

TECHNICAL UNIVERSITY OF DENMARK

DTU SPACE

# Identification and Analysis of sprites in ASIM-data

*Rasmus Christian Jørgensen (s164044)*

Supervisors: Torsten Neubert & Olivier Chanrion



December 12, 2020

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Data Processing</b>	<b>2</b>
2.1	Pre-processing: Noise Removal . . . . .	2
2.2	Method 1 - ASIM: 777.4nm removal from 337nm . . . . .	3
2.3	Method 2 - ASIM: Delayed UV peak detection . . . . .	6
2.4	Method 3 - ISUAL . . . . .	9
<b>3</b>	<b>Results</b>	<b>11</b>
<b>4</b>	<b>Discussion</b>	<b>16</b>
<b>5</b>	<b>Conclusion</b>	<b>24</b>
	<b>References</b>	<b>25</b>
	<b>Appendices</b>	<b>26</b>
<b>A</b>	<b>User-Guide: Running Method 3 and reproducing figures in Results section</b>	<b>26</b>

## 1 Introduction

The Modular Multispectral Imaging Array (MMIA) is one of two instruments forming the Atmosphere-Space Interaction Monitor (ASIM) mounted on the Internation Space Station (ISS). ASIMs main goal is to gather information on lightning, Transient Luminous Events (TLEs), and Terrestrial Gamma-ray Flashes (TGFs) [Chanrion, 2019]. The ASIM instrument has been gathering data since its installation in April 2018, which results in very large quantities of data. In order to get a proper understanding of TLEs, a systematic approach to detecting them in the dataset is necessary. Therefore, the focus of this report has been on developing an algorithm for the detection of Sprites in the MMIA dataset.

Three different approaches have been tested in an attempt to detect Sprites, and one method has been determined as the most promising, and evaluated on the 2019 dataset from MMIA. Two of the methods are based on an event that is assumed to be a sprite found in the ASIM dataset, and the third is an adaptation of the method presented by [Adachi *et al.*, 2016], which is based on observations from The Imager for Sprites and Upper Atmosphereic Lightning (ISUAL).

It is important to keep in mind, that the focus of this report is on building, testing, and validating the algorithms in question. The report will not cover the physical processes behind TLEs, nor the physical implications of observations found in the dataset. Rather, a classification and justification of the classes is presented in the Discussion section of some of the most interesting events found, which would enable possible later studies into the events. Additionally, some figures are presented which are more suited to be viewed in a running python environment, than to be observed in a report as here. Therefore, a section of the Appendix is dedicated as a user-guide for running and reproducing the figures presented here.

## 2 Data Processing

In an attempt to identify sprites in the ASIM-dataset, three different kind of approaches has been tested out. The two first are based on observations of an event found in the ASIM dataset, and the other approach is an adaptation of the method presented by [Adachi *et al.*, 2016]. All three methods will be described in depth the following section, and a brief evaluation on whether they have been pursued further is included.

### 2.1 Pre-processing: Noise Removal

In an attempt to remove some of the noise present in the ASIM dataset, a Butterworth Filter has been applied, working as a low-pass filter. A first order filter has been applied with the coefficients:

$$f_s = 10 \cdot 10^6, \quad f_c = 5 \cdot 10^5, \quad w_n = \frac{f_c}{f_s/2}$$

Where  $f_s$  is the sampling frequency,  $f_c$  is the cutoff frequency and  $w_n$  is the normalized cut-off frequency. The proves very efficient at removing fluctuations from sample to sample, revealing what appears to be the underlying true signal. A comparison of the raw data and the effects of applying the filter can be seen on fig. 1 and fig. 2 respectively. This proves especially effective in the UV band, where the signal is very weak compared to the noise level in the signal.

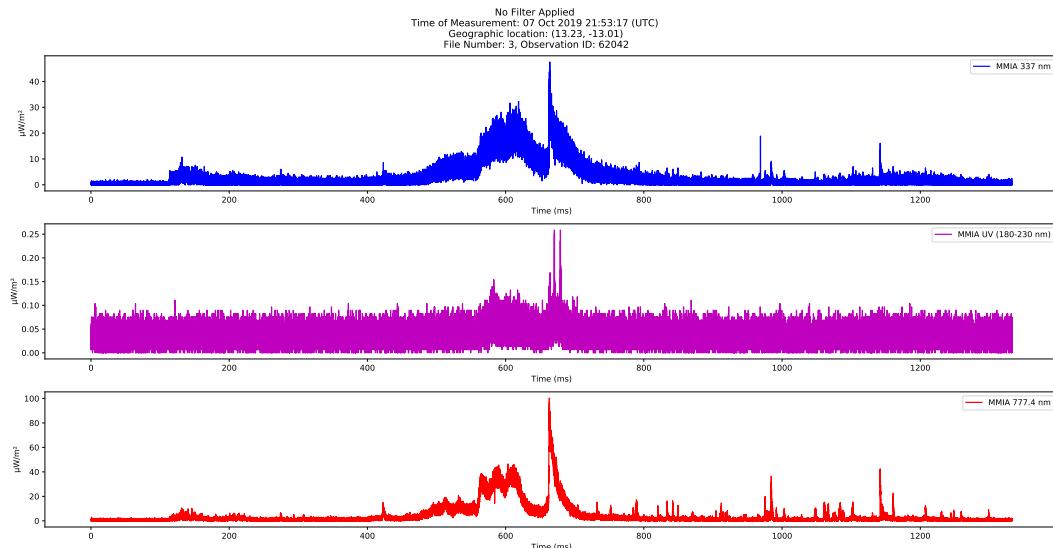


Figure 1: Example of a raw dataset with no filter applied. This is two consecutive data files put together to show the entire signal.

## Data Processing

Rasmus C. Jørgensen Method 1 - ASIM: 777.4nm removal from 337nm

12/12-2020

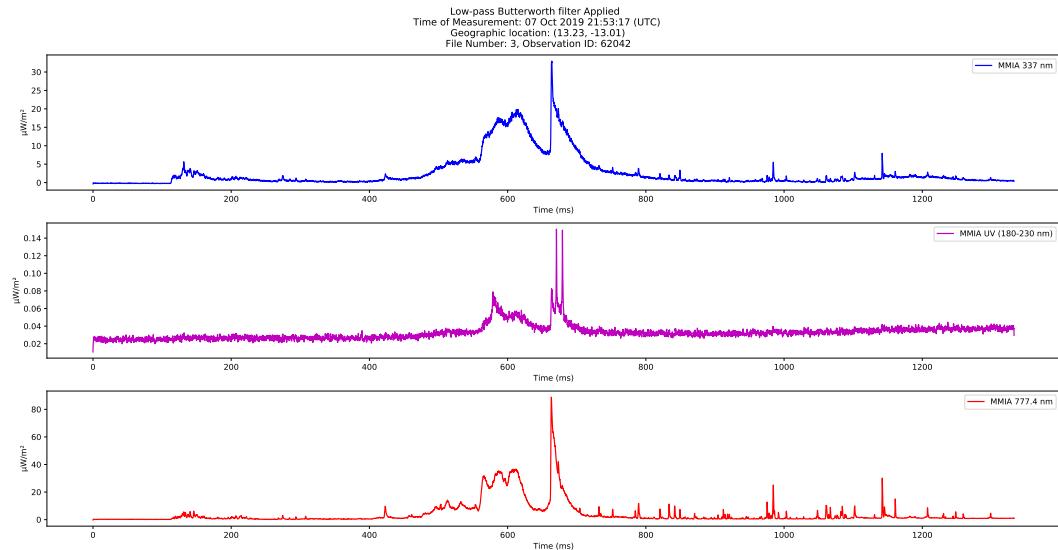


Figure 2: The same event as shown on fig. 1 after applying the Butterworth low-pass filter.

## 2.2 Method 1 - ASIM: 777.4nm removal from 337nm

The first method builds on the assumption that the measured signal in the 337nm band contains contributions from the 777.4nm band as well, and attempts to remove the 777.4nm contribution. The idea is to determine a scale factor between the two signals, scale the 337nm signal to match the 777.4nm signal, and subtract the two. In order to determine the scale factor, the slope of the peak in 337nm and 777.4nm is determined. Figure 3 shows the features we are interested in, which is the signal from the observed peak, until the signal starts to flatten out. On fig. 3 this is shown as only 1 sample, but in practice this can be multiple samples.

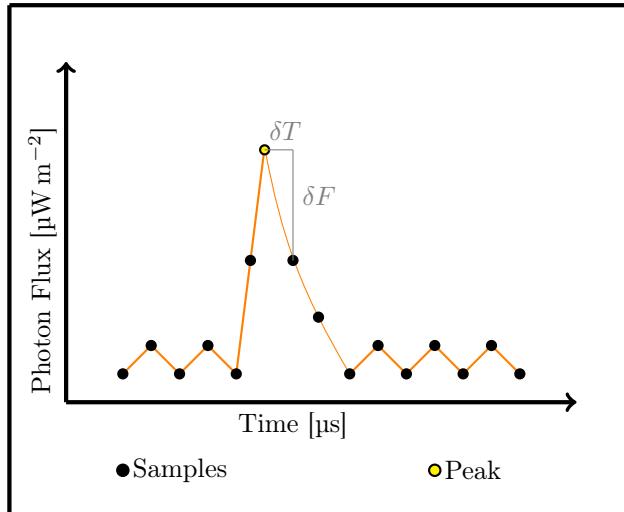


Figure 3: A sketch of the idea behind determining the scale factor for Method 1. Please note that this figure is not supposed to represent any particular signal, but rather the shared features observed in the 337nm band and 777.4nm band. The slope is determined by looking at the difference in time,  $\delta T$ , and difference in the photon flux,  $\delta F$ .

The slope of each signal is then determined as:

$$S_{337} = \frac{\delta F_{337}}{\delta T_{337}}, \quad S_{777.4} = \frac{\delta F_{777.4}}{\delta T_{777.4}} \quad (1)$$

Where  $S$  is the slope,  $\delta F$  is the change in photon flux and  $\delta T$  is the change in time respectively, as shown on fig. 3, and the subscript indicates the signal in question.

The scale factor can then be determined as the ratio between the slopes:

$$R_{scale} = \frac{S_{337}}{S_{777.4}} \quad (2)$$

Finally, the 337nm signal can be scaled:

$$F_{337,scale} = \frac{F_{337}}{R_{scale}} \quad (3)$$

Since the 337nm signal should now be scaled accordingly to the 777.4nm signal, an attempt to remove the 777.4nm contribution can be made as:

$$F = F_{337,scale} - F_{777.4} \quad (4)$$

This method has been applied to the event shown on fig. 2, and the results are shown on fig. 4. Looking at the scaled 337nm signal, it appears to work as intended; the signal matches the 777.4nm signal as expected. However, looking at the difference between the two signals, no real structure is revealed. The 777.4nm signal still appears to dominate, which gives a negative spike of roughly  $20 \mu\text{W m}^{-2}$ . A slight positive signal is shown following the negative peak, which could come from the 337nm signal. As an additional attempt, the scaled 337nm signal is plotted against the 777.4nm signal, to see if it reveals any useful structures. Again the high

peak values dominate the plot, shown on the far right of the plot. Small bumps along the line appear to be visible, which could be contributions from the 337nm signal, but these are too weak to give any meaningful information, and their implications with regard to the 337nm signal is mostly speculations.

On a final note, as of now, this method is abandoned in favor of testing the other methods, due to the inconclusive results.

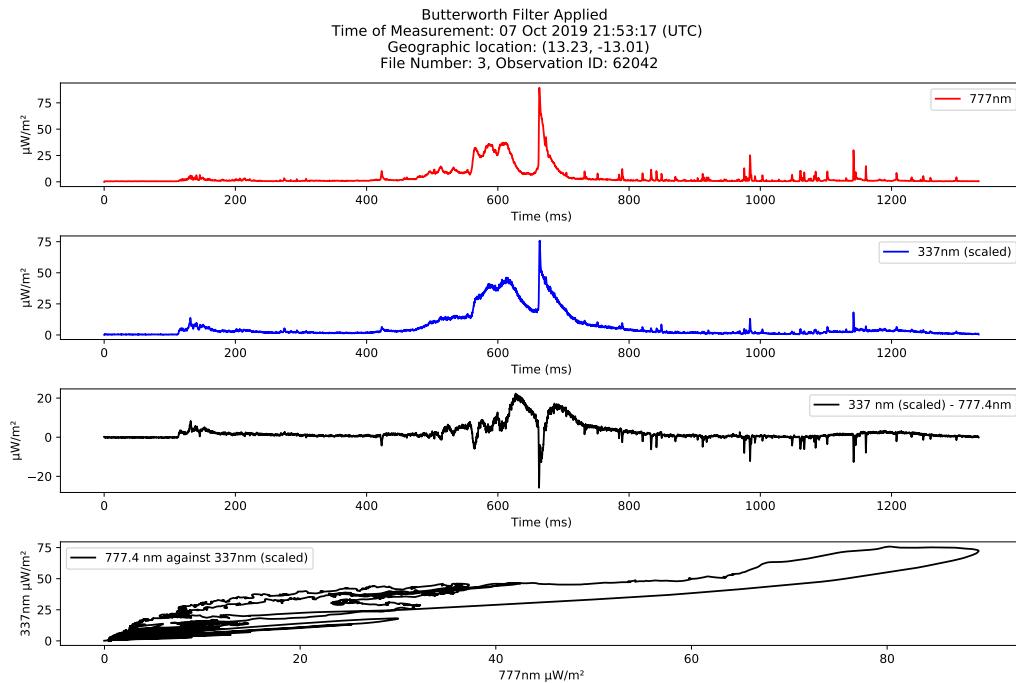


Figure 4: The result from applying Method 1 to the event shown on fig. 2. From the top: (1) the observed signal in 777.4nm, (2) the scaled 337nm signal, (3) the difference between the scaled 337nm signal and the observed 777.4nm signal, (4) the scaled 337nm signal plotted against the 777.4nm signal

### 2.3 Method 2 - ASIM: Delayed UV peak detection

Method 2 builds on the observations from fig. 2, which shows two distinct delayed signals in the UV band, compared to the peaks found in both 337nm and 777.4nm. In an attempt to detect similar events, the idea has been to relate the two signals to each other in a single signal. In order to do this, the signals have been normalized with respect to their maximum value, which results in all three signals being in the [0,1] interval:

$$\bar{F}_{norm} = \frac{\bar{F}}{F_{max}} \quad (5)$$

In this case  $\bar{F}$  is the signal measured in either 337nm, UV, or 777.4nm,  $\bar{F}_{norm}$  is the normalized signal, and  $F_{max}$  is the maximum value in the given signal. Subtracting either 337nm or 777.4nm from the UV band, results in the relation between the UV band and the other signals:

$$\bar{F}_{Rel} = \bar{F}_{norm,UV} - \bar{F}_{norm} \quad (6)$$

Where  $\bar{F}_{Rel}$  is the relation between the signals,  $\bar{F}_{norm,UV}$  is the normalized UV signal, and  $\bar{F}_{norm}$  is either the normalized 337nm signal or 777.4nm signal, depending on which relation is desired.

Since the peaks are delayed with respect to each other, the assumption is, that no peaks will result in a signal around 0, a negative peak will indicate a peak in either 337nm or 777.4nm depending on which  $\bar{F}_{norm}$  is used, and a positive peak will indicate a peak in UV. A peak finding algorithm can then be applied to search for the sequence of negative peaks followed by positive peaks, to indicate the delayed UV peaks. Of course, this could also be used to find the other case of UV peaks that are faster than the corresponding 337nm or 777.4nm peaks, by searching for positive peaks followed by negative peaks.

The method has been applied to the event shown on fig. 2 again, and the applied peak finding algorithm is *Filter 1* presented in [Olesen & Jørgensen, 2020]. This algorithm is not intended to be used on negative values, however as a proof of concept, the filter is instead applied twice, once on the actual signal  $\bar{F}_{Rel}$ , and once again after inverting the signal,  $-\bar{F}_{Rel}$ . The results are shown on fig. 5, and a close up of the peak on fig. 6.

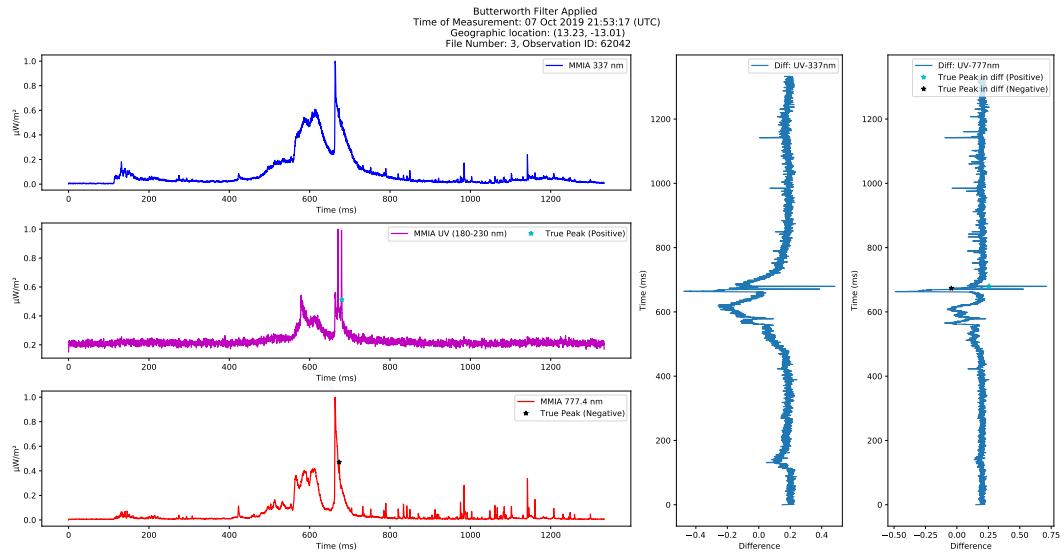


Figure 5: Left: The three signals normalized with respect to their maximum value. Right: The UV band with 337nm (middle-right) and 777.4nm (far right) subtracted respectively. On the far right the peak finding algorithm, Filter 1, has been applied.

The signal,  $\bar{F}_{\text{Rel}}$ , behaves as expected: a clear positive peak followed by a clear negative peak. Additionally, the UV-777.4nm signal appears more stable around 0 than the UV-337nm signal in this case. This is due to the strong signal in 337nm before the actual peak. For this reason, the peak finding algorithm, Filter 1, is applied to the UV-777.4nm signal. However, the assumption of applying the filter first on the positive  $\bar{F}_{\text{Rel}}$  and afterwards on the negative  $\bar{F}_{\text{Rel}}$  does not appear to work as intended. The algorithm is not able to accurately detect the 777.4nm peak and the UV peak, even though it appears to be close. For this reason, this method has not been pursued further. A proper peak finding algorithm, which is able to detect both positive and negative peaks is necessary, if this method is to be investigated further, but has not yet been made.

Another concern regarding this approach, is that the true values between the signals are not considered. By normalizing the signals, the information of the ratio between the signals is lost, which must be assumed to potentially lead to false triggers.

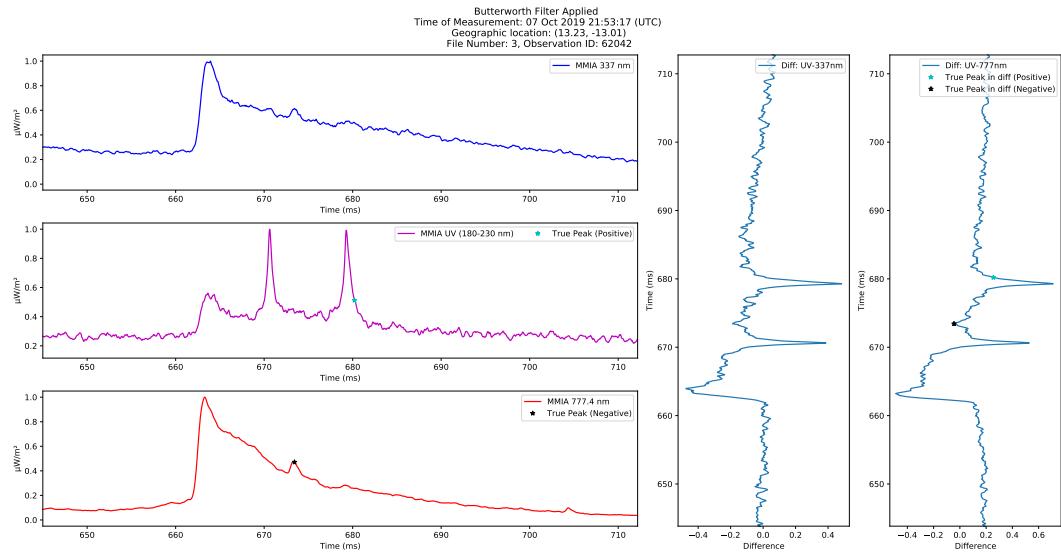


Figure 6: A close of up the peak shown in fig. 5.

## 2.4 Method 3 - ISUAL

The second method that has been tried out, is an adaptation of that presented by [Adachi *et al.*, 2016], using the data provided by The Imager for Sprites and Upper Atmospheric Lightning (ISUAL). They present an approach, in which they search for peaks in the Far UV (FUV) spectrum. They conclude that the ratio of the 337nm band and the broadband 609-753nm band, at the corresponding time of the UV peak, is a good indicator for determining TLEs. They also reject the idea of using the ratio of the 777nm band and the broadband 609-753nm band, concluding that it is not a good indicator for TLEs.

Since ASIM does not measure the broadband 609-753nm band, an adapted approach is tested; finding the UV peak, and determining the ratio of the 337nm band and the 777.4nm band. On fig. 7 the general idea shown: The first step is to find the peak in UV. In order to do this, the peak finding algorithm, *Filter 1*, presented in [Olesen & Jørgensen, 2020], is used on the UV band with the filter coefficients:

$$\text{NS} = 500, \text{threshold} = 0.03$$

This allows even very slow peaks, with risetimes up to 500 samples, and very weak signal changes down to  $0.03 \mu\text{W m}^{-2}$ . This is another reason that the noise removal from the initial low-pass Butterworth filter is important; without it, the natural noise fluctuations will start influencing the peak finding, and it would be necessary to increase the threshold above  $0.03 \mu\text{W m}^{-2}$ . After finding the peak, the ratio between the 337nm and 777.4nm at the same time as the UV peak is determined (for reference, see the orange markers on fig. 7).

As seen from [Adachi *et al.*, 2016] a significant time delay between the UV peak and 777.4nm peak can be present with TLEs, which is also the case as seen on fig. 2. Under the assumption that the delay and peak value might have some sort of correlation, a simple search for the maximum value in the 777.4nm around the UV peak is made, in a  $\pm 5$  ms interval. It is important to note, that this search does not take into account whether or not a 777.4nm peak is actually present; the maximum value found might very well just be the highest noise level present in the data. Due to the low-pass filtering, this should be very close to 0. Therefore, for 777.4 peak values close to 0, the time delay should not be interpreted as anything meaningful, as it would be assumed to be random due to the nature of random noise.

As a final condition to Method 3, the only desired triggers are those corresponding to strong positive Cloud to Ground (CG) lightning. In order to achieve this, a comparison to the Global Lightning Detection Network (GLD360) dataset is used. If a positive CG lightning, above 15 kA, is recorded within a 300km radius of the ASIM trigger, the ASIM trigger is accepted. Since we are working with Level 1 data from ASIM, which has a 20 ms uncertainty in the stored date [ASIM Science Data Centre, n.d.], an uncertainty is necessary to include when comparing it to the GLD360 dataset. By searching the GLD360 for a positive CG corresponding to the peak found on fig. 2, which will be assumed to correspond to a positive CG, a time difference of roughly 26ms was found. This could indicate some uncertainty in the GLD360 dataset as well. Therefore, a search within  $\pm 30$  ms of the ASIM UV peak was chosen, to hopefully mitigate the uncertainty. This could potentially give false triggers, but will have to do until a better uncertainty measure can be determined.

Due to the inconclusive results of the first two methods presented, and the need for further developing a peak finding algorithm for Method 2, it was decided to move forward with Method 3, as described in the Results section.

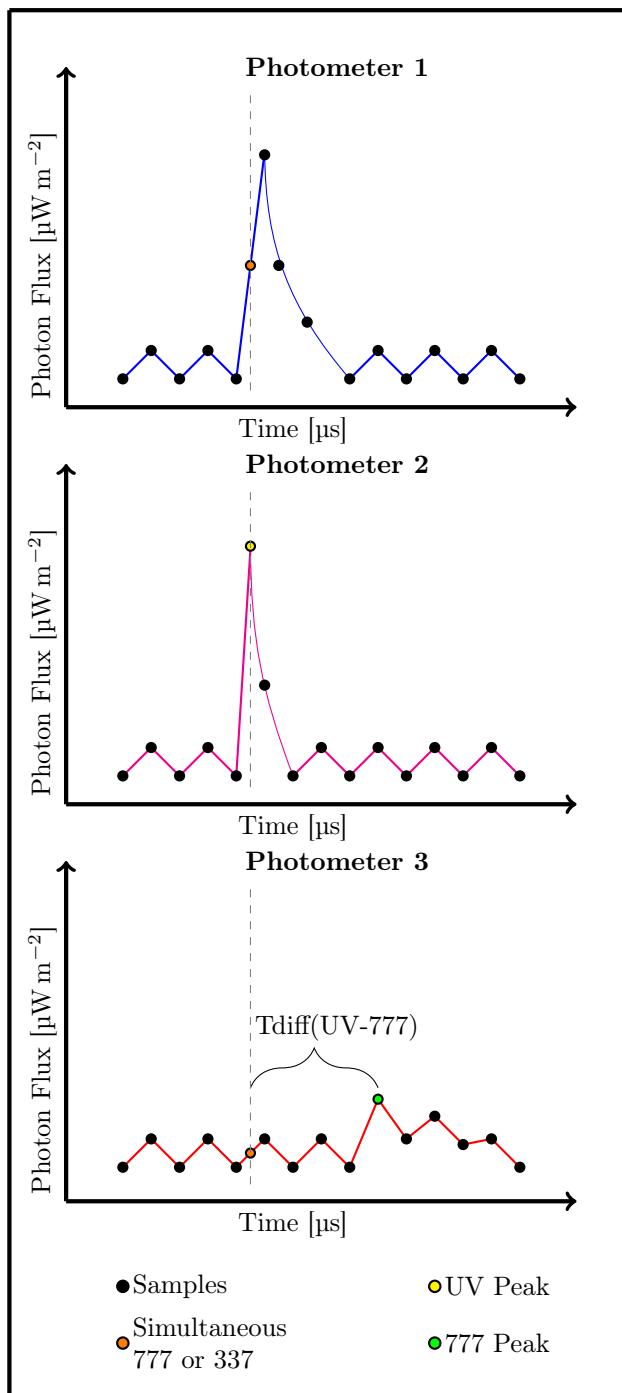


Figure 7: A sketch of the idea behind Method 2. A peak is first found in UV (Photometer 2). The corresponding value of the 337nm (Photometer 1) and the 777.4nm (Photometer 3) band is then used to determine the ratio 337nm / 777.4nm. Finally, the actual peak is found in 777.4nm, and the time delay, Tdiff, is determined as the time between the 777.4nm peak and the UV peak.

### 3 Results

After deciding the move forward with Method 3, the method has been applied to the full 2019 dataset from ASIM. This corresponds to a total of 295476 files, and returns 229 files which each containing atleast 1 trigger that satisfies all criterias, as described in section "Method 3 - ISUAL". For all 229 files, the following 4 parameters are logged: UV Peak value, Ratio between the simultaneous 337nm and 777.4nm peak values at UV peak time, time difference  $Tdiff$ , and the 777.4nm peak value is saved. Please refer to fig. 7 for precise definitions of the parameters.

First a plot of the Ratio can is made in order to follow the same approach as described in [Adachi *et al.*, 2016]. Additionally, the thresholds determined by [Adachi *et al.*, 2016] of how likely a TLE is to occure, depending on the ratio of their 337nm divided by their broadband 609-753nm is also included as a reference. The result of this is shown on fig. 8. Due to the very large differences between the extremes of the ratios, a log10 plot is made aswell, as seen on fig. 9. An important thing to highlight, is that as seen from fig. 8, a few negative ratios show up. Intuitively this should not be possible, as the instrument only measures values above, and including, zero. The problem is most likely due to the applied Butterworth filter, but this will be investigated further in the Discussion. However, this also means that the negative values cannot be included in the log10 plot, which should be kept in mind when comparing the two plots.

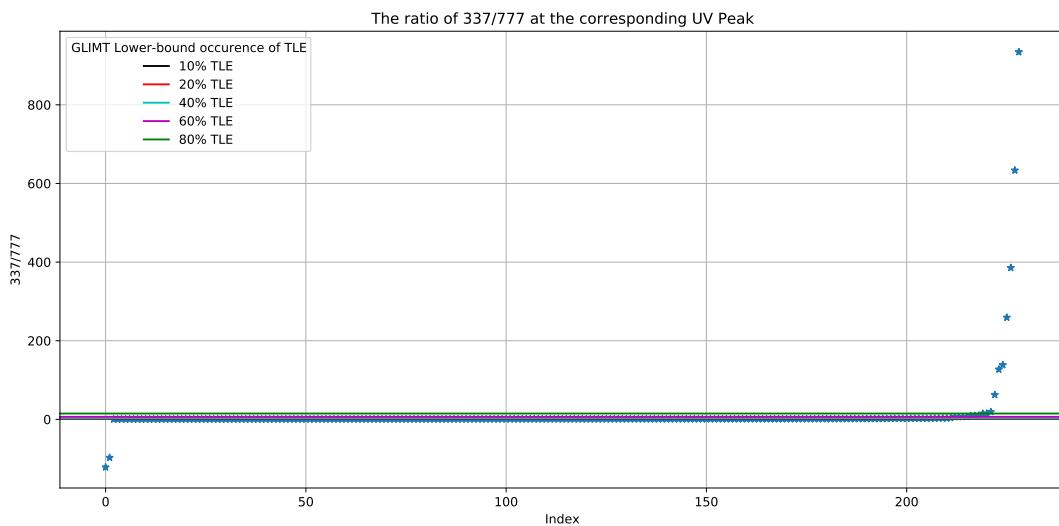


Figure 8: The ratios of 337nm divided by the 777.4nm at the UV peak time.

## Results

Rasmus C. Jørgensen

12/12-2020

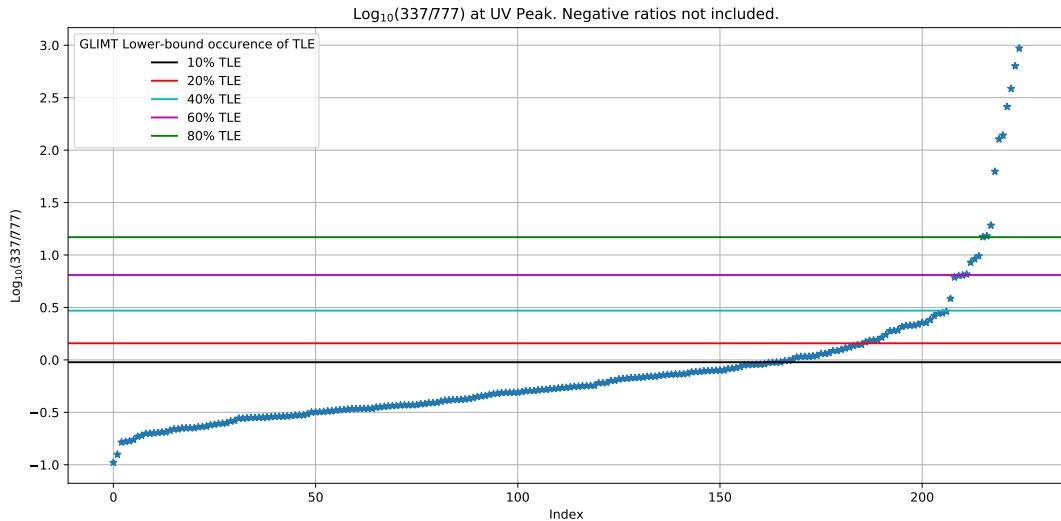


Figure 9: The log10 of the ratios of 337nm divided by the 777.4nm at the UV peak time.

Afterwards, it is possible to look at the relation between the four parameters that are logged for the triggers. The easiest way to get a quick overview is to plot the four parameters against each other, as shown on fig. 10. A semi-log10 plot is made to better showcase the true correlation between the parameters. This means that the Ratio, 777 true peak value, and UV peak value are converted to log10 scale, but the time difference is kept in the normal domain. The resulting plot is shown on fig. 11. A similar plot including the 60% GLIMT Lower-bound occurrence of TLEs, as shown on fig. 9, also exists, but is not included in the report. Please run the "Plot\_SpritesOutput.py" script and see figure 6, if this is of any interest. It is omitted in the report due to a bug with exporting 3D plots including surfaces and point plots in the matplotlib package, where the surface overwrites the points, rendering the resulting exported figure useless. However, analysing the plot in a Python environment can be of interest, and is therefore kept in the code. Please refer to Appendix A for a user-guide on how to reproduce the plots presented in this section, along with some of the implemented features for studying the events.

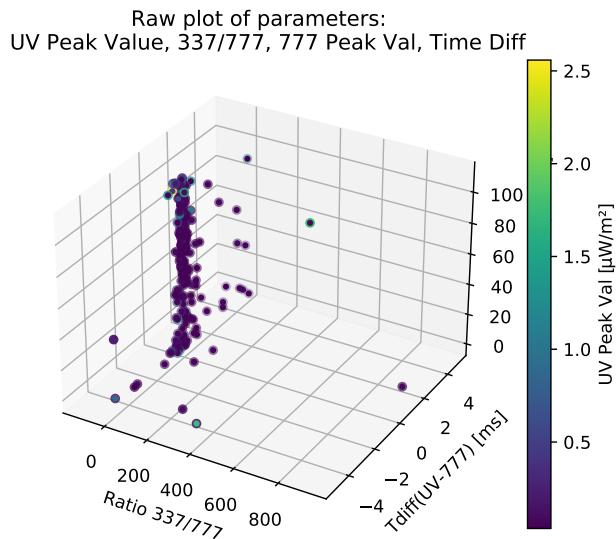


Figure 10: A plot of all 4 logged parameters from Method 3. The x,y,z axis show the Ratio,  $Tdiff$  and 777 true peak value respectively, with the color-bar showing the UV Peak intensity. Please be aware that the z-axis label is missing due to a bug in the python matplotlib package with exporting 3D-plots.

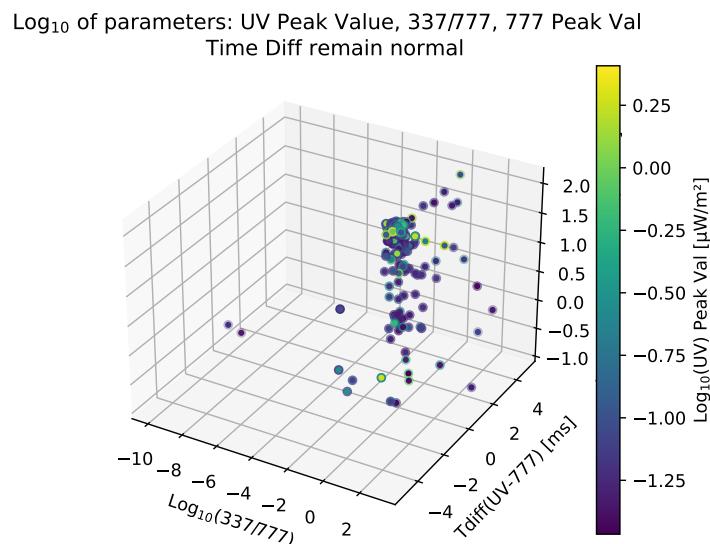


Figure 11: A semi-log10 plot of all 4 logged parameters from Method 3. Be aware that the axis  $Tdiff$  is kept in the normal domain. The same bug with the missing z-axis label applies here, but it represents the  $\log_{10}(777)$  peak value.

## Results

Rasmus C. Jørgensen

12/12-2020

Another interesting approach could be to look at a 2D plot of the relation between the  $\log_{10}(777)$  peak value,  $\log_{10}(\text{UV})$  peak value, and  $T_{\text{diff}}$  to see if there is any correlation here, as shown on fig. 12. Once again, we can use the 60% GLIMT lower-bound occurrence of TLEs to filter the plot, which results in fig. 13.

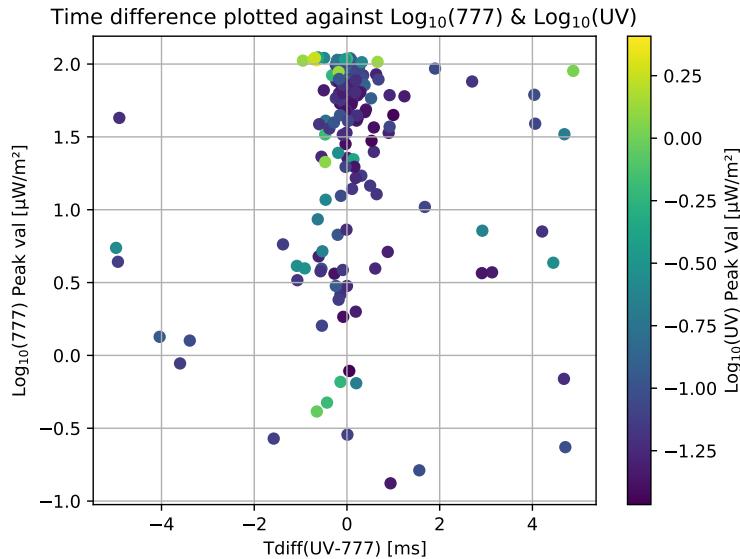


Figure 12: Plot showcasing the relation between the  $\log_{10}(777)$  peak value,  $T_{\text{diff}}$ , and the  $\log_{10}(\text{UV})$  peak value

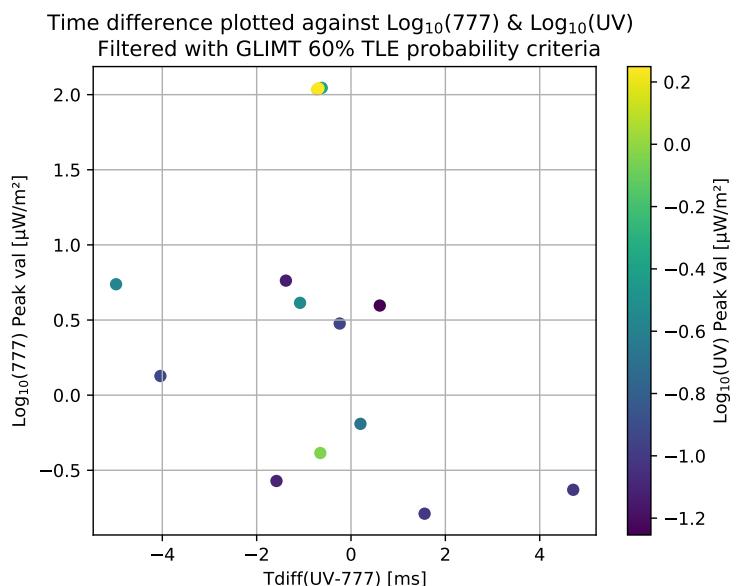


Figure 13: Filtering fig. 12 with the 60% GLIMT lower-bound occurrence of TLEs

On a final note, we can check the spatial distribution of the 229 files passing Method 3, and plot them with respect to geographic location. In other words, creating a 2D histogram of the files with respect to latitude and longitude. This is more or less a sanity check of the method, checking whether the distribution is somewhat in agreement with what might be assumed to be the case.

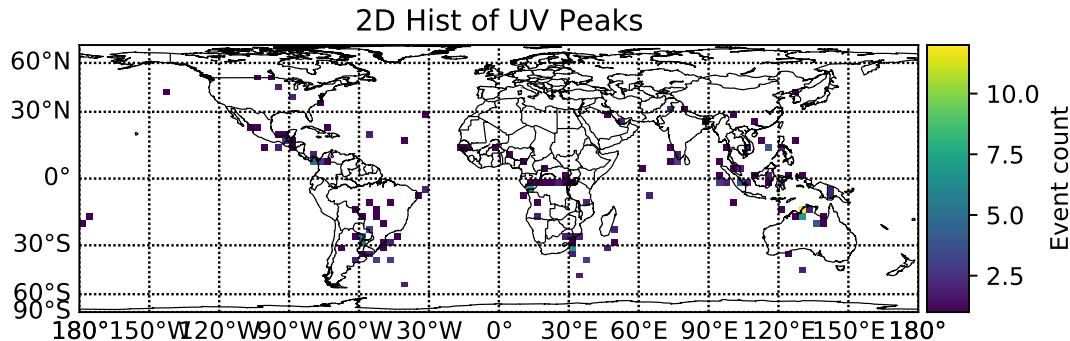


Figure 14: 2D histogram showing the spatial distribution of files passing Method 3

## 4 Discussion

To properly interpret the results, it is necessary to inspect the events causing the respective parameters. To start off with, the two negative ratios found on fig. 8 will have to be looked at. Negative ratios should not be expected to occur, since the ASIM instrument only measures positive and zero values. Plotting the observation, it becomes apparent that the peak which has triggered the event, is the very last sample in the datafile, as seen on fig. 15. The spike in UV can be a consequence of the applied Butterworth filter, and issues with boundary conditions of the filter. However, since it has passed, a positive CG lightning must also be closeby, but does not appear in the file. It can perhaps be seen in the following datafile, but might not have passed the peak finding filter. The negative value is definitely an artifact from the Butterworth filter, which appears to smooth the 777.4nm values just below zero at the end, which is an unfortunate side effect. For later table values, these issues will be classified as "Butterworth Filter Error". The same error appears in the other negative value, where the boundary value also triggers.

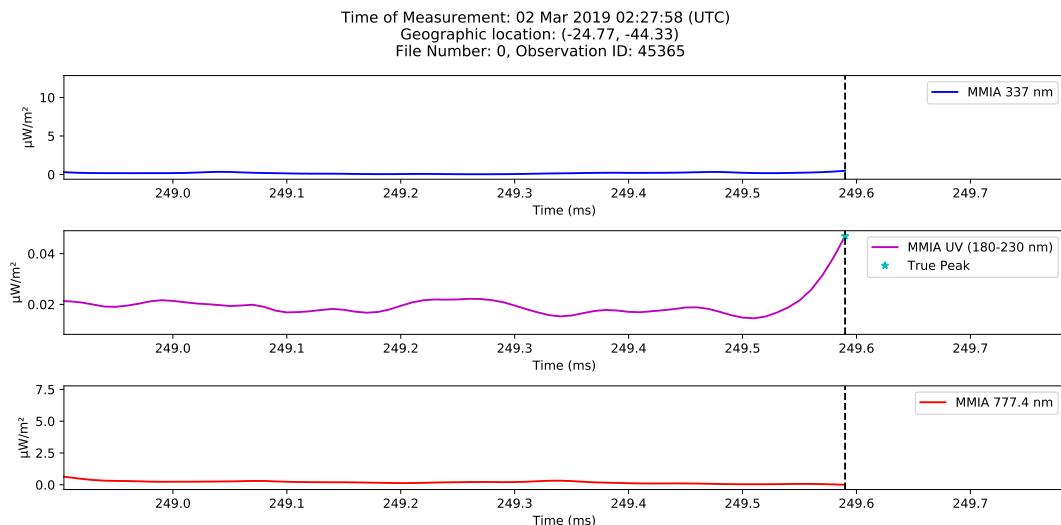


Figure 15: The peak causing the lowest measured ratio, which is well below 0. Note that we are zoomed in on the very last frame of the dataset, and the last samples appear to cause the anomaly in the UV spectrum.

Next up, let us inspect the highest values instead, passing the 80% lower bound occurrence of TLEs from [Adachi *et al.*, 2016]. A total of 10 events from the 2019 dataset has passed this lower bound, clearly shown on the Log10 plot fig. 9. All 10 events have been inspected and classified, and can be found in table 1. Please note that interesting events do occur below the 80% lower bound, but since 228 triggers have been found, only the highest will be inspected for now. The photometer and camera head unit (CHU) data from each classification will be presented, but not more once, as events within each classification share the same features. The instrument timestamp and (latitude,longitude)-coordinate from the ASDC server is also included in table 1, so all events are easy to find again. The table value "#Ratio from Top" refers to the ratios location with respect to the highest ratio, which means that number 1 corresponds to the highest ratio.

#Ratio from Top	Instrument Time	Lat	Lon	Classification
1	2019-11-05 10:49:04.37	40.637	-143.082	Butterworth Filter Error
2	2019-02-19 02:08:51.84	-51.666	-41.612	777.4 Saturation
3	2019-01-30 17:02:58.51	-13.408	130.426	Blue Jet Structure
4	2019-01-30 17:03:01.19	-13.575	130.554	Blue Jet Structure
5	2019-01-30 17:03:20.52	-14.538	131.297	Blue Jet Structure
6	2019-01-30 17:02:47.84	-12.909	130.044	Blue Jet Structure
7	2019-09-27 17:25:42.38	13.3003	114.598	Pulsating 337 & 777.4
8	2019-01-03 14:08:11.73	-17.442	130.682	Blue Jet Structure
9	2019-09-23 20:37:38.44	6.772	94.751	Buterworth Filter Error
10	2019-02-11 02:30:09.82	-36.816	-53.903	Pulsating 337 & 777.4
17	2019-04-29 00:51:28.34	-3.692	12.790	Pulsating 337 & 777.4

Table 1: Classification of the events causing the 10 highest ratios shown on fig. 9, with the additional number 17 from the top. The 17th one is mostly a proof that interesting events do occur below the GLIMT 80% lower bound occurence of TLEs, but the frequency appears to significantly decrease.

Since the highest trigger is also due to an error with the Butterworth Filter, matching that of fig. 15, except the 777.4 barely stays positive, this event will not be looked further into for now. However, number 2 from the top is the only event in the top 10 with the "777.4 Saturation" classification. The trigger causing this, can be seen on fig. 16. Comparing the 337nm peak to the 777.4nm peak, it is apparent that the 777.4 saturates, which effectively drop the value to zero. However, this is not the case in the 337nm spectrum, which causes the ratio between the two to explode. Inspecting the CHU data of the event, nothing significant stands out for now, and the event will be assumed to be a very strong lightning strike.

## Discussion

Rasmus C. Jørgensen

12/12-2020

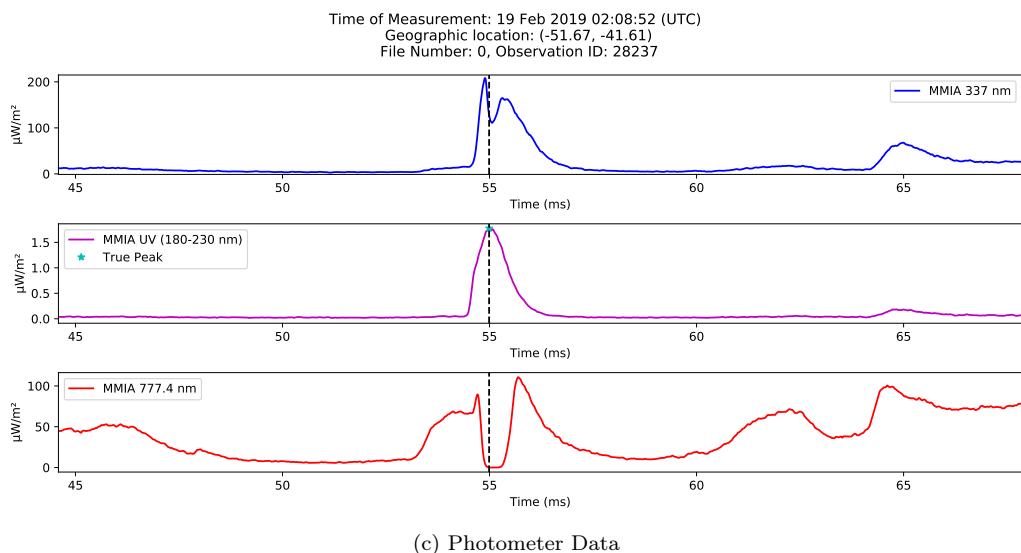
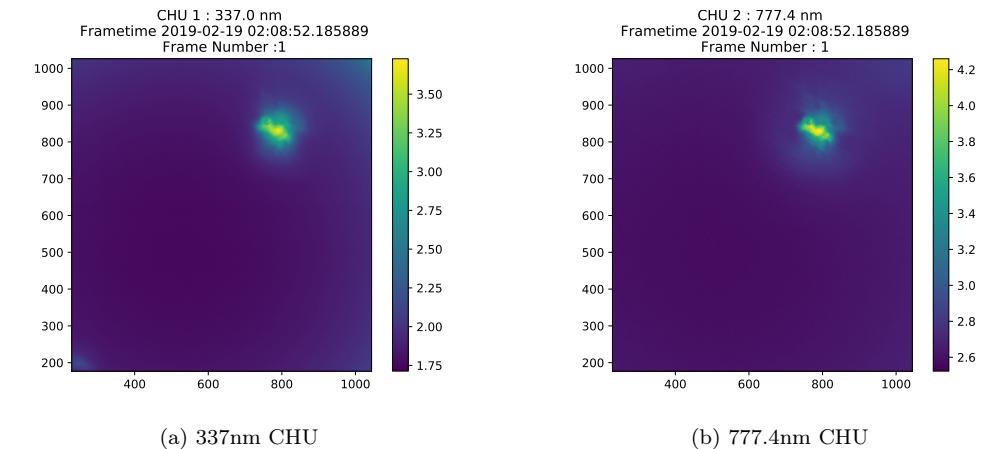


Figure 16: The trigger causing the second highest ratio. The 777.4nm spectrum appear to saturate causing it to drop to effectively zero, causing the ratio to explode, leading to the "777.4 Saturation" classification.

Next up is the classification "Blue Jet Structure". This is is by far the highest represented classification in the top 10, corresponding to 50% of the triggers. The classification is based on the Nauru event presented in [Torsten Neubert, n.d.], which means that we are looking at a fast and strong peak in the 337nm spectrum, with a long tail consisting of small pulses, a weak 777.4nm signal, and a strong UV peak aswell. Such an event, can for instance be seen on fig. 17, which shows the fourht ratio peak on fig. 9. Comparing the photometer data to the CHU data, the upward going peak from the source is also very apparent in the 337nm spectrum, and does not appear in the 777.4nm spectrum, which is consistent with the event seen in [Torsten Neubert, n.d.].

## Discussion

Rasmus C. Jørgensen

12/12-2020

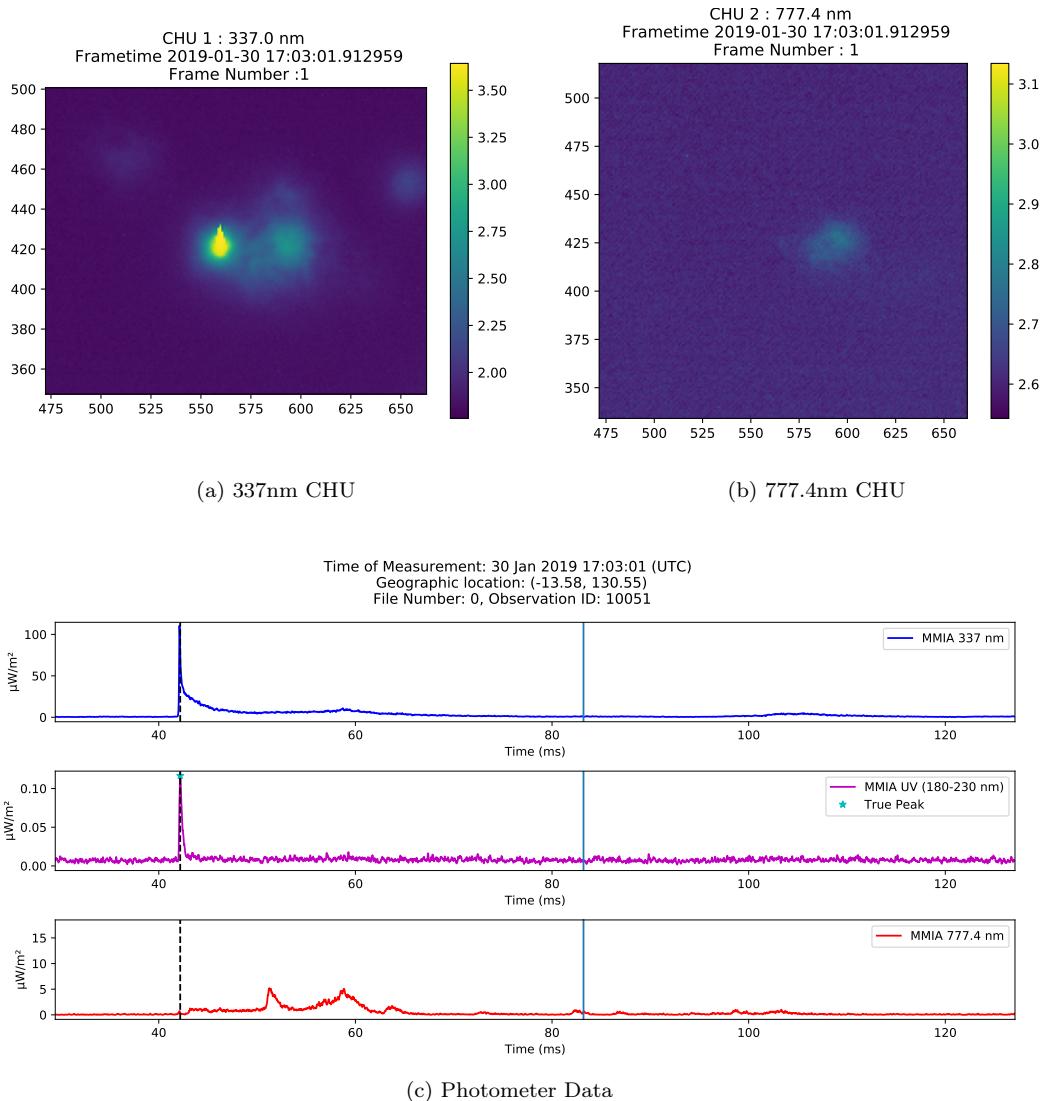


Figure 17: The fourth highest ratio from fig. 9. This event appear to be of the same nature as that presented in [Torsten Neubert, n.d.], leading to the classification of "Blue Jet Structure".

## Discussion

Rasmus C. Jørgensen

12/12-2020

The last classification is the "Pulsating 337 & 777.4". This is based on the fact, that these contain events which seem to pulsate with a frequency somewhere between 70ms and 120ms, decreasing for each subsequent peak. They contain very strong signals in all three spectra, as seen on fig. 18. Additionally, the 777.4nm spectrum appear to remain much stronger than the 337nm spectrum, which is also reflected in the CHU data, as shown on fig. 19 and fig. 20. The initial 337nm peak is almost double the amplitude than that of the 777.4nm, however the amplitude of the 777.4nm peak appear to remain stable for the subsequent 3 peaks, and does not drop off before the last fourth peak. This is not the case for the 337nm, where the subsequent peaks after the initial peak has a much lower amplitude than the initial peak. Additionally, the UV peaks appear to behave similar to that of the 337nm, with a strong initial peak, that quickly drops off. Surprisingly, the filter is still able to accurately locate the very weak subsequent UV peaks aswell.

Disregarding the initial peak, the subsequent peaks seem to be in agreement with the single peak sprite presented in [Adachi *et al.*, 2016], with strong 777.4nm, weaker 337nm, and a small peak in UV. However, the initial peak with stronger 337nm than 777.4nm and the strong pulsating nature is unaccounted for in [Adachi *et al.*, 2016].

Disregarding the source and classification of these events, the main focus is the evaluation of the applied method. Only looking at the ratios of the 337nm and 777.4nm at the UV peak time, appears to be promising. The passing files could easily be filtered with a ratio cutoff, similar to the 80% lower bound found by [Adachi *et al.*, 2016]. By inspecting all 228 passed events in fig. 9, a statistical approach could be used to make proper estimations of the lower bound probabilities based on the 337/777.4 ratio, instead of relying on the 337/broadband Red from [Adachi *et al.*, 2016].

The boundary error with the Butterworth Filtering is also an unfortunate side effect, as it appears to influence the extrema of the ratios, which would ideally be the ones we are also interested in. It is interesting that the false triggers are not filtered out by the 360 GLD positive CG check, which would indicate that a more thorough evalution of the error is probably necessary.

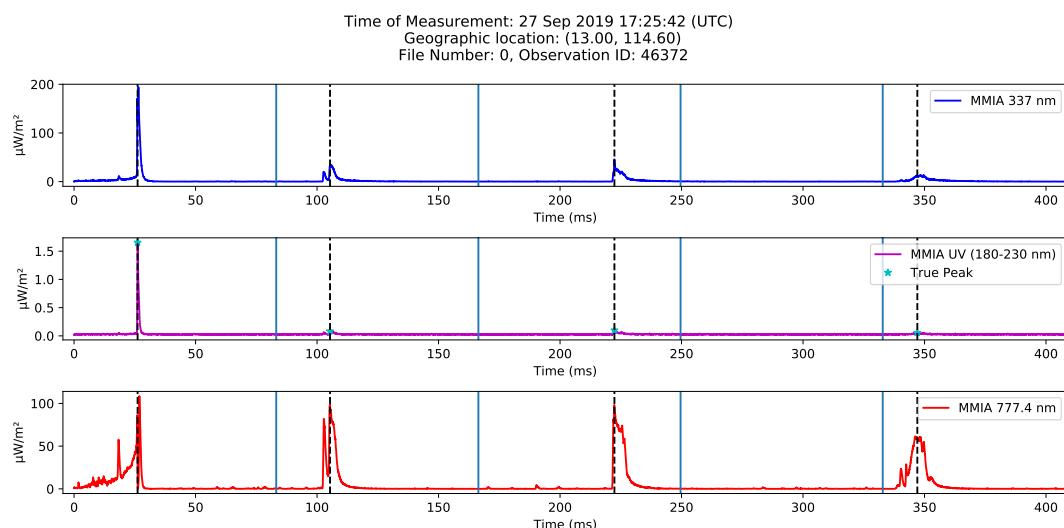


Figure 18: The photometer data of the 7th highest ratio from table 1. Note that each frame appear to contain its own peak, roughly 100ms after eachother.

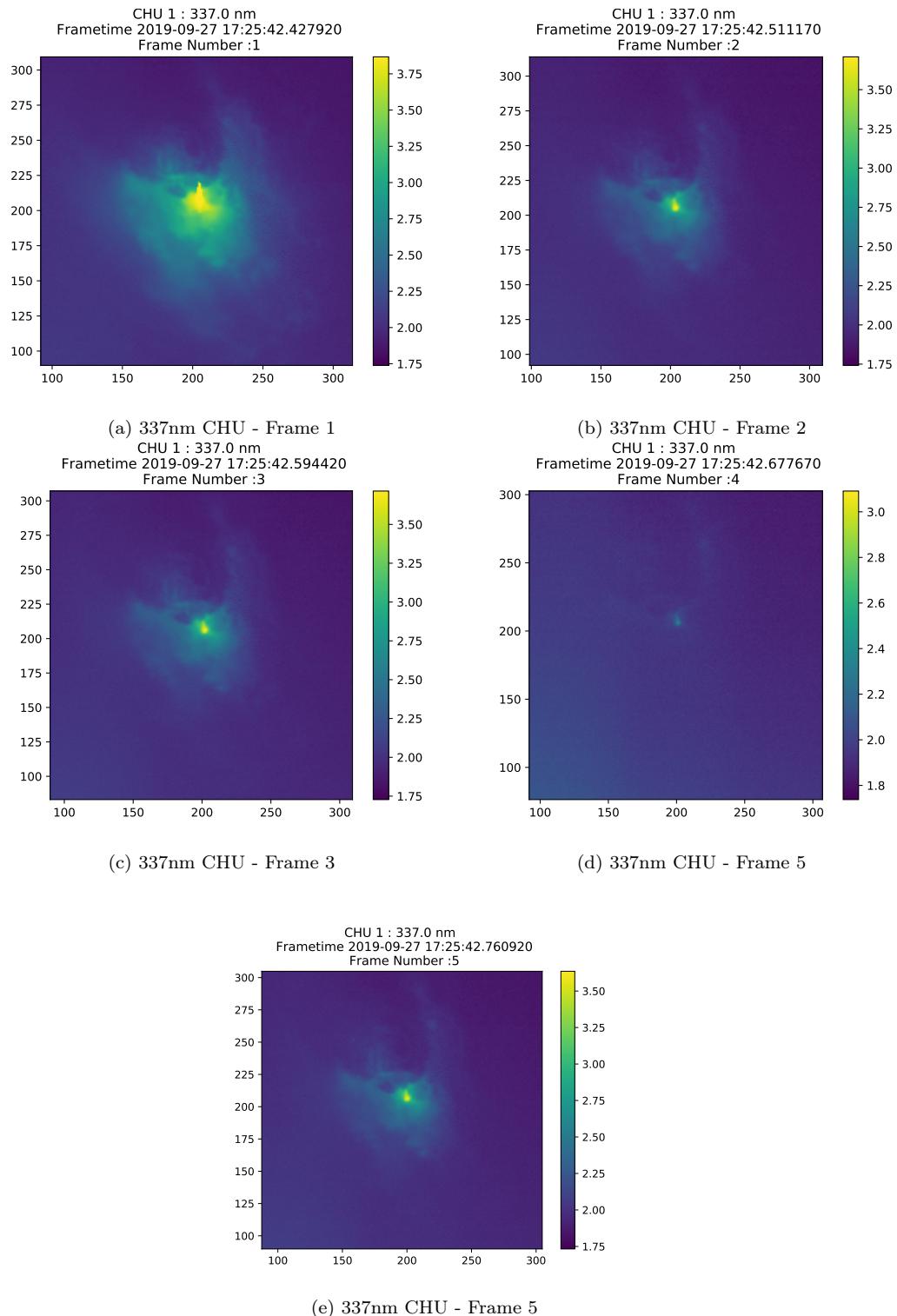


Figure 19: The 337nm CHU data from the event shown on fig. 18.

## Discussion

Rasmus C. Jørgensen

12/12-2020

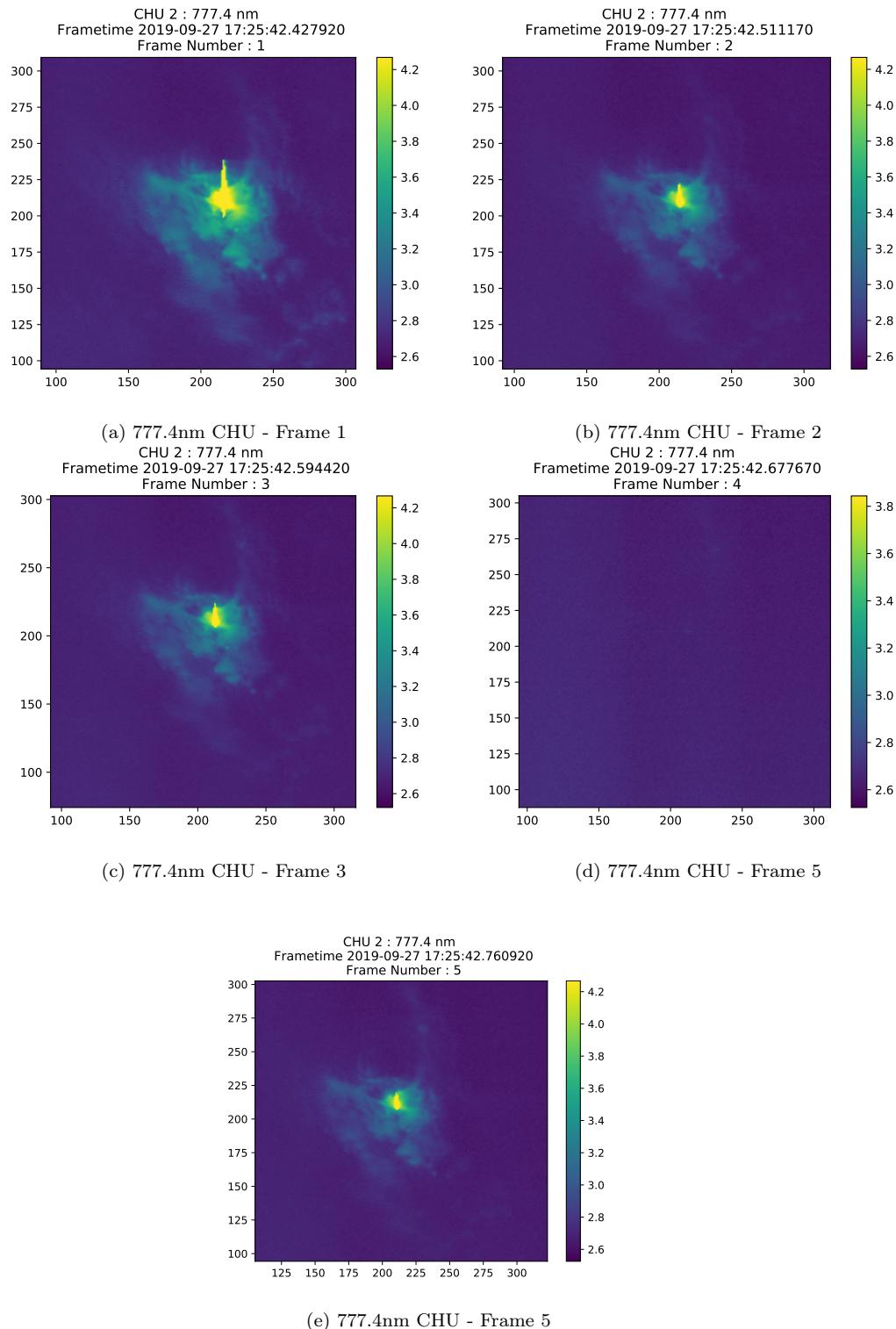


Figure 20: The 777.4 nm CHU from the event shown on fig. 18

Moving on to the correlation between the four logged parameters, as shown on fig. 11, and keeping the cross section in mind without the ratio as seen on fig. 12, a clear grouping of strong UV peaks corresponding to strong 777.4nm peaks, with a slight delay in the UV peak compared to the 777.4 nm peak is shown, at roughly  $\log_{10}(777) = 2$ ,  $T_{\text{diff}} = -0.5$ ,  $\log_{10}(\text{UV}) = 0.1$ . Upon filtering with the GLIMT 60% lower bound as shown on fig. 13, 3 events from this group remain. Zooming in on these three events, see fig. 21, and inspecting the events, reveal that the weakest  $\log_{10}(\text{UV})$  peak (right most) corresponds to the 10th event in table 1 of classification "Pulsating 337 & 777.4". The middle event corresponds to the 2nd event in table 1 of classification "777.4 Saturation", and the left event correspond to the 7th event from table 1, which is again of classification "Pulsating 337 & 777.4", and is the event shown on fig. 18, fig. 19, fig. 20. Further inspecting the events from fig. 13, it turns out that the only event from table 1 with a positive  $T_{\text{diff}}$  value is #6, with the classification Blue Jet Structure. The remaining triggers all have a negative  $T_{\text{diff}}$  indicating that the UV peak is before the 777.4nm peak.

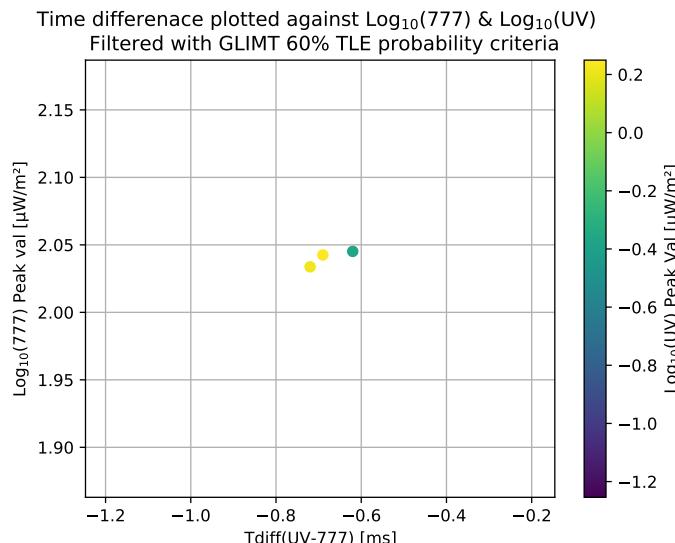


Figure 21: A zoom of fig. 13, focusing on the group of strong  $\log_{10}(\text{UV})$  values located around  $\log_{10}(777) = 2$ ,  $T_{\text{diff}} = -0.5$ ,  $\log_{10}(\text{UV}) = 0.1$

Finally the 2D histogram of the geographic location of the events, fig. 14, is very much in agreement with areas of high lightning acitivity [King, n.d.]. This is not necessarily any statistical evidence of the distribution of the events. It rather shows that they correlate with high lightning acitivity, which statisticly makes sense, as these should also be the areas with most total triggers in the ASIM dataset. Still, it is reassuring that nothing is out of the ordinary with the geographic distribution.

## 5 Conclusion

In an attempt to develop an algorithm to detect Sprites, three different methods have been tested out. Method 1 and 2 are based on an event observed in the MMIA dataset which is assumed to be a sprite, as seen on fig. 2. Method 1 attempted to remove the 777.4nm signal from the 337nm signal, but proved unsuccessful. Method 2 attempted to detect delayed UV Peaks with respect to the peaks in 337nm and 777.4nm, by normalizing and subtracting the signals from each other. This results in negative peaks corresponding to 777.4nm or 337nm peaks, and positive peak to UV peaks. The peak finding algorithm used was "Filter 1" from [Olesen & Jørgensen, 2020], but unfortunately this algorithm does not appear to be able to handle negative peak values. The method is not completely ruled out, but has not been pursued further, due to the necessity of developing a new peak finding algorithm. Finally, an adaptation of the method described by [Adachi *et al.*, 2016] was tested. They use the broadband red 609-753nm as a reference, and determine the ratio of 337nm and 777.4nm with respect to the broadband. Since MMIA does not measure the broadband red, the ratio between the 337nm and 777.4nm is determined instead. This appears to be a promising approach to detecting TLEs in general, and using the 80% lower-bound ratio probability for TLEs from [Adachi *et al.*, 2016], shows that 7 out of 10 total passing events are indeed TLEs, with 3 potential false triggers. The time difference between the UV peak and 777.4nm peak is also determined, and 9 out of the 10 events passing the 80% lower bound proved to have a negative time difference, indicating that the UV peak happens before the 777.4nm peak. Therefore, a robust approach to searching for TLEs could be to search for UV peaks, where the ratio between 337nm and 777.4nm is high, and the UV peak happens before the 777.4nm peak. The 777.4nm peak value and UV peak value is also logged, but no concrete correlation is found here.

The original goal was to specifically detect sprites in the dataset, which has not been successful. However, the approach in Method 3 appears successful in a broader way, being able to detect TLEs in general. The majority of the top 10 highest ratios proved to have Blue Jet structure consistent with the event presented in [Torsten Neubert, n.d.]. A pulsating category is however also present, but a concrete classification is not given to these events in this report. These could perhaps prove to be sprites as some of the pulses share similar structure to the single-peak sprite presented in [Adachi *et al.*, 2016], however the strong pulsating nature and strong initial 337nm peak is not accounted for.

Finally, it is important to keep in mind that only the most promising candidates have been examined as of now, due to the limitations and scope of the project. A more thorough inspection of all events passing Method 3 should be carried out, to properly estimate the lower bound probabilities of TLEs based on ratios, and the true ratio of true to false triggers using the method. Changing the filter variables in the peak finding algorithm *Filter 1* could perhaps also prove to yield better results.

## References

- Adachi, Toru, Sato, Mitsuteru, Ushio, Tomoo, Yamazaki, Atsushi, Suzuki, Makoto, Kikuchi, Masayuki, Takahashi, Yukihiro, Inan, Umran S., Linscott, Ivan, Hobara, Yasuhide, Frey, Harald U., Mende, Stephen B., Chen, Alfred B., Hsu, Rue-Ron, & Kusunoki, Kenichi. 2016. Identifying the occurrence of lightning and transient luminous events by nadir spectrophotometric observation. *Journal of Atmospheric and Solar-Terrestrial Physics*, **145**, 85 – 97.
- ASIM Science Data Centre. *MMIA Observation Download*. [https://asdc.space.dtu.dk/mmia\\_observation/](https://asdc.space.dtu.dk/mmia_observation/).
- Chanrion, Olivier; Neubert, Torsten; Rasmussen Ib Lundgaard; Stoltze Christian; Tcherniak Denis; Jessen Niels Christian; Polny Josef; Brauer Peter; Balling Jan E.; Kristensen Steen Savstrup; Forchhammer Søren; Hofmeyer Peter; Davidsen Peter; Mikkelsen Ole; Bo Hansen Dennis; Bhanderi Dan; Petersen Carsten Lunde; Lorenzen Mark. 2019. The Modular Multispectral Imaging Array (MMIA) of the ASIM Payload on the International Space Station. *Space Sci Rev*, 1–25.
- King, Hobart M. *World Lightning Map*. <https://geology.com/articles/lightning-map.shtml>.
- Olesen, Emilie Petrea Petäjämaa Wiinberg, & Jørgensen, Rasmus Christian. 2020. Analysis of blue lightning measured by ASIM. 1.
- Torsten Neubert, Olivier Chanrion, Victor Reglero Nikolai Østgaard Krystallia Dimitriadou Matthias Heumesser Lasse Husbjerg Ib Lundgaard-Rasmussen. The genesis of blue lightning into the stratosphere.

# Appendices

## A User-Guide: Running Method 3 and reproducing figures in Results section

### Reproducing plots in Results section

In order to reproduce the plots presented in the Results section, the first step is to unzip the provided zip folder *Method3.zip*. The folder contain three sub-directories: *plotResults*, *\_pycache\_*, and *runMethod3*. The directory *\_pycache\_* is irrelevant, and could probably be deleted. The directory *plotResults* contain the necessary python scripts for creating the plots, along with the datafiles located in the subdirectory *PositiveCG\_OutputFiles*. The directory *runMethod3* contains the python scripts used for actually running the algorithm described in section "2.4 Method 3 - ISUAL".

In order to plot the results, acces the *plotResults* folder and open the file *Plot\_SpritesOutput.py* in a python IDE of choice, such as for instance Spyder. The scripts are developed in Spyder 4.1.3, running Python 3.7.7 and runs for sure with the matplotlib package version 3.3.1, but will most likely run on older versions aswell.

After opening *Plot\_SpritesOutput.py* it is only necessary to change 3 lines of code, with an additional optional fourth one:

- **Line 18 (optionally changed):** The *fileName* variable is optionally changed, and is used to name the output directory where figures will be automatically saved when running the script. It is also included in the file name of the output files. Consider this as a bookkeeping name, making life easier when looking through the output.
- **Line 19 (mandatory to change):** The *dataDirFile* variable is the path to the "output-PositiveCG.txt" file. This file contains all the logged variables described in section "2.4 Method 3 - ISUAL", and are the actual data used for the plots in the Results Section. If the directory is unpacked directly on the Desktop on a windows system, this should be "C:\Users\user\Desktop\Method3\plotResults\outputPositiveCG.txt". Where *user* should be replaced with the user name of the system.
- **Line 20 (mandatory to change):** The variable *outputDir* gives the directory where the output files will be saved. This can be any path of choice, for instance if you wish to save them directly on the Dekstop, this will be " 'C:\Users\user\Desktop\Figures\_'" + *fileName* + '\'', where *user* should be replaced with the user name of the system. This will create a directory on your desktop named Figures\_*fileName* (where *fileName* is the variable from line 18), where the figures are automatically saved.
- **Line 22 (mandatory to change):** The variable *plotFileDir* is the path to the directory where the datafiles are saved. If you unpack the zip file directly on the desktop on a windows system, the path should be "C:\Users\user\Desktop\Method3\plotResults\PositiveCG\_OutputFiles\" where *user* should be replaced with the user name of the system. These are the files used to plot the photometer and CHU data presented in the Discussion section.

Afterwards simply run the script, and all figures from the section Results will be created. An advantage of running the figures in a python environment, is that the 2D plots are interactive. This means, that you can press any point on the plot, for instance any point on fig. 9, and it will bring up the photometer and CHU data corresponding to the event. It might take a little

while before it appears, since the figures are not saved anywhere, but rather created when pressing the plot point. An indication that the script has registered a mouse click is, that it will print the file name in the console. This should enable the user to easily get an overview of the events, and analyse the results. If no CHU plots are created, it is because the given event does not contain CHU data, which happens from time to time. It is *not* possible to open multiple events at the same time, which means that before pressing a new event on a plot, all photometer and CHU plots should be closed. A few final notes regarding the different scripts in the *plotResults* folder:

- The script *Click2Plot\_Func.py* contains the python script called when pressing a point. It calls the peak finding algorithm and creates the plots. In order to enable interactive plots, the *picker* argument for matplotlib figures should be set to True, as seen in *Plot\_spritesOutput.py*. An "onPick" function is then defined, which calls Click2Plot and connects it to the plot through the line *fig.canvas.mpl\_connect('pick\_event', onPickFunction)*
- The script *Filters\_RCJ.py* is a file containing different designed filter types, such as for instance the Butterworth Filter and the "Filter 1" peak finding algorithm. It is intended as a library of filters.
- The script *PlotCHU\_Pix.py* is provided by Krystallia Dimitriadou and is used for plotting the CHU data. This is used in the Click2Plot\_Func.py script.
- The *ReadMMIAfits\_OCRE.py* script is provided by Olivier Chanrion and is used to read the MMIA files.

### Running Method 3

If the user instead wishes to run the algorithm the directory "runMethod3" should be accessed. The main script is the python script *Sprites\_Filter\_New.py* which contains the algorithm itself. This script also uses the *ReadMMIAFits\_OCRE.py* and *Filters\_RCJ.py* scripts as described above. Additionally it contains the *Vaisala.py* script, which is used to connect and compare to the GLD360 dataset, and has been provided by Oliver Chanrion, created by Lasse Husbjerg. Finally, the script "getFrameTime.py" is an optional script, that is used in combination with the *PlotCHU\_Pix.py* script. As shown on line 279, it is used to get the "FramesOfInterest" variable. This can be used as an input to *PlotCHU\_Pix.py* script, see line 282. It finds the frame where the peak is detected, and returns the frame right before it, the current frame of the detected peak, and the frame right after the peak frame. If the peak is detected in either the first or last frame, it will return only the frame before, or after, whichever is possible, along with the frame of the peak itself. If more than one peak are detected, it will return all frames.

When running *Sprites\_Filter\_New.py* simply change the variable *dataDir* on line 31 to the directory where the raw files are saved, and run the script.