
x = np.arange(len(raw_integr))

bkg_est = np.nanmedian(raw_integr)

#####
#remove cosmic rays and UV noise
data = np.copy(raw_data)

cr_col_frac = np.zeros(len(LAMBDA)) #fraction of remaining px
for cr_col,col in enumerate(data.T):
    col_avg = np.nanmean(data[:,cr_col])
    cr_line,_ = find_peaks(col,
                           prominence = cr_prominence*col_avg,
                           width = (0,cr_width))

    cr_widths = peak_widths(col, cr_line, rel_height=0.5)[0]

#set left and right boundaries of the source region along the slit
left_width = cr_line-cr_widths - cr_pad
right_width = cr_line+cr_widths + cr_pad

#scan each column and remove peaks
cr_sel = np.zeros(np.shape(col), dtype=bool)
for i in range(np.shape(col)[0]):
    for peak,width in zip(cr_line,cr_widths):
        if abs(i-peak) < width+cr_pad:
            cr_sel[i] = True

#counts how many pixels are left in a column
saved_px = (NAXIS2 - np.sum(cr_sel))/NAXIS2
cr_col_frac[cr_col] = saved_px
if saved_px >= data_col_frac: #if enough, take the masked column
    data[cr_sel, cr_col] = np.nan
else: #else discard the entire column
    data[:, cr_col] = 0.

#####
#use noise/bkg info to find peaks
integr = np.nansum(data, axis = 1)
bkg_est = np.median(integr)

#detrend: global trend (including peaks)
_,trend_raw = flatten (x,
                       integr ,
                       method = 'biweight',
                       window_length =200 ,
                       cval = 10, return_trend = True )

#trim removing peaks, i.e. data fare above the global trend
integr_trim = np.where(integr <= trend_raw+0.05*bkg_est,
                       integr, trend_raw)

```

```

#detrend the trimmed data , much less sensitive to the peaks
_,trend = flatten (x,
                    integr_trim ,
                    method ='biweight',
                    window_length =50 ,
                    cval = 10, return_trend = True )
'''
more plot about detrending
plt.plot(x,integr, label='original profile')
plt.plot(x,trend_raw, label='raw de-trend')
plt.plot(x,integr_trim, label='trimmed profile')
plt.plot(x,trend, label='final bkg trend estimation')
plt.legend()
plt.show()
'''

#detrend residuals: original peaks are highlighted wrt the bkg profile
diff = integr-trend

#find peaks
peaks,properties = find_peaks(diff, height=0.05*bkg_est, width = cr_width)
peak_FWHM = peak_widths(integr, peaks, rel_height=.5)[0]/2.

if len(peak_FWHM)== 0:
    no_source = " WARNING: no sources were detected"
    log_append(no_source)

#generate a booleana mask True outside the peaks
bkg_sel = np.full(np.shape(x), True)
for i,peak in enumerate(peaks):
    width = (int(peak_FWHM[i])+1)*width_mult
    for w in range(-width,width):
        bkg_sel[peak+w]=False
    w = width -1
    while integr[peak+w] >= trend[peak+w]:
        bkg_sel[peak+w] = False
        w += 1
    w = width
    while integr[peak-w] >=trend[peak-w]:
        bkg_sel[peak-w]=False
        w += 1

#####

#plot the luminosity profile , show source and bkg regions
if 1 == plot_profile:
    plt.title(year+'/' +name[: -8]+' : wavelenght integration')
    plt.plot(raw_integr, alpha=0.2, ls='dashed', c='C1')
    plt.plot(integr, alpha=0.4) #integrated flux
    plt.scatter(x[bkg_sel],
                integr[bkg_sel],
                s=0.2, c='green') #select bkg

#esimate the bkg of the filtered regions only
bkg_est_filt = np.mean(integr[bkg_sel])

plt.plot(x, trend_raw+0.05*bkg_est, ls='dashed', c='grey', alpha=0.5)
ima = np.zeros(np.shape(bkg_sel))
plt.fill_between(x, ~bkg_sel*1.1*max(integr), color='red', alpha=.1)

#plt settings
plt.ylim(min(integr)*0.9, max(integr)*1.1)
plt.legend(['raw signal','cleaned signal',
            'bkg signal only', 'detrended treshold',
            f'peak regions ({width_mult}*FWHM)'])

if save_plots is True:
    plt.savefig('./plots/integr/'+year+'_'+name[: -8]+'.png')
    plt.close()
else:
    plt.show()

#####
#integrated spectrum (along the slit)
total = np.nanmean(data, axis = 0) #integration along the slit

```

```

sky = np.nanmean(data[bkg_sel,:], axis = 0) #integration of bkg rows only

total[total == 0] = np.nan

#plot the spatially integrated spectrum of the bkg
if 1 == plot_spec:
    plt.title(year+'/' + name[: -8] + ': bkg spectrum')
    plt.plot(LAMBDA, total, label='full frame', color='gray', alpha=0.3)
    plt.plot(LAMBDA, sky, label='sky only')
    plt.legend()
    if save_plots is True:
        plt.savefig('./plots/sky_spec/' + year + '_' + name[: -8] + '.png')
        plt.close()
    else:
        plt.show()

#####
#extract only the bkg rows

ma_data = data #set masked data
for i,row in enumerate(bkg_sel):
    #cancel data from the source rows
    if row == 0:
        ma_data[i,:] = np.nan

#plot as image the bkr rows only
if (1 == show_ima) and (save_plots == 0):
    plt.title('image (sky selection only)')
    plt.imshow(ma_data, extent = [LAMBDA[0], LAMBDA[-1], NAXIS2, 0])
    plt.show()

#####
#save masked data in a new FITS file
if 1 == save_FITS:
    now = datetime.now()
    now_str = now.strftime("%Y-%m-%d %H:%M:%S")

    hdr.set('BKGEXTR', now_str, 'Time of bkg extraction')
    hdr.set('UVLIM', LAMBDA_lim, 'A')
    hdr['NAXIS1'] = len(data[0])
    new_hdu = fits.PrimaryHDU(ma_data)
    new_hdul = fits.HDUList([new_hdu])
    new_hdul[0].header = hdr

    file_new = file[: -5] + '.bkg.fits'
    new_hdul.writeto(file_new, overwrite=True)

f = open("bkg_extr.log", 'r')
warnings_count = len(f.readlines()) - 1
if warnings_count != 0:
    print(f'WARNING: {warnings_count} warnings occurred (see the log)')

```

4.2 Bkg analysis code

```

import numpy as np
import numpy.ma as ma
import matplotlib.pyplot as plt
from astropy.io import fits
import astropy.coordinates as coord
from astropy import units as u
from scipy.signal import find_peaks, peak_widths, detrend
from scipy import ndimage
from datetime import datetime
import glob
import os
from sklearn.decomposition import PCA
from sklearn.preprocessing import StandardScaler

#####
#OPTIONS
savefiles = False
plot_profile = False
plot_spec = False
show_ima = False

#PARAMS

```

```
#####

#browse all the *.fc.fits files in a directory and its subdirectories
main_path = './Asiago_nightsky/2011/'
main_path = './'
file_ls = glob.glob(main_path+'/**/*.fc.bkg.fits', recursive= True)
names = [os.path.basename(x) for x in file_ls]

LAMBDA_bins = np.arange(3500, 8500, 100)

df = np.zeros((len(names),len(LAMBDA_bins[:-1])))

#process all the files found
file_id = 0
for name,file in zip(names,file_ls):

    #load the frame
    hdul = fits.open(file)
    hdr, data = hdul[0].header, hdul[0].data

    #take wavelength info from the hdr
    NAXIS1, NAXIS2 = hdr['NAXIS1'], hdr['NAXIS2']
    LAMBDA0, DELTA = hdr['CRVAL1'], hdr['CDELTA1']
    LAMBDA_lim = hdr['UVLIM']

    #the (eventually) UV-limited wavelengths array
    LAMBDA_start = max(LAMBDA_lim, LAMBDA0)
    LAMBDA = np.arange(LAMBDA_start, LAMBDA_start+NAXIS1*DELTA, DELTA)
    if len(LAMBDA) == NAXIS1+1:
        LAMBDA = LAMBDA[:-1]

    spec = np.nanmean(data, axis=0)
    #bin_indices = np.digitize(LAMBDA, LAMBDA_bins)
    #print(bin_indices)

    hist,_ = np.histogram(LAMBDA, bins=LAMBDA_bins, weights=spec)
    df[file_id] = hist
    file_id +=1
df[29,:]=0
plt.imshow(df)
plt.xticks([0,len(LAMBDA_bins)-2],[LAMBDA_bins[0],LAMBDA_bins[-2]])
plt.yticks([0, file_id-1],[2006,2021])
plt.xlabel('binned bkg spectrum')
plt.ylabel('frame (chronological order, yr)')
plt.show()
```