UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Escola Superior d'Enginyeries Industrial, Aeroespacial i Audiovisual de Terrassa

Master's Degree in Space and Aeronautical Engineering

Tools for validation of the Lightning Imager sensor on the 3rd generation of the METEOSAT weather satellite

Master Thesis

Author: M. Núria Partal Camps
Supervisor: Joan Montanyà Puig

Spring term 2019/20

Index

Table of figures

Table of tables

Declaration of honour

I declare that,

the work in this Master Thesis is completely my own work, no part of this Master Thesis is taken from other people's work without giving them credit, all references have been clearly cited.

I'm authorised to make use of the research group related information I'm providing in this document.

I understand that an infringement of this declaration leaves me subject to the foreseen disciplinary actions by The Universitat Politècnica de Catalunya - BarcelonaTECH.

M. Núria Partal Camps	_____	09.11.2020
Student Name	Signature	Date

Title of the Thesis:

Tools for validation of the Lightning Imager sensor on the 3rd generation of the METEOSAT weather satellite

Abstract

The 3rd generation of the METEOSAT (MTG) will be equipped with a lightning imager sensor (LI) to detect and locate lightning flashes. The field-of-view of the LI-MTG will cover Europe and Africa.

Since the LI-MTG will observe optical emissions from lightning from cloud tops, it will be able to provide total lightning activity in thunderstorms including intra-cloud and cloud-to-ground.

At ground, the unique system that detects total lightning with high efficiency is the Lightning Mapping Array (LMA). The UPC Lightning Research Group operates the Ebro LMA network at the Delta de l'Ebre area. This network will be used to validate the MTG LI performance when it will be in operation.

To do that, the UPC LRG group has recently compared the ISS-LIS imager sensor on the International Space Station and the needs have been defined.

The aim of this project is to follow the comparison between the ISS-LIS imager and LMA to define the needs for LI-MGT. To do that, data from the Geostationary Lightning Mapper (GLM) will be compared with the Colombia LMA at Barrancabermeja, as both LI-MGT and GLM are geostationary satellites.

With the results obtained the bases for the final validation tool of MTG LI will be set and also will be reached a better understanding of the data gathered by a lightning detection system on board of a geostationary satellite.

1. Lightning Detection Systems

It was Greek philosophers who started what would be the predecessor to what nowadays is meteorology [1]. Aristotle wrote “Meteorological” around 340 B.C, where he tried to deduce some things about weather, and for about 2000 years that was the only work on this subject.

In 1593, Galileo invented the thermometer and years later, in 1643, Evangelista Toricelia invented the barometer. From that moment, various inventions were created to measure wind speed, levels of precipitation and other meteorological phenomena.

But it wasn’t until the 1740’s that Benjamin Franklin saw similarities between the electricity he was using in his experiments and lightning. This led him to make experiments to prove storm clouds were electrically charged and years later to the invention of the lightning rod [2].

A long time has passed from that days, and now buildings, cars and aircrafts are made taking into account the damage a lightning strike can do, and therefore using the appropriate protective measures that have been created as more information about lightning and their characteristics has been discovered.

This information is obtained with different systems that are able to track lightning from the ground or from space and obtain their characteristics.

While ground-based lightning locating systems have been used since the 1920s, space-based ones are quite new considering they have been operating for almost two decades [3][4].

Space lightning observation systems have become the key piece to achieve real-time global measurements of lightning activity, making them the future of lightning observation.

Nowadays, both ground and space lightning detection systems work together, complementing each other and in numerous occasions ground systems are used as tools for validation of the new space systems created.

In this chapter will be explained three of the systems currently used to obtain information about lightning: the Lightning Mapping Array, the GOES-R Geostationary Lightning Mapper and the Meteosat Lightning Imager Sensor.

1.1. Lightning Mapping Array

The Lightning Mapping Array or LMA, is a three-dimensional lightning location system. LMA measures the time of arrival of the very high frequency (VHF) emissions produced by lightning to determine the location of their sources and with that information creates a three-dimension map of the lightning activity.

The system was developed by researchers at the New Mexico Institute of Mining and Technology and was originally used in Oklahoma in June 1998.

There are various networks currently operating, for example in Europe there are the Ebro LMA (ELMA) and Corsica's SAETTA; and in USA there are the Oklahoma LMA (OKLMA), the North Alabama LMA (NALMA), the West Texas LMA (WTLMA) among many others. The network studied in this project is located in Barrancabermeja, Colombia (COL-LMA).

LMA have a very high detection efficiency and can detect both cloud to ground and intracloud discharges, which makes it a very useful tool to perform calibrations for satellite-based sensors as GLM or the future Meteosat LI. [5]

1.1.1. System characteristics and operation

LMA consists in a set of VHF antenna stations (between 7 and 20) surrounding a central station, this station is the one that calculates the location and emission power of the lightning source using the Time of Arrival (TOA) technique. These stations are usually placed over a region of 60-80 km diameter and there is a distance of 15-20 km between each other. The VHF antennas can detect sources at a distance of 200 km around the centre of the antenna network.

Lightning discharges radiate over a broad range of radio frequencies, between VLH and VHF; for VHF, they cover from 30 MHz up to 300 MHz. The frequency in which the antennas operate will depend on the usage of VHF frequencies in the region, and the threshold at which they identify a possible source will be determined by the noise expected in the area.

As an example, a LMA network located in Ebro's Delta in Spain operates with frequencies between 60 to 66 MHz, as those belonged to the old TV system and now are unused.

Therefore, when in normal operation mode, a VHF antenna is able to detect every 80 μs a source which signal has a magnitude above their established threshold, they send the time when the signal was detected (time of arrival) to the central station. Then the central station calculates the time, altitude, latitude and longitude of the lightning with the already known distances between antennas

and the differences between the received times of arrival, following TOA technique. Other parameters as the power of the source or the number of stations that detected the lightning may be saved to in order to make a post-processing of the data. [6]

1.1.2. Colombia Mapping Array COL-LMA

The Lightning Research Group at Universitat Politècnica de Catalunya manages two LMA networks, one located at Ebro delta in the southwest of Catalonia and one located at Barrancabermeja in the north region of Colombia, called COL-LMA. In this project, all LMA data used will be from the COL-LMA network.

The Colombia Mapping Array, or COL-LMA started operating in April 2018. Initially was a network of six VHF antennas separated between 8 and 18 km around Santa Marta area, but in July 2019 the network was updated: two more stations were added and the location changed to Barrancabermeja. Both old and new networks are shown in Figure 1.

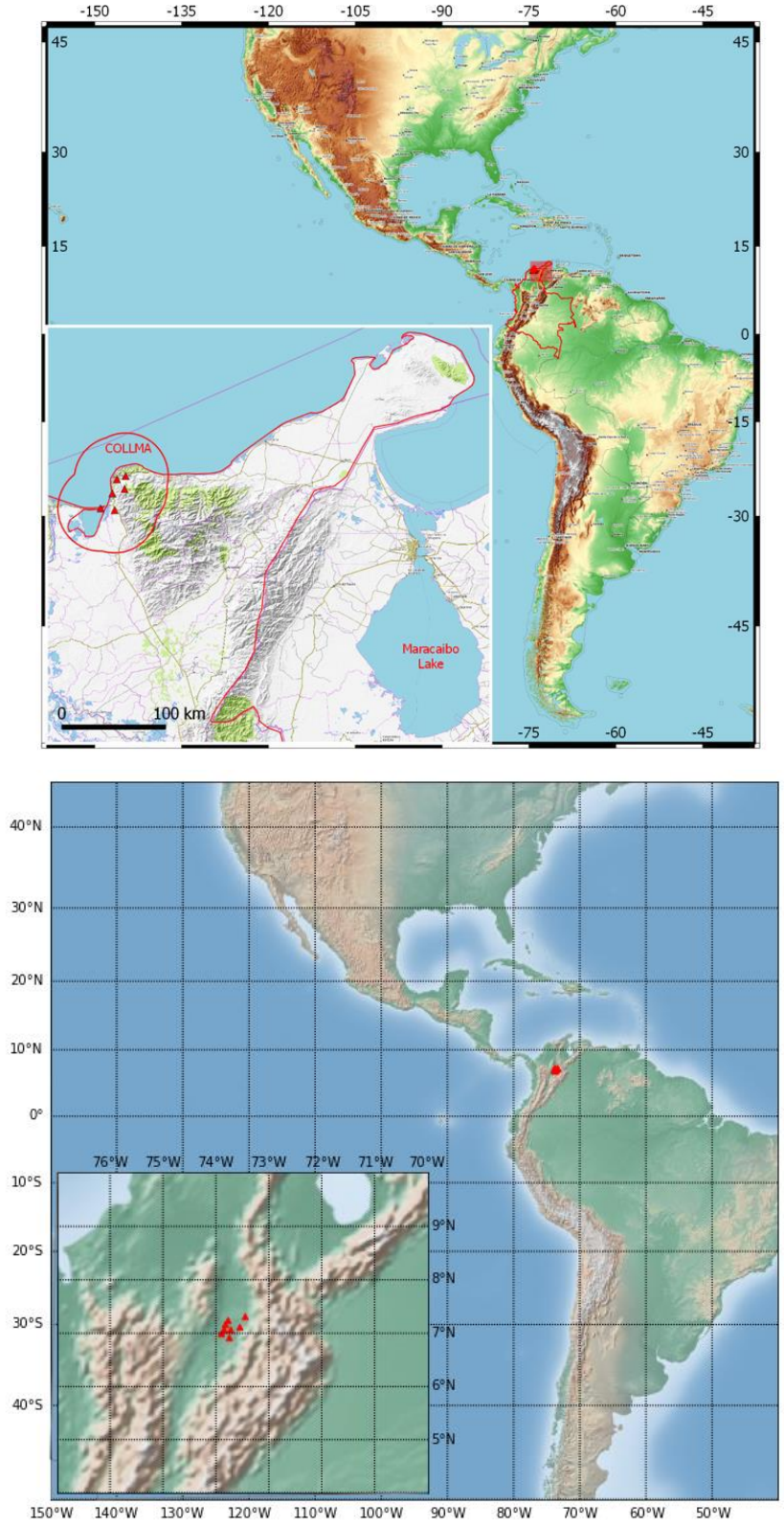

Figure 1. Geographical location of Colombia Lightning Mapping Array network COL-LMA in Colombia. Upper image is for old location in Santa Marta [7], lower image is for the new location in Barrancabermeja.

1.2. GOES-R Geostationary Lightning Mapper

The Geostationary Lightning Mapper, or GLM, is an optical instrument currently on board the GOES-R satellites GOES-16 and GOES-17. GLM maps the total lightning activity during day and night over the Americas and their oceanic regions on the western hemisphere.

This system was developed by Lockheed Martin for NASA Goddard Space Flight Center and is currently operated by the National Oceanic and Atmospheric Administration (NOAA).

The data sets created for GLM data were derived from the NASA Lightning Imaging Sensor and the Optical Transient Detector [8], were both instruments use 0.5º x 0.5º latitude-longitude grids to identify and locate the detected flashes.

GLM data is available within seconds of occurrence, therefore it is used to monitor severe weather as formation of tornadoes among other weather events. Other important uses of GLM data are to aid in air traffic management and support of sea transportation logistics. [9]

1.2.1. GOES-R Satellites

The Geostationary Operational Environmental Satellites—R Series or GOES-R Series is a collaborative program between NASA and NOAA.

The program consists in a fleet of four geostationary weather satellites: GOES-R (now in orbit as GOES-16), GOES-S (now in orbit as GOES-17), GOES-T and GOES-U.

A two-satellite operational system is implemented, where the locations of the satellites are 75.2º W and 137.2º W. A third on-orbit spare satellite is maintained at 105º W in order to assist in case of an anomaly or failure of the main satellites. Figure 2 shows an outline of the location and coverage of the satellites.

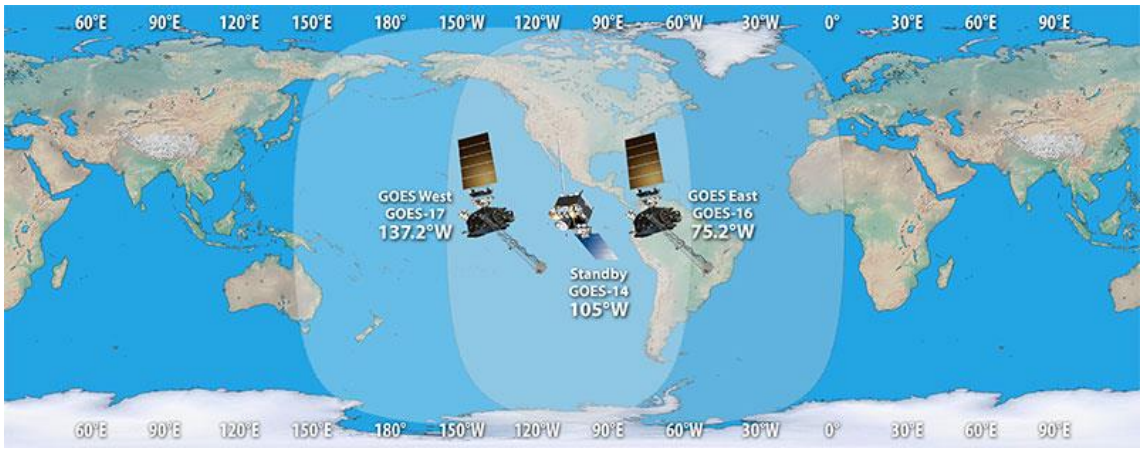

Figure 2. Outline of the GOES-16 and GOES-17 position and region coverage [10].

On board the satellite there is a six-instrument suite consisting on:

- Advanced Baseline Imager, ABI: the primary instrument for imaging Earth's weather, oceans and environment.

- Extreme Ultraviolet and X-ray Irradiance Sensors: detect and monitor solar irradiance in the upper atmosphere.

- Geostationary Lightning Mapper: the first operational lightning mapper in geostationary orbit.

- Magnetometer: provides measurements of the space environment magnetic field.

- Space Environment In-Situ Suite: an array of sensors that monitor proton, electron, and heavy ion fluxes in the magnetosphere.

- Solar Ultraviolet Imager: a telescope that observes and characterizes complex active regions of the sun.

All this instrument technology will allow the detection of environmental phenomena and provide more timely and accurate forecasts and warnings.

1.2.2. System characteristics and operation

The GLM measures radiances at cloud top from all types of lightning (in-cloud and cloud-to-ground) during day and night.

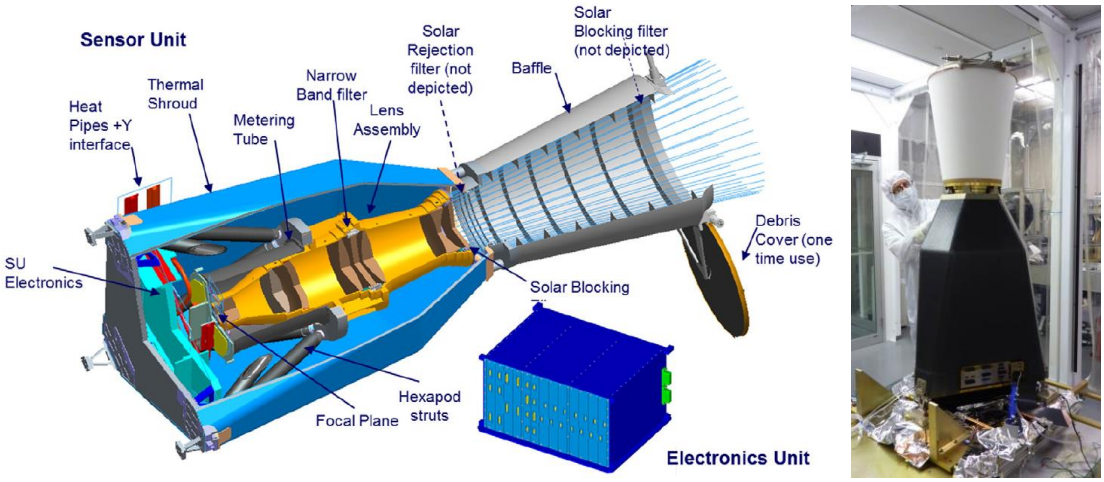

Figure 3. Right, sketch of the sensor unit and electronics unit of GLM [11]. Left, GLM being prepared to undergo vibration testing [12].

To do so, GLM is a high-speed event detector that operates in the near infrared. Its data handling and processing is what makes it different from other imagers as the characteristics of lightning and the difficulty to identify them during day time need different requirements.

Table 1. GLM performance characteristics [11].

CCD imager	1372x1300 pixels
FOV (across)	Full disk
Pixel FOV (nadir)	8 km
Pixel FOV (corner)	14 km
Wavelength	777.4 nm
Frame rate	2 ms
Downlink data rate	7.7 mbps
Product latency	< 20 s
Total mass	125 kg
Average operational power	405 W
Volume (height, width, depth)	149 cm, 63.5 cm, 65.8 cm

GLM is formed by a wide field of view lens combined with a narrow band interference filter that is focused on a high-speed Charge Coupled Device (CCD) focal plane. It reads the signals from the focal plane and processes them in real time for event detection and data compression. The events detected are then formatted and sent to the satellite's local area network. In Table 1 are all the performance characteristics of GLM.

Since GLM is an operational instrument, the latency must be minimal. In 10 seconds, the instrument collects, filters, geo-locates and time tags the raw data into Level 1B events. Then, Level 2 filtering performs the clustering of lightning

event data into groups and flashes, this filter is called Lightning Cluster Filter Algorithm or LCFA and the resulting lightning data is used for scientific research.

1.2.3. Events, groups and flashes

An event is the basic unit of data from the GLM, it is defined as a pixel that exceeds the background threshold during a single frame. It has to be noted it is possible that multiple pulses happening in the same integration window contribute to the same event.

The Level 1b data contains time, x and y pixel location, latitude and longitude locations, and calibrated amplitude of an event. With this information, the Lightning Cluster Filter Algorithm or LCFA determines the groups and flashes related to each event.

LCFA also determines if an event could be product of a non-lightning source, if that is the case the data is still clustered with the lightning data after being consequently marked.

Figure 3 shows how LCFA clusters events into groups and flashes. A group will consist on events adjacent to each other. Flashes will be groups separated in time by 330 ms or less and in space by 16.5 km (two pixels aprox.).

A flash could be formed by one single group or many of them, and at the same time groups can contain only one or many events.

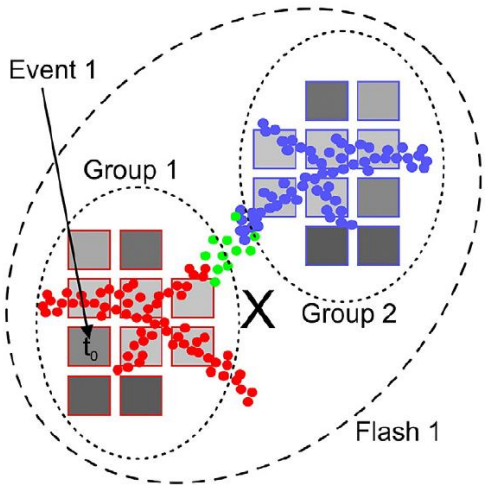

Figure 4. Illustration of a GLM flash formed by 2 groups and 20 events. The dots represent LMA sources [11].

1.3. Meteosat MTG Lightning Imager

The Ligtning Imager or LI, is an optical instrument that will be on board the Meteosat Third Generation satellite. LI will map the total lightning activity during day and night over more than the 80% of the visible Earth disc. [13]

This instrument will complement the Geostationary Lightning Mapper onboard GOES-R and the China Meteorological Administration Lightning Mapper onboard the FY-4.

The information provided by the Lightning Imager will be used for nowcasting, aid to air traffic safety and to understand the physical and chemical processes in the atmosphere.

1.3.1. MTG Satellites

Meteosat Third Generation or MTG is a cooperative program between EUMETSAT and ESA to produce the next generation of geostationary satellites. This generation will improve their imaging service as it will include a new Lightning Imager and a state-of-the art atmospheric sounding service.

The program consists in six geostationary satellites that will launch from 2022 onwards, four of them will be Imaging Satellites (MTG-I) and two will be Sounding Satellites (MTG-S).

Their operational configuration implemented will consist of one MTG-S satellite and two MTG-I satellites operating in tandem, one will scan Europe and Africa every 10 minutes and the other one will scan Europe every 2.5 minutes.

On board the imaging satellites, MTG-I, will be found the Flexible Combined Imager and the Lightning Imager. On the sounding satellites, MTG-S, an interferometer, the Infrared Sounder and the Copernicus Sentinel-4 will be found.

1.3.2. System characteristics and operation

The Lightning Imager will detect energy received from photons. Same as GLM, If the energy detected exceeds the trigger threshold established, it is identified as LI triggered Event.

The instrument is formed by four identical cameras, each one covers a different domain of the field of view, as shown in Figure 4.

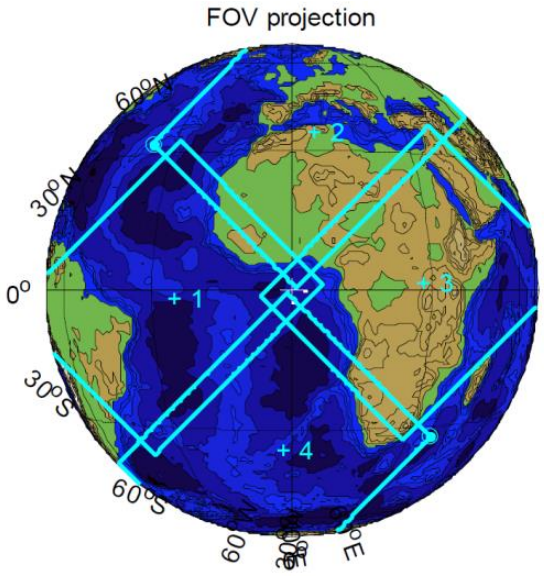

Figure 5. Field of View of the Lightning Imager instrument. [14]

Again, a GLM, the data produced by this instrument is given as a set of events, groups and flashes. Events will be the triggered pixels in the detector grid, groups will be those events in the same integration period that are next to each other and finally flashes will be collections of groups in temporal and spatial vicinity.

2. Methodology

In this chapter will be explained the data received from the different lightning detection systems discussed before: Lightning Mapping Array and Geostationary Lightning Mapper.

Both systems provide a raw file that is processed before using it in the comparing program made in this project. Both files, raw and processed will be explained and compared to understand why the processing was necessary.

The programs created to analyse and compare both systems will also be explained in this chapter, in order to understand the creation process and the capabilities that the program has.

2.1. Data files

2.1.1. Lightning Mapping Array Data

As explained in section 1.1.1., each LMA VHF antennas give information of the distance and timing of the sources detected and afterwards the information of each antenna is processed to determine the 3D location of the source.

In COL-LMA station, each antenna dumps this information into a "*.dat*" file every 10 min. With this information, an algorithm made by UPC's Lightning Research Group crosses all the different antenna data and returns another "*.dat*" file containing the position, timing, power among other values for each source detected for a 10 min period.

With this algorithm it is possible to change the number of antennas to be used for the process. This is important as, a very high number of antennas give less noisy results but it can cause a loss of information, as a lightning event may be detected by less antennas than required. On the other hand, if a low number of antennas is selected, the noise will be higher and more false positive events would be considered as valid.

Before using the data, another algorithm also made by the Lightning Research Group finds and identifies the flashes and the noise in the data. This algorithm returns a "*.txt*" file containing an identifier for noise (0) and lightning flash ID (≠0), time, latitude, longitude, height and power. With this program is also possible to display the events, and has been the reference for the display program made in this project.

2.1.2. Geostationary Lightning Mapper Data

NOAA shares the GLM data online on Google Cloud Platform, there can be found the data organized in folders per year, day of the year and hour. Then, files with data for every 200 milliseconds are available to download.

This GLM data files have netCDF format "*.nc*". NetCDF is a data model capable to store large amounts of array-oriented data in a way that is self-describing and portable. This type of dataset contains dimensions, variables, and attributes, and all have both a name and an ID number in order to be identified. [15]

Therefore, the GLM files contain a lot of information that may not be of interest. Figure 5 shows the structure of a GLM netCDF file, it contains different variables, and their descriptions. There is information related to event, flash and group levels but also about the satellite status at that moment (not shown in picture).

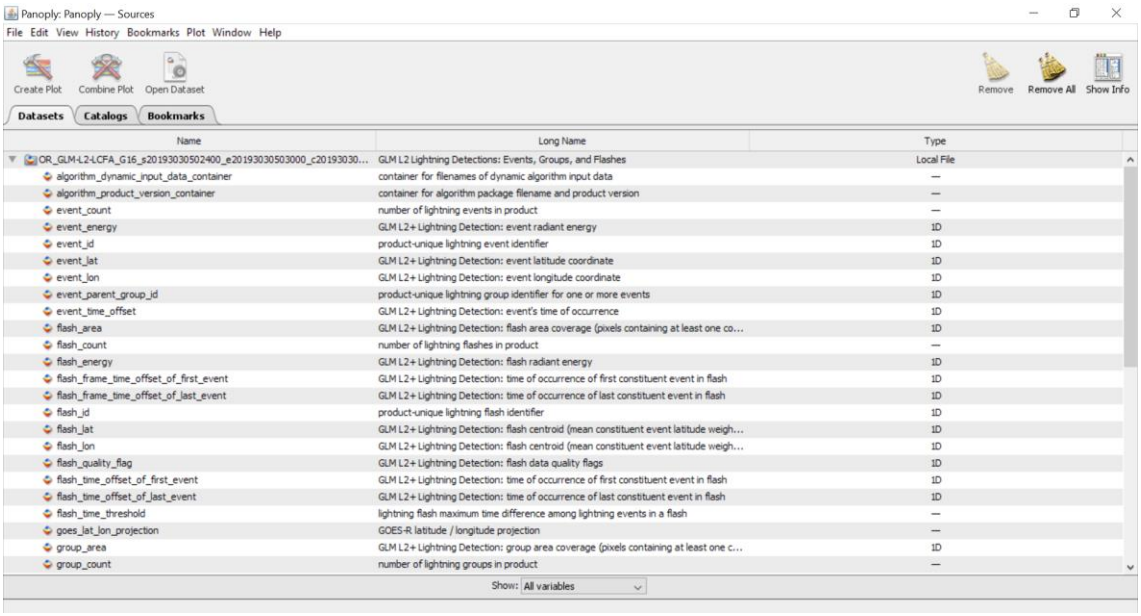

Figure 6. NetCDF data file structure for a GLM file.

In order to extract only the information required for location and identification of lightning, a python code has been developed by UPC's Lightning Research Group. This program takes all the GLM data files saved in a specified folder and returns a "*.txt*" file with time (in seconds), group latitude and longitude, flash ID, flash latitude and longitude, and energy for each event.

A second processing is needed for this information, as the first results of the data processed showed an error where more than one flash could have the same ID.

In order to avoid this bug, a python algorithm was created to find the problematic IDs, separate the flashes and assign them new unused IDs.

2.2. Data Analysis

2.2.1. Pre-processing: LMA and GLM data normalization

In order to compare the data between LMA and GLM, the data of LMA will be taken as reference for time and space delimitation, as GLM covers more space range and if it's not delimitated some results could be misleading.

To delimitate the GLM data in time it will be used the time of the first and last LMA sources. To delimitate the area of evaluation, a fixed value of 150 km from the LMA network centre will be set, which is the double the value of the estimated LMA range used.

2.2.2. Comparison parameters

Once GLM data has been delimited it will be evaluated against LMA data. In order to do so, a series of parameters have been defined to evaluate and compare the data correlation between both lightning location systems. This parameters have been determined using as reference a *ISS-LIS data analysis based on LMA networks in Europe* [16].

The comparison parameters to evaluate are:

- **Flash Detection Efficiency**:

A correlation between GLM flashes and LMA flashes based on the distance between their centroids and the time of the events/sources.
Once GLM flashes have been assigned to a LMA flash (or not), the Detection Efficiency for GLM (DE_f) is calculated following the formula

$$DE_f = \frac{Number\ of\ LMA\ flashes\ detected\ by\ GLM}{Total\ number\ of\ LMA\ flashes\ evaluated}$$

Using the flash correlations found with these criteria various items are calculated:

- Qualitative maps
- Flash False Alarm
- Number of GLM events and flashes per LMA flash
- Distribution of number, power and height of LMA sources for flashes detected and not detected by GLM

- Flash detection efficiency vs maximum height of LMA flashes

- **Flash duration**:

An average of the duration of LMA flashes, GLM flashes and of the correlated flashes is made to compare their differences.

- **Location accuracy**:

The Detection efficiency flashes are compared using the distance between their centroids, but this could lead to mistakes in the assignation of flashes. To determine the accuracy in the location of correlated flashes, LMA data should be rearranged to fit a 0.075º square grid that would make easier the comparation with the grided data of GLM.

With LMA data rearranged in a grid, a series of qualitative maps to visually compare the LMA and GLM are obtained.

3. Results

Following the methodology presented in Chapter 2, in this chapter are explained the results obtained after evaluating the LMA and GLM flash data for four different cases.

3.1. Cases evaluated

The LMA and GLM data of four cases are evaluated to determine the correlation existing between the two lightning detection systems.

Table 2. Data of cases evaluated.

Case	Date	Time interval (Local time)	Sources detected	Events detected
A	30 Oct. 2019	00:30 - 05:30	697955	318134
B	31 Oct. 2019	20:30 - 23:00	2535	4806
C	01 Nov. 2019	23:40 - 05:40	307376	189356
D	01 Nov. 2019	05:40 - 08:30	104306	42971

Case A is one of the longest in terms of time and is the one with more sources and events detected. This case and case B happen during the night, when it should be easier for GLM to detect events, since the amount of light interfering with the detection of lightning is lower.

Cases C and D belong to the same meteorological event, as can be observed in Figure 7. This event is divided in two parts to separate the data detected before and after sunrise, which was at 05:43 local time.

It will be interesting to evaluate if there are any differences in the results between C and D, and D, A and B, as the sunlight during those cases will be different and may be affecting the GLM detection.

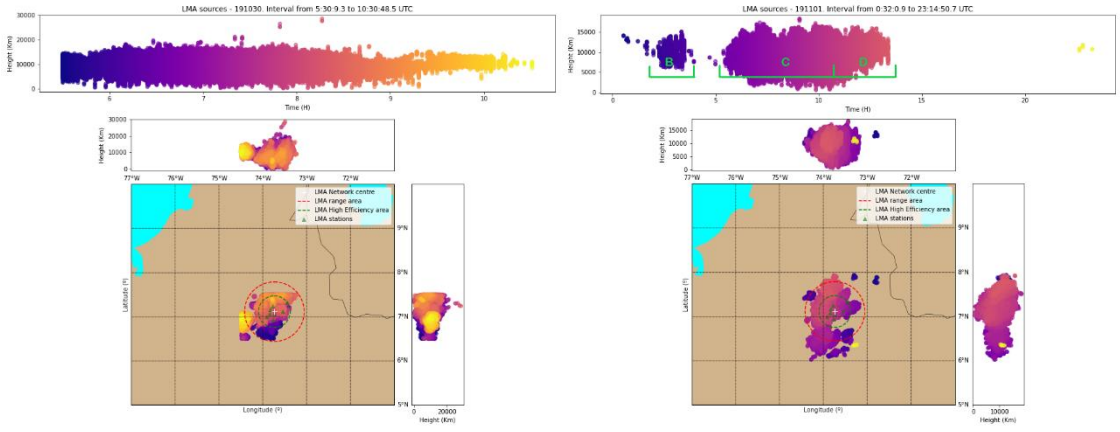

Figure 7. Preview of LMA data to evaluate in scenario A (left) and B, C and D (right)

3.2. Flash Detection Efficiency with LMA as reference

As explained in section 2.2.2., in order to calculate the flash detection efficiency, the GLM flashes are assigned to a LMA flash. In order to do so, various parameters have to be stablished to determine if two flashes are corelated.

Five parameters are stablished to discard LMA sources out of range, to define the margins allowed to assign a GLM flash to a LMA flash and to assess the quality of a LMA flash and its detection by GLM.

LMA detects sources within a range from the centre of the station, and this range may change depending on the number of stations available or the orography of the surroundings. A good process to set the range would be to create a grid to evaluate the power detection of the LMA data. The cells with lowest power values would indicate the most sensitive areas, which would be the ones were the detection is more precise. Then, with the most sensitive areas detected a good range value can be determined.

For this project, an estimated value of 75 km from the centre of the LMA network has been defined as the range of the LMA network. This value ensures that the flashes within this range have been detected in a good efficiency area.

To determine if a GLM flash is correlated to an LMA flash it has to be checked if both happen at the same time and at the same location. To compare the location, an initial value for the maximum distance allowed between centroids of both flashes is set to 50 km. In order to allow some margins to the LMA flash time window, a time tolerance is applied to the start and end of the flash. The initial value of this parameter is set to 0.1 seconds.

For LMA flashes, it would be interesting to evaluate only flashes with a minimum number of sources per flash, as they would represent larger and better detected flashes. To do so, an initial value for the minimum sources required to evaluate a LMA flash is set to 50. Then, from those LMA flashes that are evaluated, only those who are related to at least one GLM event will be considered detected.

A summary of the parameters defined is shown in Table 3.

Table 3. Flash detection efficiency parameters for data evaluation

Area range for LMA	75 km
Max. distance between centroids of GLM and LMA flashes	50 km
Time tolerance	1 s
Minimum number of LMA sources to accept a flash	50
Minimum number of GLM events to detect a flash	1

3.2.1. Detection Efficiency Parameter Results

With the parameters established in Table 3, the following results are obtained.

Table 4. Detection efficiency for initial parameters

Case	Average flash rate (min^{-1})	Average LMA source rate (s^{-1})	Number of flashes	Number of flashes detected by GLM	Detection efficiency
A	3.85	18.45	2231	2117	0.95
B	0.1	0.11	20	20	1
C	2.14	6.88	1379	1229	0.89
D	0.41	2.1	325	292	0.9
Total number of LMA flashes				3955	
Total number of LMA flashes detected				3658	
Total Detection efficiency				0.92	

For this initial result, the detection efficiencies obtained are really high. In case B, all flashes evaluated are corelated with a GLM flash, probably due to the low number of flashes and the low intensity of the event, as the flash and source rates are very low too.

As it was expected, cases C and D, the ones with possible sunlight affecting the GLM detection, have by little the lowest values of detection efficiency.

Calculating the detection efficiency for different values of range, allowed distance between centroids and minimum number of sources in flash, the following results are obtained.

Table 5. Detection efficiency for Range increase to 150 km

Range 150km	Average flash rate (min^{-1})	Average LMA source rate (s^{-1})	Number of flashes	Number of flashes detected by GLM	Detection efficiency
A	3.69	17.6	2242	2128	0.95
B	0.1	0.12	22	22	1
C	2.17	6.92	1395	1245	0.89
D	0.41	2.10	325	292	0.9
Total number of LMA flashes				3984	
Total number of LMA flashes detected				3687	
Total Detection efficiency				0.93	

For the first change the range is increased, therefore more LMA sources can be evaluated and more LMA flashes should be detected. This increase in LMA flashes is obtained, but it's so low that it does not affect the detection efficiency.

The range has been increased to double the estimated range of the LMA in order to evaluate the change in the detection efficiency but, as explained before, determining a good value for the range is more complicated.

Table 6. Detection efficiency for elimination of minimum sources required to evaluate LMA flash

No min sources	Average flash rate (min^{-1})	Average LMA source rate (s^{-1})	Number of flashes	Number of flashes detected by GLM	Detection efficiency
A	9.26	18.67	5731	4372	0.76
B	0.35	0.17	74	71	0.96
C	5.74	7.86	3701	2885	0.78
D	0.79	2.16	631	505	0.8
Total number of LMA flashes				10137	
Total number of LMA flashes detected				7833	
Total Detection efficiency				0.77	

For the second parameter to change, the requirement of 50 sources to evaluate a LMA flash is removed. This means that all flashes detected, even if they are formed by only one source are evaluated.

The number of flashes with this change has increased, in more than two times the initial results in all cases. The flash rate has also increased greatly, while the source rate has done it in a more contained way.

But the other side, the detection efficiency has decreased. The number of detected flashes has not increased as much as the number of flashes evaluated, and that could be an indicator that flashes with less than 50 sources could be not detected by GLM or are VHF emissions detected and not related with a flash.

In this scenario, the lowest detection efficiency is in A, one of the night cases.

Table 7. Detection efficiency for distance between centroids decrease to 25 km

Centroid dist. 25 km	Average flash rate (min^{-1})	Average LMA source rate (s^{-1})	Number of flashes	Number of flashes detected by GLM	Detection efficiency
A	3.85	18.45	2231	2077	0.93
B	0.1	0.11	20	20	1
C	2.14	6.88	1379	1223	0.89
D	0.41	2.1	325	290	0.89
Total number of LMA flashes				3955	
Total number of LMA flashes detected				3610	
Total Detection efficiency				0.91	

Table 7 shows the values when the distance between the centroids is decreased to 25 km. This is a more restrictive parameter, as the number of correlations decreases, but again as in the range change, the decrease in flashes detected is so low that it does not impact the detection efficiency.

In this case, the distance is decreased to 25 km because a 93% of the correlated flashes in the first evaluation had a distance lower than 25 km. In section 3.2.4 the different problems related to determine a correct distance between centroids is further elaborated.

Table 8. Detection efficiency for Range increase to 150 km, elimination of minimum sources required to evaluate LMA flash and distance between centroids decrease to 25 km

All changes applied	Average flash rate (min^{-1})	Average LMA source rate (s^{-1})	Number of flashes	Number of flashes detected by GLM	Detection efficiency
A	9.45	18.4	5974	4441	0.74
B	0.4	0.18	96	93	0.97
C	6.09	7.95	3922	3061	0.78
D	0.79	2.17	638	485	0.76
Total number of LMA flashes				10630	
Total number of LMA flashes detected				8080	
Total Detection efficiency				0.76	

Finally, when all the parameters are changed at the same time, changes in all the results are obtained. The number of flashes evaluated has increased, along with the flash and source rates, probably due to the removal of the minimum sources restriction.

The number of detected flashes is at its maximum for B and C cases, but on A and D it was when only the minimum number of sources was removed (Table 6). This could mean that they lost a number of detections due to the more restrictive distance between centroids, because the LMA range change could only help as it increases the number of detections.

In all cases the detection efficiency has dropped, for cases A and D it's the lowest value obtained. B and C cases both had their lowest detection efficiency when the minimum number of sources was changed. On the other hand, all cases had their best detection efficiency value with the first set of parameters.

Therefore, after evaluating different changes in the parameters and seeing that the initial values give the best results for detection efficiency, the next sections will be evaluated with the initial parameters in Table 3.

3.2.2. Qualitative maps

Two types of qualitative maps are obtained for each of the cases, one shows the location of the LMA sources and GLM events and the other their time-height values. Since GLM data not provides height, it is set to 1 for all events.

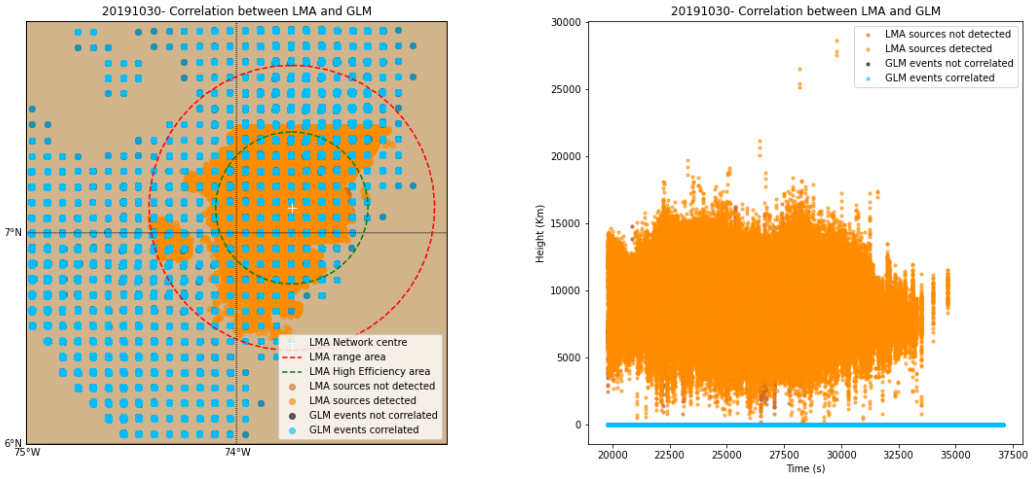

Figure 8. Qualitative maps for LMA sources and GLM events for case A. Right: Location. Left: time-height relation (with value 1 for GLM).

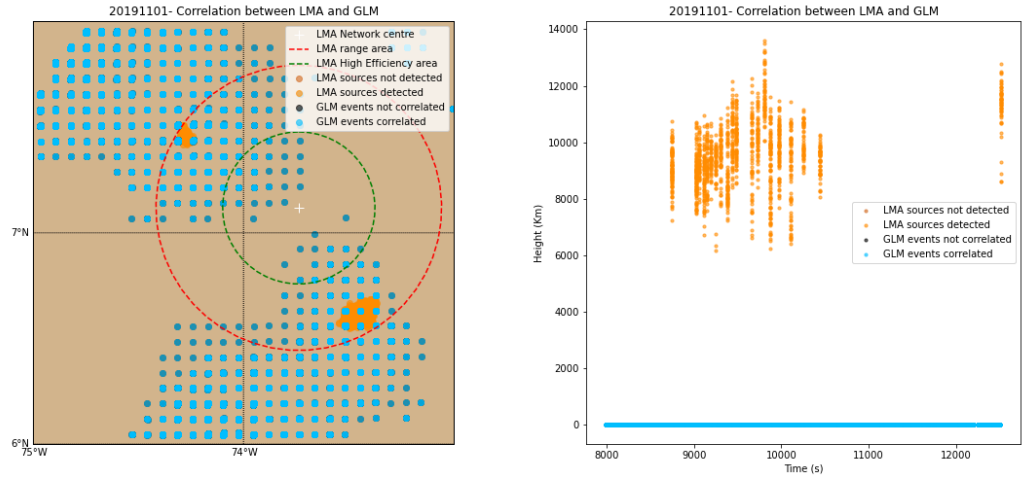

Figure 9. Qualitative maps for LMA sources and GLM events for case B. Right: Location. Left: time-height relation (with value 1 for GLM).

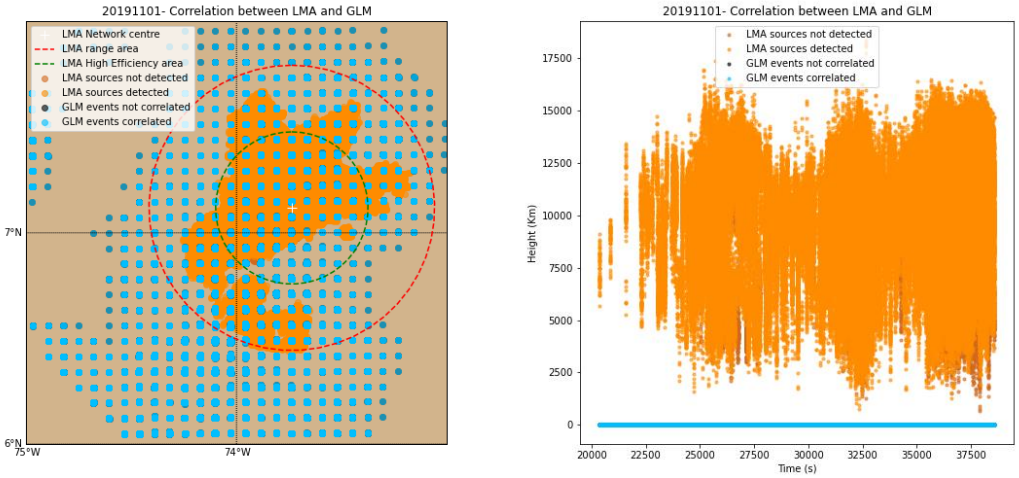

Figure 10. Qualitative maps for LMA sources and GLM events for case C. Right: Location. Left: time-height relation (with value 1 for GLM).

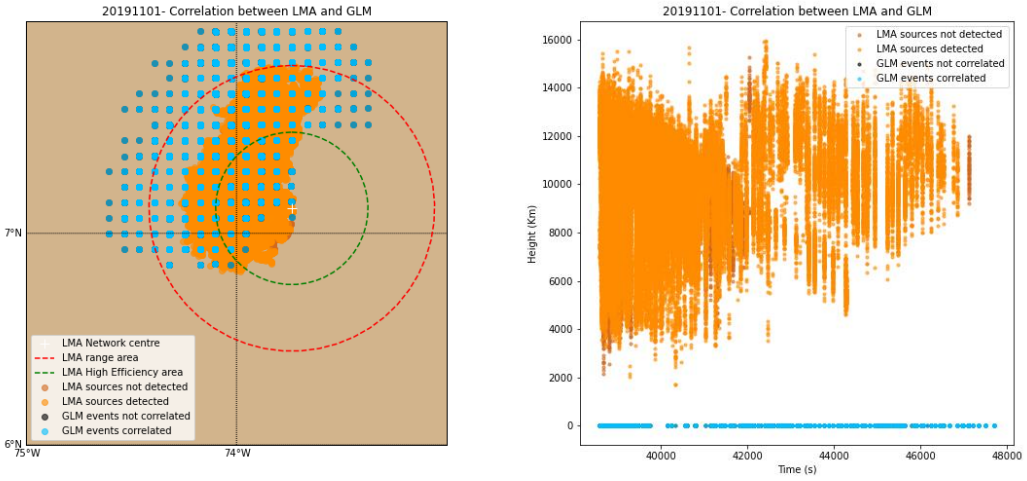

Figure 11. Qualitative maps for LMA sources and GLM events for case D. Right: Location. Left: time-height relation (with value 1 for GLM).

In these figures are presented both correlated and not correlated LMA sources and GLM events.

For LMA sources, it is quite difficult to see the not correlated sources in the map. On the time-height plots, especially in case C and D (Figures 10 and 11) little glimpses of the not detected sources can be appreciated every now and then.

For GLM events, since their data is evaluated over a grid, the map plot for all cases shows various events superposed. Since not correlated events are under the correlated ones, not very much information can be extracted from these

graphs. Case B map is the one that shows a notable change of colour in the events, indicating they aren't correlated. This can be seen on the GLM events closer to the centre of the LMA network, in an area with no LMA sources in this case.

In the next section, GLM events not assigned to a LMA flash will be studied in more detail.

3.2.3. Flash False Alarms

In all cases studied, there are some GLM events and GLM flashes that are not correlated to any LMA flash. This could happen because some LMA flashes are not evaluated (due to the minimum source number restriction or a heavy filtering during noise reduction) or because the LMA does not detect a flash (if it has a low frequency emission it would not be detected, if not enough stations detect the flash…).

There is also the possibility that GLM has made a false detection, a flash false alarm. To consider that a not correlated flash is a false alarm, the candidate should be inside the range of the LMA network, should happen during a thunderstorm and should be compared with raw LMA data or data from other instruments that can confirm that a flash was indeed happening at that moment.

In Figure 12 can be found the maps showing the location of not correlated events. These maps show almost the same data as the qualitative maps on section 3.2.1. Looking at the values for the total number and the number of not correlated events and flashes it seems that most events were not correlated and therefore it makes sense that the maps look so similar.

Table 9. GLM events and flashes not correlated to a LMA flash.

Case	GLM events	Not correlated	(%)	GLM flashes	Flashes not correlated	(%)
A	790521	600706	76	10636	8667	81
B	161281	159720	99	2090	2070	99
C	466760	346939	74	3913	2765	70
D	47657	9711	20	404	125	31
Total number of GLM flashes					17043	
Total number of not correlated flashes					13627	
Percentage					80	

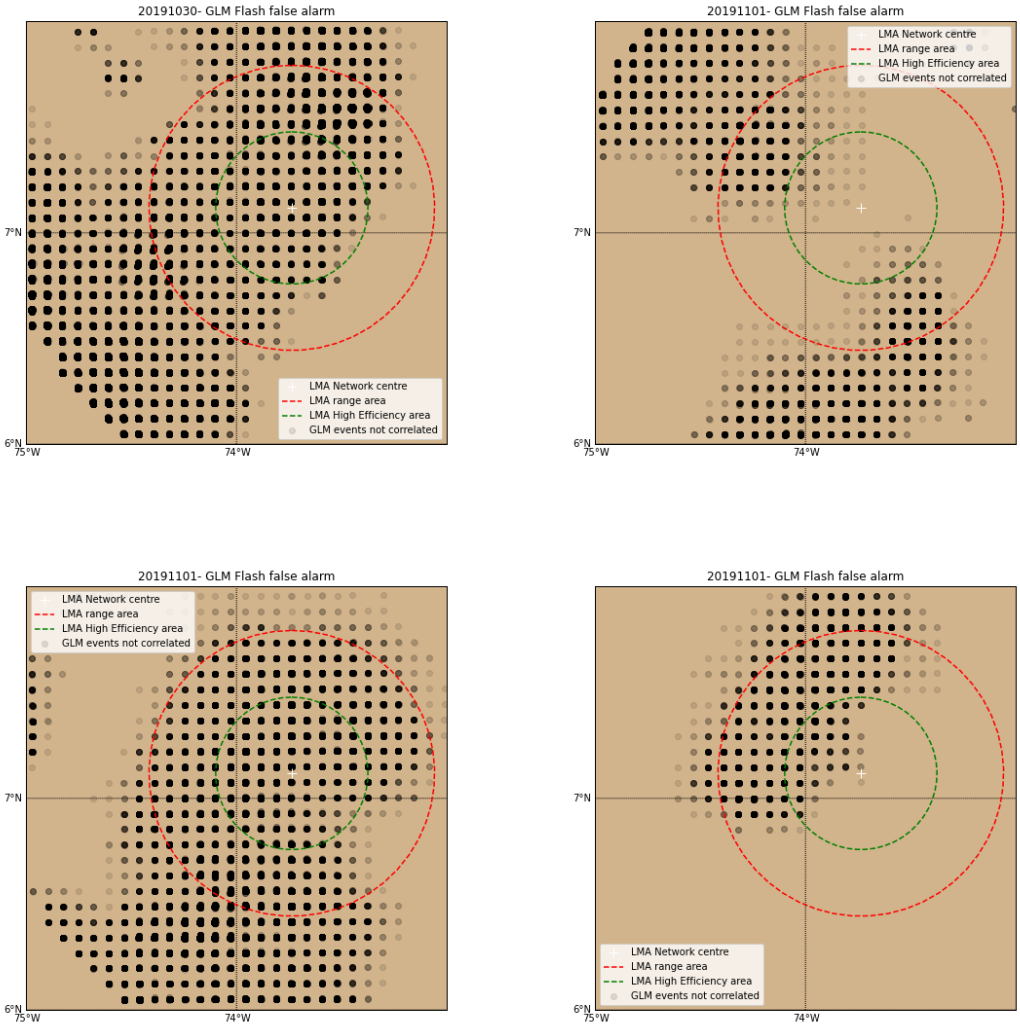

Figure 12. Qualitative maps for GLM events not correlated to a LMA flash. From left up corner, in order: Case A, case B, case C and case D.

Even when the not correlated flashes have not been compared to raw LMA data to evaluate if they truly are flash false alarms, the data obtained has been useful to find out interesting aspects of the GLM flashes.

In the first place, when evaluating the found and not found flashes, a series of flash duplicates were found. All these GLM flashes were assigned to more than one (or two) LMA flashes. It could be a question of distance, that various LMA flashes were close enough to be assigned to their correlated flash and one of another LMA flash, or that two LMA flashes “formed” a GLM flash. Various examples of this can be found in section 3.2.4.

It was also interesting to see that in total, around an 80% of the flashes were not assigned to a LMA flash. This is possibly the result of evaluating a large area of GLM data, were most of it has no LMA flashes detected.

3.2.4. Distance between correlated flashes centroids

In order to determine the most appropriated value for the maximum distance that could be allowed to consider a LMA flash and GLM flash correlated, an evaluation of the distances found when the maximum distance is set to 50 km has been performed.

In total, a 93% of the flashes that are correlated are separated by a distance below 25 Km, with a median of 2.94 Km. Therefore, it would be safe to assume that decreasing the allowed distance from 50 to 25 would not affect the number of detections, as only a 7% of the correlations would be lost. This is confirmed for the cases studied in Tables 4 and 7, section 3.2.1

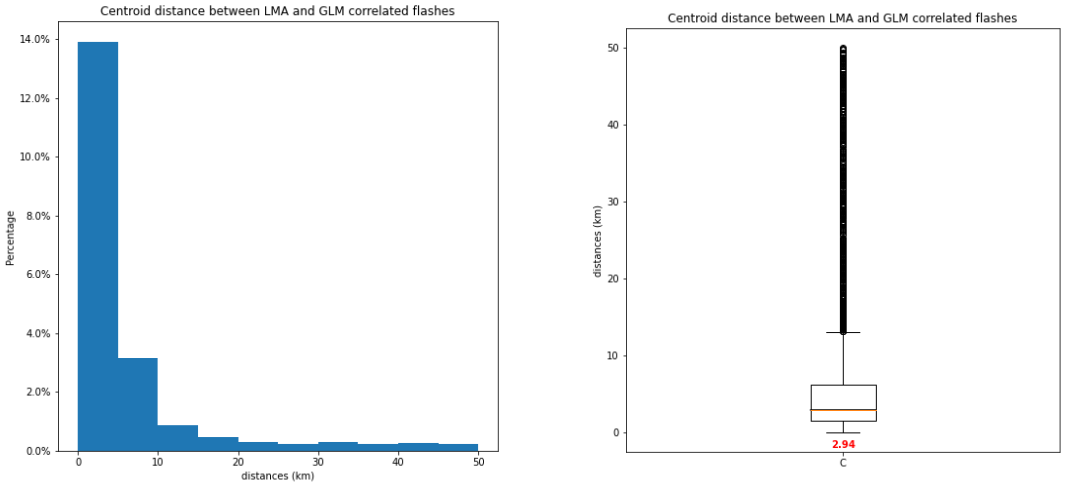

Figure 13. Histogram and boxplot for the distance values of all the flashes correlated.

But looking at those GLM flashes that are assigned to more than one LMA flash, it can be seen that there is not only one way for the LMA flashes to share a GLM flash. And the distance criteria could erase correlations that are in fact correct.

For instance, the most common way for two flashes to share a GLM flash is because they are close in time and space.

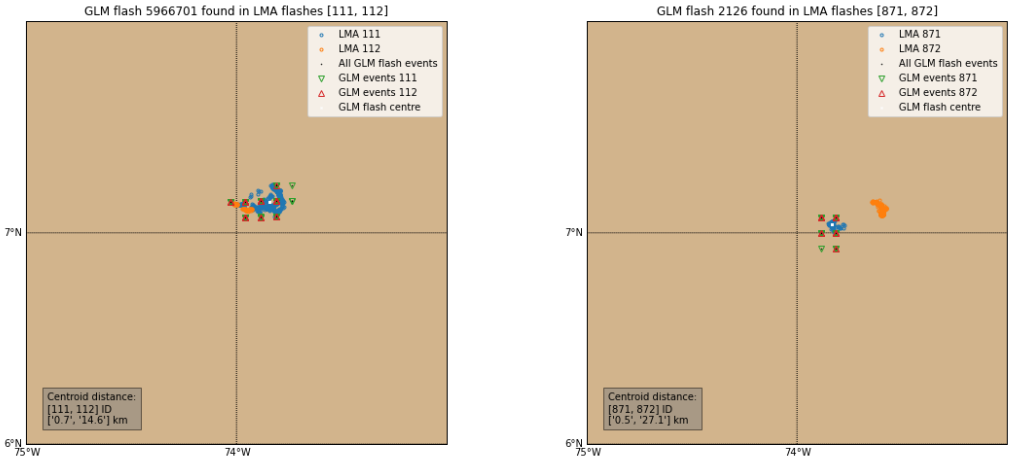

Figure 14. Two examples of LMA flashes sharing one GLM flash. The ones in the left are both separated from the GLM flash by less than 25 km, while one of the ones in the right is separated by 27 km.

For the example in Figure 14, reducing the distance would be beneficial as the flash in the right would be assigned only to the correct flash and will count only once on statistical studies.

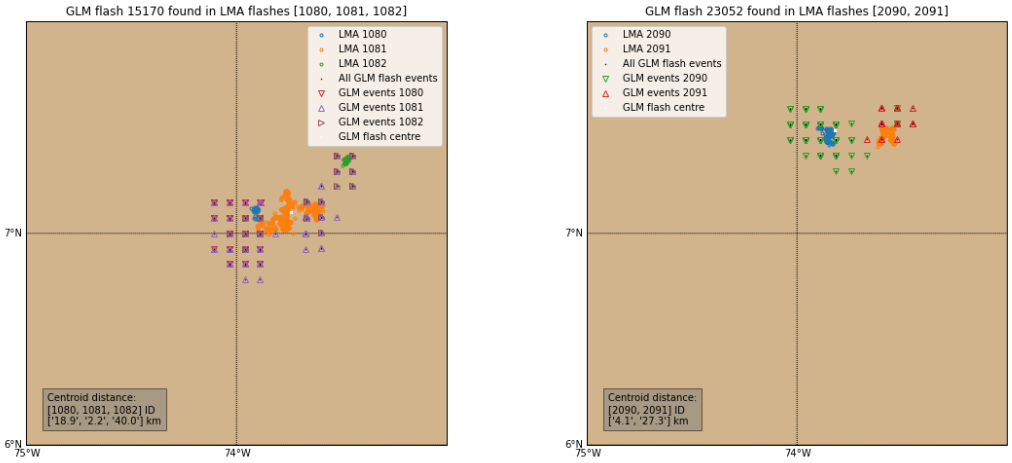

Figure 15. Two examples of LMA flashes sharing one GLM flash. In both examples there is one LMA flash separated from the GLM flash by more than 25 km, but it's a correct correlation.

But now, in the examples in Figure 15, the LMA flashes do belong to the GLM flash, and if the distance is reduced those correlations would be lost.

Another interesting aspect is that some GLM flashes are divided between various LMA flashes, not sharing all the events in both correlations but only part of them.

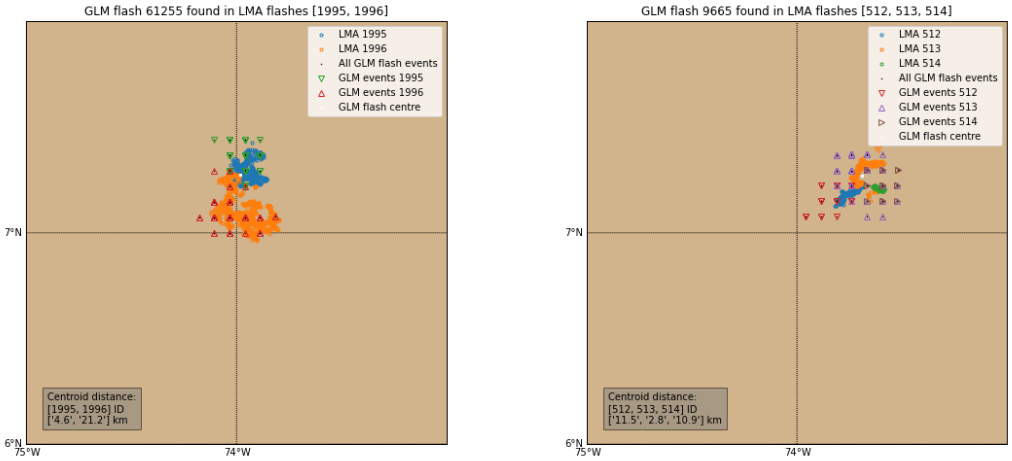

Figure 16. Two examples of LMA flashes partially sharing one GLM flash. The events relation with each LMA flash can be quickly seen as not all GLM events have the same markers at all times.

Since GLM not only merges events in flashes but also groups, it would be interesting to make an assignation of GLM groups to LMA flashes to see if those have a higher detection efficiency decreasing the distance between centroids.

3.2.5. Number of GLM events and flashes per LMA flash

Contrary to the results for ISS-LIS, for GLM it is not usual to have more than one GLM flash corelated with a LMA flash.

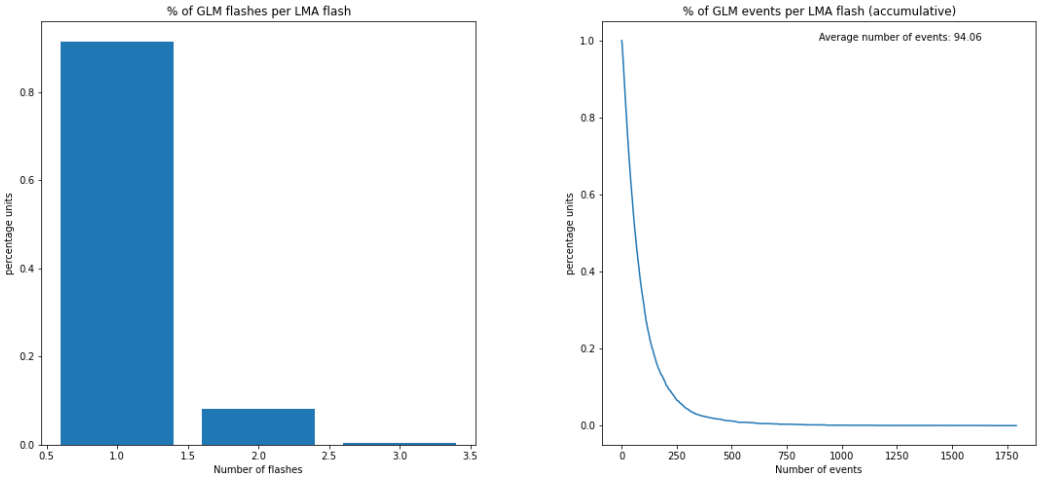

Figure 17. Right: distribution of the number of GLM flashes per LMA flash. Left: Accumulated number of GLM events per LMA flash.

This also makes sense with the found GLM flashes related to various LMA flash, as it indicates that it may be possible that one GLM flash is formed by various LMA flashes.

This change between ISS-LIS detection and GLM detection could be due to the difference between being in a low orbit, were the field of view is changing and possibly the identified flashes need to be shorter, and being in a geostationary orbit, that would not limit the time window for detection of flashes.

3.2.6. Distribution of number, power and height of LMA sources for flashes detected and not detected by GLM

In order to see it there are any significant differences between those LMA flashes detected and not detected, some statistics for the number of sources, power and height of the flashes has been computed.

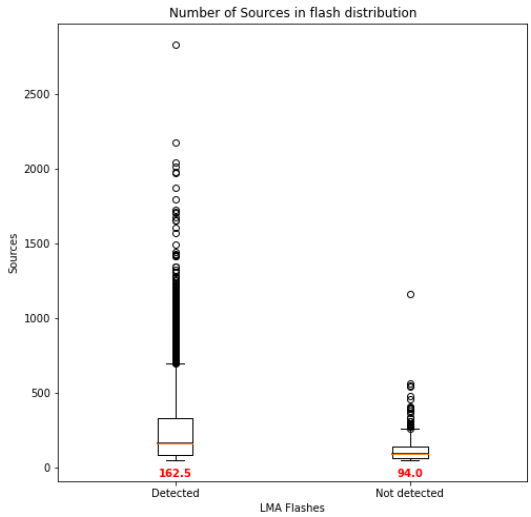

Figure 18. Number of LMA sources for flashes detected and not detected by GLM.

For the number of sources, it seems like those flashes that are detected have a higher number of sources. The median for detected is 162.5 and for not detected is 94.

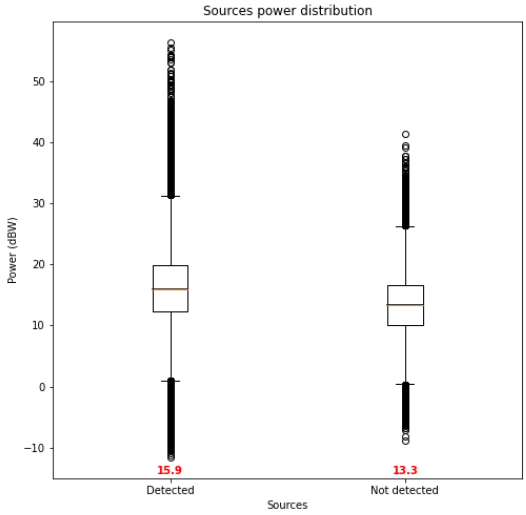

Figure 19. Power of LMA sources for flashes detected and not detected by GLM.

In terms of VHF power, the values obtained for detected and not detected flashes are quite similar. The median is 15.9 and 13.3 respectively.

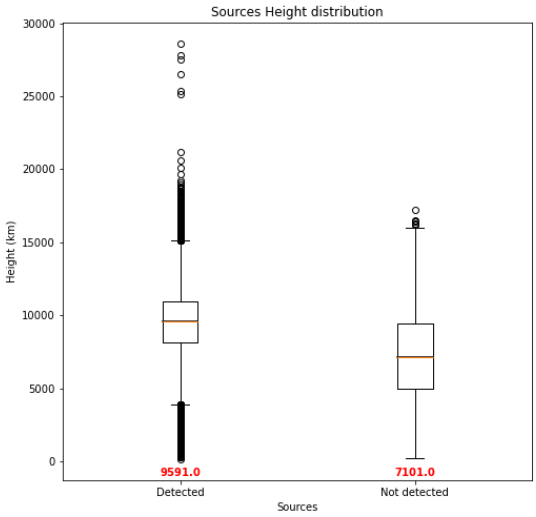

Figure 20. Height of LMA sources for flashes detected and not detected by GLM.

Height has the biggest difference between detected and not detected mean, being 9591 and 7101 respectively. This difference could indicate that deep clouds on top of a lightning could largely affect the detection of this flash from space.

In order to have another look to the possible effect of flash height in GLM flash detection, the detection efficiency is calculated for various height ranges in the next section.

3.2.7. Detection Efficiency vs maximum height of LMA flashes

In order to see the influence of flash height in flash detection, the Detection efficiency has been calculated based on their height, using the maximum height value as reference.

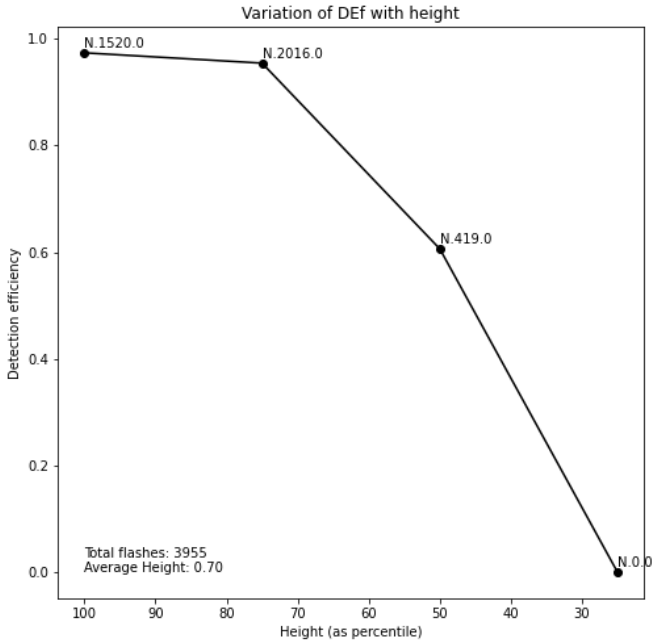

Figure 21. Influence of flash height to Detection efficiency.

The heights in the cases studied are quite high, as the average value is 0.7 times the maximum height, this results in a slow decrease of the detection efficiency for heights above 0.5. For heights under 0.25, there are no flashes to evaluate.

3.3. Flash Detection Efficiency with GLM as reference

After seeing the results of sections 3.2.3 and 3.2.5, it seems as GLM flashes could be assigned to more than one flash. For this reason, it seems interesting to perform the flash detection efficiency procedure again using GLM flashes as reference.

Therefore, in the next results, the assignation of LMA flashes to GLM flashes has been done with the same parameters as established before. GLM flashes will only be evaluated if they contain more than 50 events, distance between centroids of the correlated flashes will be less than 50 km, and with one source the GLM flash will be considered detected.

Note that all the GLM data used is inside the 75 km range of the LMA centre, as no region delimitation can be applied to LMA with GLM restrictions (as GLM covers more area than LMA), and that way the GLM is more restricted and the detection efficiency better represented.

Table 10. Detection efficiency for established parameters

Case	Average flash rate (min^{-1})	Average GLM event rate (s^{-1})	Number of flashes	Number of flashes detected by LMA	Detection efficiency
A	6.54	14.42	1954	1386	0.71
B	0.46	0.84	31	25	0.81
C	3.71	9.17	1126	1058	0.94
D	1.37	4.28	208	207	0.99
Total number of GLM flashes				3319	
Total number of GLM flashes detected				2676	
Total Detection efficiency				0.80	

Again, the best results of detection efficiency are obtained for the initial values (Table 3), and the next results are obtained using them.

Qualitative maps are omitted, as they do not provide any additional information than that commented in section 3.2.2.

For LMA flashes not correlated to GLM flashes, the following results are obtained.

Table 11. LMA sources and flashes not correlated to a GLM flash

Case	LMA sources	Not correlated	(%)	LMA flashes	Flashes not correlated	(%)
A	712435	196249	28	5974	3649	61
B	2535	1144	45	96	69	72
C	311581	93001	30	3922	2006	51
D	106242	38037	36	638	310	49
Total number of LMA flashes					10630	
Total number of not correlated flashes					6034	
Percentage					57	

As with LMA, an important part of the flashes are not correlated. In this case, more than the 50% of the flashes are not correlated, while events not reach the 50%. The quantity of LMA data outside the range made for GLM is less than the GLM data in the previous case, therefore less flashes are out of range and with no possibility for a range.

Repetition of LMA flashes on various GLM flashes is also found. Discarding the ones were the problem could be solved with a reduction of the maximum distance

between centroids, this time the interesting aspect of these repeated flashes is that the GLM flashes that share a LMA flash are coincident in time.

During the assignation of GLM flashes to LMA, the problem was that two LMA flashes close in time had assigned the same GLM flash. Therefore, those LMA flashes happened at different times and were different flashes.
But in this case, the GLM flashes that share a LMA flash also share the time interval, or part of it. It would be interesting to analyse the flashes with this characteristic and discard if it could be possible that the same flash is detected twice by GLM.

Table 12. GLM flashes repeated during LMA flash detection and LMA flashes repeated during GLM flash detection.

Case	GLM flashes repeated	LMA flashes repeated
A	327 (14.7%)	86 (1.4%)
B	0 (0 %)	0 (0 %)
C	138 (10%)	58 (1.5%)
D	44 (13.5%)	24 (5.5%)
Total	**509 (12.9%)**	**168 (1.6%)**

Table 12 shows the number of GLM flashes that were assigned to more than one LMA flash in section 3.2, and the LMA flashes that were assigned to more than one GLM flash in the current analysis.

As expected, the number of repeated flashes is lower when using GLM flashes as reference, as the issue seems to be more like a punctual identification problem.

Figure 22 also helps to validate the assumption made for the repetition of flashes in the LMA assignation, since more than 50% of GLM flashes are correlated to various LMA flashes.

It would be really interesting to calculate the detection efficiency for GLM groups, in order to compare the number of LMA flashes associated to a group, and see if groups are closer in size to LMA flashes.

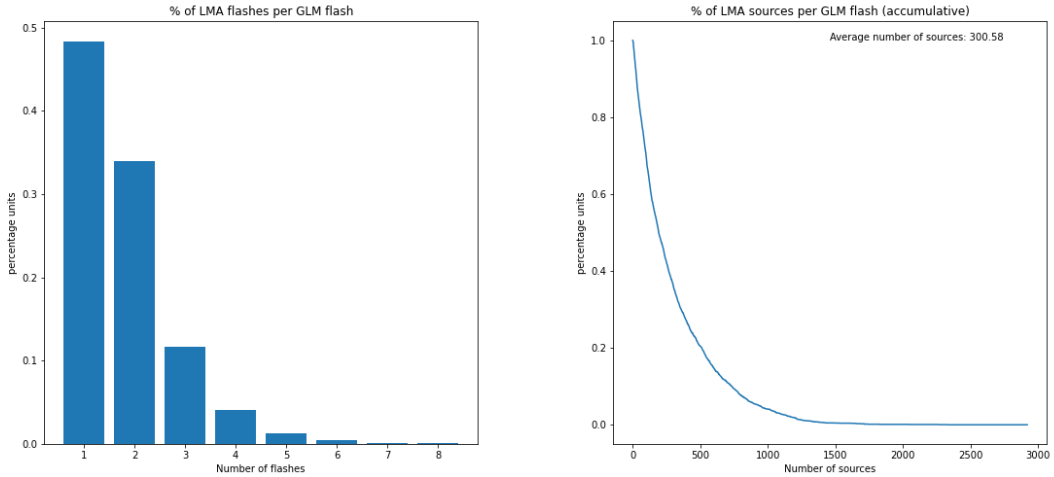

Figure 22. Right: distribution of the number of LMA flashes per GLM flash. Left: Accumulated number of LMA sources per GLM flash.

For comparison of parameters between GLM flashes detected and not detected, only the number of events and the radiance of the events is calculated, as GLM does not give information related to the height of the events.

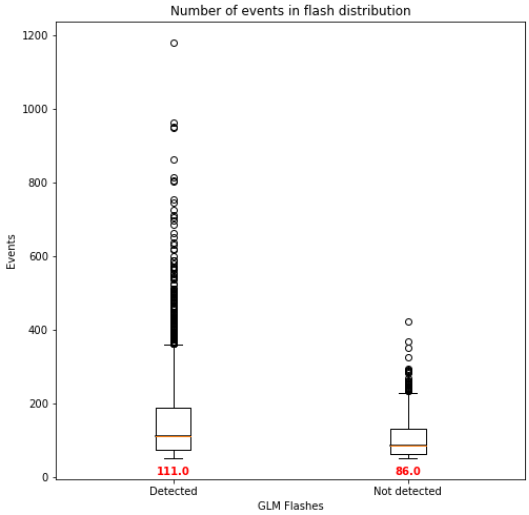

Figure 23. Number of GLM events for flashes detected and not detected by LMA.

For the number of events per flash, as for LMA sources, it seems that detected flashes have a higher number of events. The median number of events for detected is 111 and for not detected is 86.

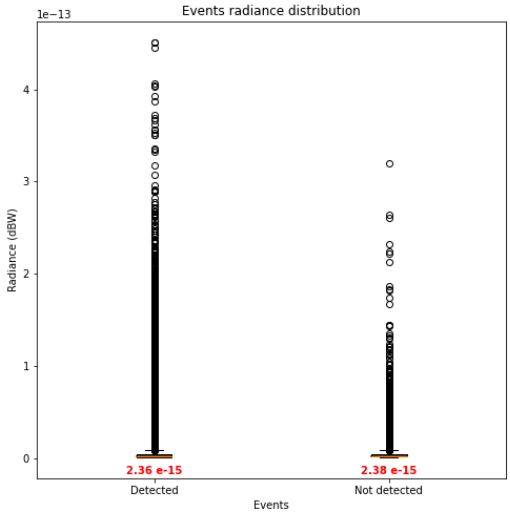

Figure 24. Radiance of GLM events for flashes detected and not detected by LMA.

The difference on radiance between detected and not detected flashes is very small and seems to not have an impact on the detection. To check if that is true, the influence of the radiance of the flashes over the detection efficiency has been calculated.

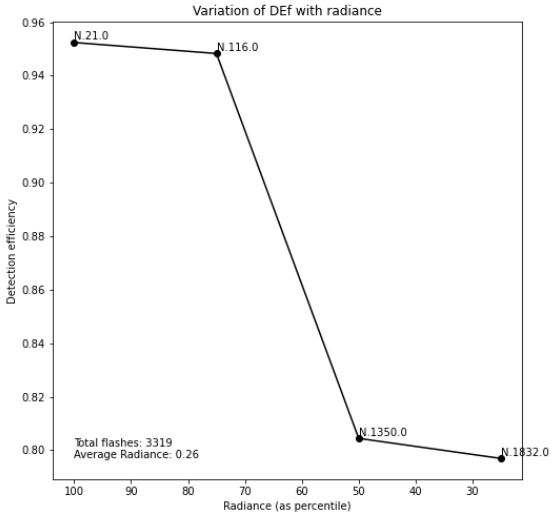

Figure 25. Influence of flash radiance (energy) to Detection efficiency.

The change of the detection efficiency with radiance is not as big as with LMA height, as it only changes a 20%, but there is still a decrease on the value related with the decrease of the radiance.

It is also interesting to see that as opposite as height, were most of the flashes had a higher value, for radiance a major part of the flashes are on the lowest levels detected (average radiance is 0.26).

3.4. Flash Duration

A general view of the differences in time duration can be obtained calculating the average duration values for GLM and LMA flashes. The duration of those flashes associated during the flash detection efficiency analysis using LMA and GLM as reference can also be calculated to see if there is a difference between their times and those of all the flashes in data.

To obtain the average duration of a flash, the time of the last and the first source or event are subtracted. When the time is calculated for all flashes, the average value for the case can be obtained. In the next table are shown the values obtained for each case.

Table 13. Mean time duration (in seconds) for LMA and GLM flashes and those flashes associated to a flash for analysis in sections 3.2 and 3.3.

Case	GLM flashes	GLM associated to LMA	LMA flashes	LMA associated to GLM
A	0.3843	0.4195	0.1770	0.2696
B	0.2840	0.2326	0.1097	0.1217
C	0.4199	0.4921	0.1813	0.2230
D	0.3466	0.4029	0.2205	0.28744
All	**0.3855**	**0.4415**	**0.1808**	**0.2495**

From these results, the average length of a LMA flash is not even a 50% of the length of a GLM flash. Knowing this information, it is quite understandable that GLM flashes end up matching with various LMA flashes.

Another remarkable thing is that flashes associated have longer durations than the average flash, it is also a very coherent result since the longer the flash is there are more possibilities that both systems detect it.

3.5. Location accuracy

Using the flash assignation from section 3.2., the sources and events belonging to a correlated flash are distributed in a grid of 0.075º lat x 0.075º lon in order to compare their location.

Doing this, a grid containing the density of flashes per cell (in percentage units) is obtained and is possible to compare the most crowded areas for LMA sources and GLM events in every case.

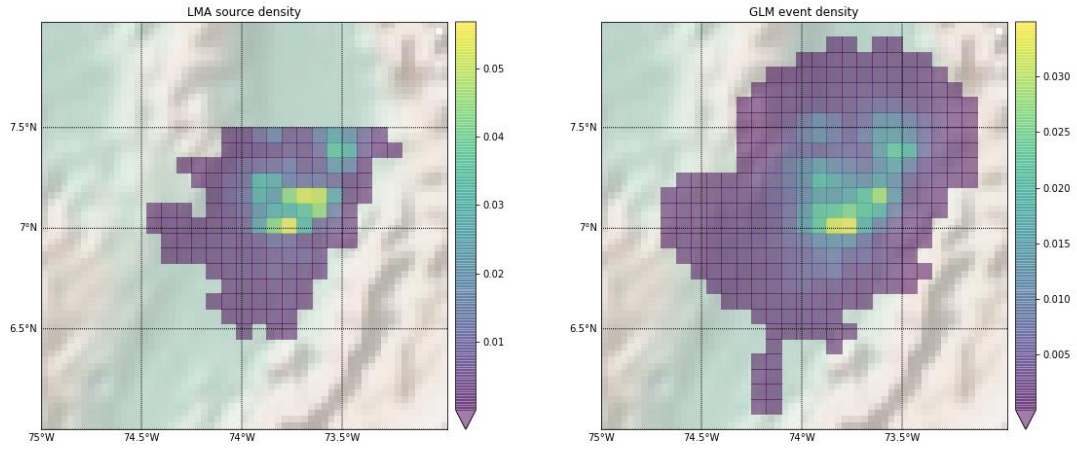

Figure 26. Density of data in grid in percentage units for case A. Left: LMA sources density. Right: GLM events density.

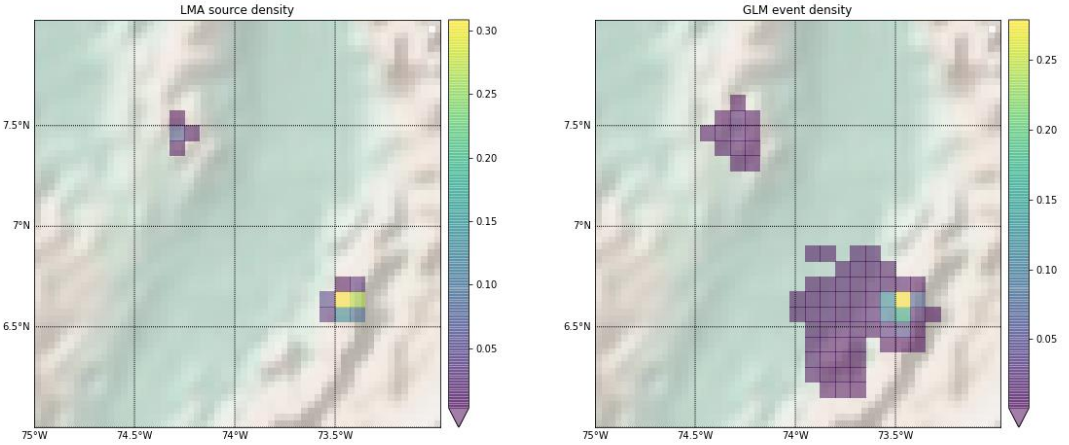

Figure 27. Density of data in grid in percentage units for case B. Left: LMA sources density. Right: GLM events density.

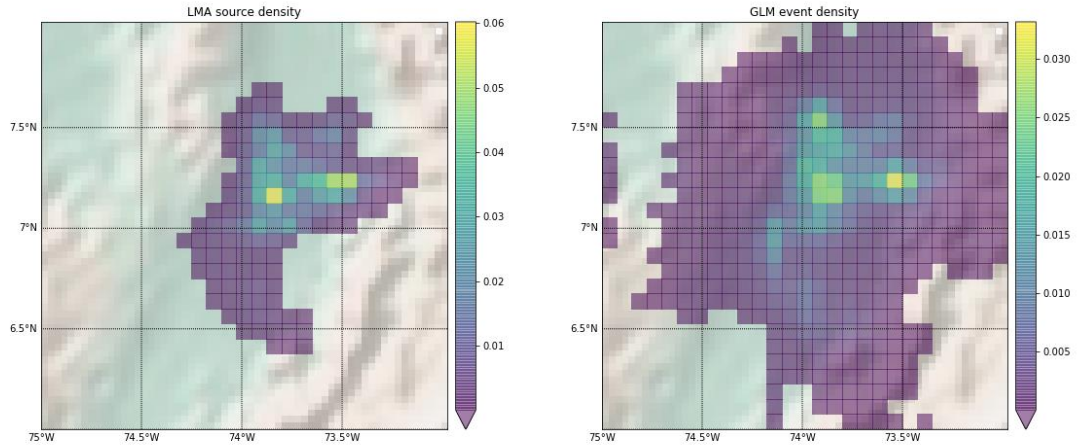

Figure 28. Density of data in grid in percentage units for case C. Left: LMA sources density. Right: GLM events density.

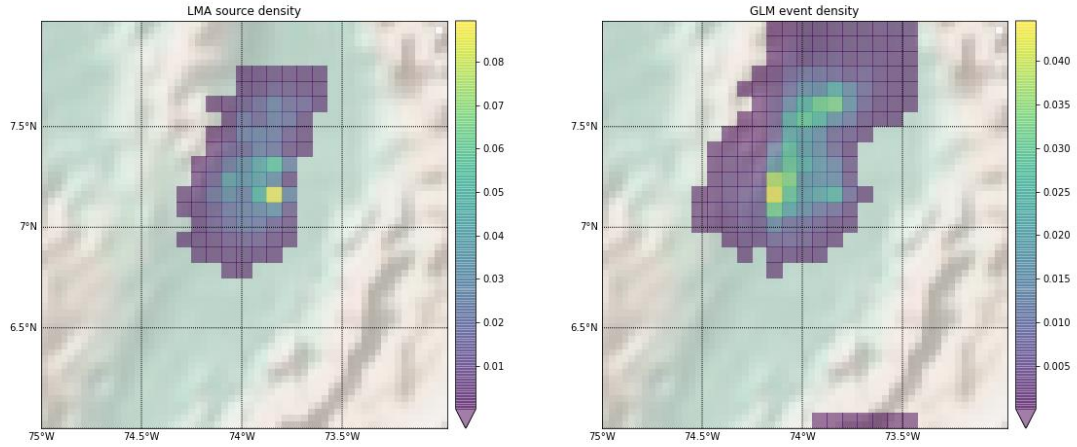

Figure 29. Density of data in grid in percentage units for case D. Left: LMA sources density. Right: GLM events density.

In all cases, GLM events cover more area than the sources of the flashes they're assigned to. In most cases, the shape of both LMA and GLM gridded data share some resemblance, except for case C (Figure 28), where the shape is completely different. Even though, the cells with high density are located at similar places. This can also be observed in cases A and B (Figures 26 and 27).

From this qualitative assessment it can be deduced that the location of the cells with more density for LMA and GLM is usually the same, but it would be necessary to do a more in deep analysis of the data to validate this first impression.

4. Conclusions

The goal of this project was to set the base for the validation tool that will be used with the METEOSAT Lightning Imager sensor, LI-MGT. To do so, it has been followed the comparison between the ISS-LIS imager and LMA made by the UPC Lightning Research Group.

Some of the comparison parameters of that study have been applied to the data obtained from the Geostationary Lightning Mapper (GLM), because LI-MGT will be also operating in a geostationary orbit. To compare the GLM data, LMA data from the Colombia LMA network at Barrancabermeja was used.

The main analysis of the project has been the study of the results obtained when calculating the Detection Efficiency of the GLM.

The most surprising finding was to have more than one LMA flash associated to a GLM flash. This was surprising because coming for the ISS-LIS comparison, the GLM flashes were expected to be smaller than LMA flashes. This was the reason to calculate the detection efficiency using GLM flashes as reference.

When changing the reference from LMA to GLM, the change when evaluating the number of LMA flashes associated to a GLM flash validated the theory that GLM flashes in are usually formed by more than one LMA flash.

The results obtained when the average time duration of GLM and LMA flashes also show that GLM flashes duration is usually more than twice the duration of a LMA flash. With all this results, the value of the detection efficiency for GLM flashes was expected to be better to the one obtained per LMA flashes, but that was not the case.

This could be for the simple reason that GLM data has a larger detection (or view) area, and generates more flashes than LMA. Therefore, it could not be possible for LMA to detect enough flashes to compare and assign to GLM flashes in order to be able to reach a higher value of detection efficiency.

In order to follow the work made in this project, it would be interesting to calculate the flash detection efficiency with LMA as reference using GLM groups of events instead of flashes.

Since groups of events are smaller than flashes, it could be possible that they match the size of the LMA flashes, then it could be obtained good values for the detection efficiency, and not loose the great assignation that is obtained when evaluating with GLM as reference.

It would be also interesting to keep analysing in detail the time duration of flashes, and also maybe group, in order to have a better understanding of the relation in time between the LMA and GLM flashes and groups. This could help to decide which data is more suitable to be used as reference when calculating the detection efficiency.

Bibliography

[1] L. Alter, “94.05.01: Meteorology.” http://teachersinstitute.yale.edu/curriculum/units/1994/5/94.05.01.x.html#b (accessed Sep. 07, 2017).

[2] “Lecture 2 - The electrical experiments of Benjamin Franklin.” http://www.atmo.arizona.edu/students/courselinks/spring13/atmo589/ATMO489_online/lecture_2/lect2_history_benjamin_franklin.html (accessed Sep. 15, 2020).

[3] “A New Era of Lightning Observations From Space: Journal of Geophysical Research: Atmospheres.” https://agupubs.onlinelibrary.wiley.com/doi/toc/10.1002/(ISSN)2169-8996.LIGHT1 (accessed Oct. 26, 2020).

[4] K. L. Cummins and M. J. Murphy, “An overview of lightning locating systems: History, techniques, and data uses, with an in-depth look at the U.S. NLDN,” *IEEE Transactions on Electromagnetic Compatibility*, vol. 51, no. 3 PART 1. pp. 499–518, 2009, doi: 10.1109/TEMC.2009.2023450.

[5] “Lecture 23 - Lightning location Pt. 2.” http://www.atmo.arizona.edu/students/courselinks/spring13/atmo589/ATMO489_online/lecture_23/lect23_lightning_location_pt2.html (accessed Sep. 21, 2020).

[6] “Instrument: Lightning Mapping Array (LMA) | Global Hydrology Resource Center (GHRC).” https://ghrc.nsstc.nasa.gov/home/micro-articles/instrument-lightning-mapping-array-lma (accessed Sep. 21, 2020).

[7] J. A. López *et al.*, “Charge Structure of Two Tropical Thunderstorms in Colombia,” *J. Geophys. Res. Atmos.*, vol. 124, no. 10, pp. 5503–5515, May 2019, doi: 10.1029/2018JD029188.

[8] D. J. Cecil, D. E. Buechler, and R. J. Blakeslee, “Gridded lightning climatology from TRMM-LIS and OTD: Dataset description,” *Atmos. Res.*, vol. 135–136, pp. 404–414, Jan. 2014, doi: 10.1016/j.atmosres.2012.06.028.

[9] S. Edgington, C. Tillier, and M. Anderson, “Design, calibration, and on-orbit testing of the geostationary lightning mapper on the GOES-R series weather satellite,” in *International Conference on Space Optics — ICSO 2018*, Jul. 2019, vol. 11180, p. 143, doi: 10.1117/12.2536063.

[10] “Mission Overview │ GOES-R Series.” https://goes-r.gov/mission/mission.html (accessed Sep. 26, 2020).

[11] S. J. Goodman *et al.*, “The GOES-R Geostationary Lightning Mapper (GLM),” *Atmos. Res.*, vol. 125–126, pp. 34–49, May 2013, doi: 10.1016/j.atmosres.2013.01.006.

[12] “Images: GLM │ GOES-R Series.” https://goes-r.gov/multimedia/instr-Glm.html (accessed Oct. 28, 2020).

[13] “ESA - Lightning Imager.” https://www.esa.int/Applications/Observing_the_Earth/Meteorological_missions/Meteosat_Third_Generation/Lightning_Imager (accessed Sep. 26, 2020).

[14] “MTG Design — EUMETSAT.” https://www.eumetsat.int/website/home/Satellites/FutureSatellites/MeteosatThirdGeneration/MTGDesign/index.html#li (accessed Sep. 26, 2020).

[15] R. Rew, G. Davis, S. Emmerson, and H. Davies, “NetCDF User’s Guide for C An Access Interface for Self-Describing, Portable Data,” 1997.

[16] J. Montanyà, O. Van Der Velde, N. Pineda, and J. A. López, “ISS-LIS data analysis based on LMA networks in Europe,” 2019. [Online]. Available: https://www.eumetsat.int/website/home/Data/ScienceActivities/ScienceStudies/ISSLISDataAnalysisBasedonLMANetworksinEurope/index.html.