Peer Review.
October 29, 2020.

Overview

The authors of this work analyzed a C7.1 class flare nicknamed the “Thin Mint Flare” that occurred at roughly 2:10 am UTC on March 12, 2015. The authors removed the preflare noise (did baseline correction), adjusted the curve to remove energy from a second solar flare, and integrated to find the total energy of the flare.

In order to remove the pre-flare noise, the authors of this work started by finding the local minima of points on either side of the peak. Assuming that these local minima were roughly the highest levels of noise, they averaged the local minima and subtracted the average of the local minima from the original data, using min.locs, mean(), and .subtract from the numpy library, scipy, and statistics libraries. This data was the baseline corrected data. After removal of the preflare, the authors had to remove the energy due to the 2nd solar flare. They started by estimating the start and end of the 2nd overlapping solar flare. Next, they approximated the exponential decay of the original flare using a linear line, converting the solar flare dataframe to a list, manipulating the lists to somehow change the original data to the line data over the period of the 2nd solar flare (this part of the method was somewhat unclear), and converting the data in list form back to a dataframe. They plotted the adjusted curve, zoomed in on the start and end of the flare, and estimated starting and ending times for the flare. Finally, they used trapezoidal integration from the scipy library to calculate area under the curve (and thus total energy) from the flare’s start to end and reported this total area as total energy in ergs.

Merits

Flare Description

- Subclass of the flare is reported correctly, as is time and peak irradiance.
- I love the name of the flare; so creative!
- I like that you include the image of the flare. It would be really cool if you could circle the flare on the sun image!

Coding Approach (Task 4, Baseline Correction)

- I like that you grouped your approach into sections and used bullet points. These do a really great job of separating out your thoughts.
- I also really like inclusion of some of the function names – it gives us a preview of what’s to come.

Baseline Correction Code

- I love how concise and clear your code is. It is well organized and well commented, especially line by line.

- It makes sense what you are doing. Easy to read code, for the most part.
- The graph looks really good; I like how you put all of the graphs on the same curve, used a legend, and had different colors
- Units are correct, in seconds.
- Sufficient information was used to find the baseline flare.

Coding Approach (Task 4, Reporting Total Energy)

- Method makes sense. I love the way that you've outlined the approach to taking out the 2nd flare. Makes way more sense than the way my team did it. Perhaps also move the bullet points out so that they are at the same level – I don't really think you need to indent so many times.
- Love the word "interpolate." Lol.

Integration & Total Energy Code

- I like that most of the lines of code are commented to the side! This helps me understand what's going on.
- The code is clearly segmented into paragraphs (?), and that helps a lot!
- "Comparison…" Graph is clear, colors, legends, and labels used well.

- I like that you zoomed into the plots for the starting and ending points – we as readers can clearly see the starting point compared to the actual curve, and it is a good check that the points you chose are good points to start and end your integration with.
- Multiple graphs add clarity and help to break up code for what would otherwise be one very lengthy section.
- All components of graphs are present, and many colors and labels are used to improve readability of graphs.
- Correct units are used – time in seconds.

- I like the way that there is a color under the curve to show the integration. The colors/points used to show the starting and ending points are also helpful.
- Seems to accurately show the total energy of the flare
- Again, code is concise and (mostly) clear. Units are correct – total energy of flare is reported in ergs, integration is done with respect to time and not the default index units
- Limits of integration appear to be reasonable.
- Trapezoidal integration was used.
- Method of integration was pretty reasonable.

Critiques

Flare description:

- I think that this flare actually starts a few minutes earlier, closer to 2:10 am, based off of the Space Data Portal dataset. The flare peak is at 2:18 am, but it starts slightly earlier.
- In your description of your flare, you talk about some of the characteristics of the flare but not at all about the environment it is in – what did the other flares that occurred before and after it look like? Were there a lot of flares in the time period? How powerful were the flares that occurred somewhat near to it, timewise?
    - For example “Thin Mint was one of several C class flares that occurred on March 12th, in a period of moderate activity. In fact, just one day prior, the sun had ‘its first superpowerful flare of the year,’ ‘aimed directly at Earth.’ The flare on March 11, peaking at 12:22, was a monster flare that led to many blackouts, and our flare, Thin Mint, followed its intense activity, along with many M class flares.”
    - It could also be useful to give the time of the peak irradiance.
- The flare description mentions that the flare chosen looked smooth and straight forward to calculate total energy… what about the baseline? Why was this specific flare chosen as opposed to another smooth flare?
- The sentence on features is slightly vague (“the features of this flare that stood out was the nice spike and curve…”); I wonder if you could try using more specific and academic language like “Although the flare’s clean spike and downward sloping curve initially seemed easy to analyze, while we were working, we realized that there was another flare that overlapped with our flare. We then had to find different data in order to cut out the smaller flare.”
    - Also, where did this “different data” come from?
    - You could also describe this smaller flare with more specific data – eg: “at time x… while Thin Mint was on the downward slope, another flare, with a peak flux of y… peaked. This smaller flare ended at time y and was relatively small/large so we found that it was pretty easy/hard to remove from Thin Mint’s data.”
- In your description of the flare, you mention the peak irradiance… how was this value calculated or identified?

Coding Approach (Task 4, Baseline Correction)

- When you refer to “local minima,” I am not sure if you mean the local minima right before the start of the flare, over the entire flare, after the flare… which minima are you finding? And over what time range? Precision here will help clarify to the reader.
- Instead of “subtract that off the dataset,” saying “subtract the average power from the original data” will be more precise, perhaps less colloquial.
- Under reasoning, when you say “Rather than the average of our dataset given over a period of time…”, do you mean to say “over the entire range of time before the start of the flare?” or something similar? What period of time are you referring to?
- What do you mean by “when a flare may become an outlier?” Do you mean to say that averaging the energy over the entire baseline would end up including flares in

the baseline, when it should really just be the background energy or noise? I'm just not sure that the word outlier is used correctly here.
- The second bullet point could be written a little more clearly… I don't think that you need the clause "Since there exists some noise in our dataset,". Instead, you could probably just write that "Local minimas represent the highest noise level read from the GOES-15 XRS. Subtracting the highest noise level will ensure that all of the noise is removed, while ensuring that no area actually due to the flare itself is removed."
- Why an average as opposed to a median? I don't know which is right, but a brief one sentence, 10 words or less explanation could also be helpful.

Baseline Correction Code

- I wonder if the local minima at about 5500 seconds should be removed from the baseline correction. This data point still seems to be along the flare – it doesn't seem to have ended quite yet, and also, I think it more attests to the fact that a second flare overlaps with your flare; the curve goes up because of that second flare, not because it is the highest noise level reading. Therefore it shouldn't be included in the calculation for the baseline.
- Also, the noise level before the start of the curve is on a slight downward slope… considering the context of the flare, that it occurred a day after a huge flare, I wonder if there is also some overlap with that flare? Maybe that's a stretch, but it's definitely not flat before the start of your flare, so perhaps there is some error there (baseline curve is somewhat biased, perhaps? Although a lot of points/data are used).
- Overall, I think your baseline is too high. Between the downward slope, the inclusion of the minima at 5500 seconds, and the inclusion of local minima on the downward slope of the curve (second half of the flare) make me uncertain that the baseline is low enough.
- This is just a little nitpicky, but in your legend and title, be sure to capitalize consistently, or not at all. Also, legend is spelled incorrectly in the comments at the bottom of your code. (plt.legend())
- I wonder if a few more comments could be added to effectively section out your code. You've already written the steps in the description, so why not add them into the code? For example: # Average the minima and print average value; this is the value we will need to subtract from the data #Now, we remove the baseline from the data, and define a new array with only the baseline corrected data #Graph Data
- Also, your graph's baseline goes from 0 to 9000, but the rest of your curves stop before 8000 time units.. maybe change the xaxis for the baseline back to 9000 and then move the legend to the top left side of your graph, in order to just give the graphs that much more room?

Coding Approach (Task 4, Reporting Total Energy)

- May I suggest that as an alternative to using a linear function for the decay, you could use the polynomial fit functions from the pandas library? These use dataframes, and we did them on Problem 12 of the Coding Packet from like Week 3

or 4. Although the line does seem to approximate the curve well, at least visually, and I also think it is a pretty reasonable assumption.
    - It is unclear – how exactly did you decide on 5580s and 6960s for the start and end of the line, again? Thank you!
- "After we had removed the 2nd flare. We …" is not a full sentence. Oops! 😊 I think you meant to have a comma. You have one more grammatical errors in the bullet points below as well "it is when it pass the…" Some more commas or other grammatical pauses may also be helpful in improving readability.
- When you write "the ending point should be near zero" in the final bullet point under integration, I am confused by what zero is referring to. Zero seconds? Zero ergs/second?
- Under integration, the first bullet point should say somewhere that the first step you took was to identify the starting and ending points. It does not say this.
- To be consistent, you should make "Then, to calculate…" a bullet point. I missed this at first glance.

Integration & Total Energy Code

- As a reader unfamiliar with def (or at least, unfamiliar in python), I would be better able to understand what the second paragraph of your code does if you commented.
- Where does the equation go to? I am unsure where you will be moving this equation down the line; commenting where the equation returns to would be helpful!
- I think adding a few more comments before a few of the sections would help readers understand what each part of the code does. You do this well for the first paragraph ("# removing the 2nd flare slope…"). This same thing could easily be copied below to right above "def interpolate…" and "removedFlare_data…" to increase clarity. Maybe also the information purely for plots could just all be under one label, no "comment header" necessary.
- I notice that there aren't any axis labels for the "comparison of baseline…" plot You should consider adding these, as well as capitalizing the words in the title, to improve clarity. Maybe the legend titles can also be made shorter? I notice that you moved it. But how to shorten them? TBA. :D The blue line legend label should also be remained to make more sense. Like "2nd Flare Pre-removal"
- I also wonder if including an x-axis (plot at y = 0) could make your graph look cleaner, so people can see that the data is, in fact, the post-baseline correction data.

- Some comments need to be added to the code on the Starting Point & Ending Point sections to explain what you are doing. Adding one header to the top of each saying "in order to integrate, we need to find the starting points and ending points of the flare." And/or "Graphing the start of the flare, we can visually identify a good starting point for the integration of Thin Mint." Or "we chose 3840s as the starting point for integration because _____."
- Moving the legend to the upper left side and shortening the labels would make the starting point graph more readable.

- Did you just estimate the starting point for the integration? Or did you find a local minimum? Either way, it would be helpful to explain why you chose the points that you did, as well as why you *didn't* use min.locs or vice versa.

- Again, more header comments would be helpful in the integrate section to explain what you're about to do. What does line 3 "flare_data = removed…" do?
- I think that the label for the red star on the right side of the curve is meant to say "Ending point," not "Starting point."
- Instead of "sets the x limit" as the comment, you could say "defines the graph domain"
- How did you decide, again, that you didn't need an end point for the integration? It is not clearly explained. I also wonder what you think on whether the slight increase/bump at 7000 seconds is another overlapping flare or not. If not, why not? If it is another flare, why did you choose not to remove this flare, but to remove the other flare? If it isn't, why is it noise?
- An x-axis or other plot would, again, be helpful to show that the data is baseline corrected (even though we already know it is, it confirms what we know).
- How would you go about figuring out/calculating/attributing the uncertainty? I know you didn't have time – the work you've done is VERY comprehensive, well thought out, and well done, and I'm sure some sort of explanation here would only improve your work and make it that much better.

Overall Recommendation

	No revisions are needed
X	Needs minor revisions
	Needs major revisions

Conclusions

Overall, the authors of this work did a good job of eliminating noise and secondary flares from the data and integrating to find total energy. Units are consistent, graphs are used effectively, and code is generally concise, clear, and advanced. The methodology is mostly well thought out; however, a few minor adjustments will help further reduce error and improve clarity/readability. Most importantly, the local minima for baseline correction could be better chosen; the authors should consider better explanation of starting and ending values for their flares; and more explanation of the methodology is necessary through clear and precise use of markdown cells and header comments.