Meteosat Third Generation Lightning Imager Level 2 expected performances

Doc.No. : EUM/RSP/REP/20/1179001
Issue : v1
Date : 20 August 2020
WBS/DBS :

EUMETSAT
Eumetsat-Allee 1, D-64295 Darmstadt, Germany
Tel: +49 6151 807-7
Fax: +49 6151 807 555
http://www.eumetsat.int

Page left intentionally blank

Change Record

<Ensure Version/Date in Change Record & Document Headers are the same as on the Profile in the DM Tool >

Version	*Date*	*DCR* No.* *if applicable*	*Description of Changes*
V1	20/08/2020		First issue

DCR = Document Change Request

Table of Contents

1 INTRODUCTION ... 7
1.1 Scope ... 7
1.2 Applicable Documents ... 7
1.3 Reference Documents ... 8
1.4 Terminology ... 8
1.5 Document Structure ... 10

2 LI REFERENCE PROCESSOR AND SIMULATED PERFORMANCES ... 11
2.1 Performance descriptors ... 11

3 INPUT DATA ... 13
3.1 Background Scene ... 13
3.2 Optical pulses ... 13
3.2.1 Flash modelling ... 14
3.2.1.1 Flash location ... 14
3.2.1.2 Number of pulses in flashes ... 14
3.2.1.3 Time difference between pulses in flashes ... 16
3.2.1.4 Location of pulses in flashes ... 16
3.2.1.5 Flash duration ... 16
3.2.2 Pulse modelling ... 17
3.2.2.1 Pulse diameter ... 17
3.2.2.2 Pulse radiance and pulse duration ... 18

4 SIMULATIONS AND RESULTS ... 20
4.1 Session `ID016 day` ... 20
4.1.1 Settings ... 20
4.1.2 Results ... 21
4.2 Session `ID017 night` ... 25
4.2.1 Settings ... 25
4.2.2 Results ... 26
4.3 Session `ID018 half` ... 31
4.3.1 Settings ... 31
4.3.2 Results ... 33

5 DISCUSSION ... 38

6 CONCLUSIONS ... 43

Table of Figures

Figure 1. Diagram describing the key processing steps used in the performance assessment simulations ... 11
Figure 2. Comparison between the pulse radiance distribution from FEGS and the event radiance distribution from LIS. Top panel: comparison between the LIS original distribution (red solid line) and the FEGS distribution (blue solid line) scaled to match the maximum of the LIS distribution at 5 µJ / (m^2 sr). Bottom panel: zoom on the high-end tail of the distribution from the bin at which the match was imposed ... 15
Figure 3. Example of distribution of time interval between pulses in a flash: reference input distribution (black solid line), distribution derived from the input (red solid line), and cumulative distributions from the two distributions (dashed lines with the respective colours). ... 16
Figure 4. Example of distribution of flash duration: reference LIS distribution (black solid line), distribution derived from the input (red solid line), and cumulative distributions from the two distributions (dashed lines with the same colour format). The peak at 2 secs is due to the flash truncation. ... 17

Figure 5. Example of the distribution of pulse area: reference input distribution (black solid line), distribution derived from the input (red solid line), and cumulative distributions from the two distributions (dashed lines with the same colour format). 18
Figure 6. Example of distribution of pulse radiance: reference input distribution (black solid line), distribution derived from the input (red solid line), and cumulative distributions from the two distributions (dashed lines with the same colour format). The minimum detectable energies for night and day are marked with vertical dashed lines. At the bottom, the cumulative fractions with respect to the minimum detectable energy values are reported. 19
Figure 7. Example of distribution of pulse duration: distribution derived from input pulse total radiance and pulse peak radiance (red solid line), distribution of total pulse duration from FEGS (black solid line), and cumulative distributions from the two distributions (dashed lines with the same colour format). 19
Figure 8. Location of the input pulses (red dots) from 10 different simulations and background scene (grey-scale images) for the session `ID016 day`. Each OC is highlighted with a different colour: OC1 is sky-blue, OC2 is orange, OC3 is green, and OC4 is dark-red. This colour convention, used to differentiate the different OCs, is consistent through the whole document. 21
Figure 9. ADP and DT variation for session `ID016 day.` Top panel – average ADP and standard deviation through the end-to-end filtering. The different colours represent the four OCs, the horizontal red-dashed line represents the Level 1b requirement for the ADP (see [SRD] for the details), the horizontal dark-blue dashed lines represent the average ADP at Level 2. Bottom panel – variation of the number of TTs and FTs (expressed as a fraction with respect to the total number at RTPP; filled circles and open circles, respectively); the colour coding of the OCs is the same used in the top panel. In both plots: the grey-shaded areas represent the results for each analysis step as if it was applied independently. The white areas of the plot represent the key sequential steps of the filtering, i.e., Level 0 on-board sequential filtering (RTPP, SDTF and MVF), Level 1b input (L1bin), Level 1b output at STC filter, Level 2 input (L2in), Level 2 final group filtering (ALL) and Level 2 final flash filtering (ALL). The different colour used to represent the results from the different OCs are in line with the colours used to highlight the OCs in Figure 8. In the top panel, the average values for the ADP, FDE and FFAR are presented. 22
Figure 10. DTs and ADP for session `ID016 day`. Top panel – Location of the Level 2 DTs (orange dots) from 10 different simulations. The rest of the image is formatted as Figure 8. Bottom panel – average ADP on a geographical grid with bins of 2.5 × 2.5 deg. 24
Figure 11. Detection threshold for session `ID016 day`. Histogram of the number of input pulses detected (red) and missed (blue) at Level 2 as a function of the radiance of the pulse (referred to the left y-axis). The green solid line and dots measure the fraction of detected pulses as a function of input radiance (referred to the right y-axis). The detection threshold is defined by the energy at which the fraction of detections reaches 50% (marked by the dashed red line). 25
Figure 12. Location of the input pulses in the simulation `ID017 night`. Same format as Figure 8. 26
Figure 13. ADP and DT variation for session `ID017 night`. Same format as Figure 9. 27
Figure 14. DTs and ADP for session `ID017 night`. Same format of Figure 10. 29
Figure 15. Detection threshold for session `ID017 night`. Same format as Figure 11. 30
Figure 16. ADP (top panel) and FDE (bottom panel) as a function of distance from the center of the LI FOV (the sub-satellite point, SSP) at Level 2 for session `ID017 night`. The histograms represent the total number of input pulses (blue solid line) and the number of input pulses detected at Level 2 (red solid line) as a function of distance from the SSP (referred to the right y-axis). The green solid line and dots measure the fraction of detected pulses at Level 2 as a function of distance from the SSP (referred to the left y-axis). Note that the distance of 5000 km corresponds to the angular distance of 45 degrees. 31
Figure 17. Location of the input pulses in the simulation `ID018 half.` Same format as Figure 8. 33
Figure 18. ADP and DT variation for session `ID018 half`. Same format as Figure 9. 34
Figure 19. DTs and ADP for session `ID018 half`. Same format as Figure 10. 36
Figure 20. Detection threshold for session `ID018 half`. Same format as Figure 11. 37

Table of Tables

Table 1. Settings for the `ID016 day` session 20
Table 2. Settings for the `ID017 night` session 25
Table 3. Settings for the `ID018 half` session 31

Table 4. The main MTG LI expected performance characteristics, averaged over the whole FOV and 10 simulation runs of each session. 43

1 INTRODUCTION

In the context of the preparation of the Meteosat Third Generation Lightning Imager (MTG LI) mission, the assessment of the expected (pre-flight) Level 2 performances is essential from a user and application perspective.

In addition, EUMETSAT and ESA need to understand where LI will stand performance-wise against GLM and LIS.

This assessment involves ESA and EUMETSAT experts, considering the sharing of responsibilities adopted for the MTG development programme:

1. EUMETSAT is responsible for the design and testing of the Level 2 processing/filtering.
2. EUMETSAT is responsible for the MTG LI System end-to-end performances and for the assessment of the impact of the Level 1b performances on the Level 2 final performances.
3. ESA is responsible of the MTG space segment overall design and procurement, including the implementation of the LI end to end Performance up to Level 1b.
4. EUMETSAT is responsible for the communication of the expected MTG LI lightning detection performances to future users of LI data.

This document presents the analysis rationale and methodology adopted by the LI Instrument Functional Chain Team (IFCT) that enabled the first pre-flight assessment of the LI Level 2 performances.

1.1 Scope

This document is addressed to different forums: the LI IFCT members of EUMETSAT and ESA, the LI Mission Advisory Group (LI MAG), EUMETSAT delegate bodies, and the public. In fact, it provides many technical details of the analysis approach that has been adopted to derive the final results, but at the same time, it communicates, in a compact fashion, the key figures to understand the expected LI lightning detection performances.

1.2 Applicable Documents

	Document Title	**Reference**
[SRD]	*MTG System Requirements Document [SRD]*	EUM/MTG/SPE/06/0032

1.3 Reference Documents

	Document Title	Reference
[LIL2ATBD]	*Algorithm Theoretical Basis Document (ATBD) for Level 2 processing of the MTG Lightning Imager data*	EUM/MTG/DOC/11/0155
[LI-9]	*LI Performance Model Description (April 30, Issue 11)*	MTG-GA-LI-DD-004
[LI-29]	*LI Performance and Calibration Analysis Issue 7 (April 30, 2020)*	MTG-GA-LI-RP-028
[ZHANG19]	*Time evolution of Satellite-Based Optical Properties in Lightning Flashes, and its Impact on GLM Flash Detection*	Zhang, D., Cummins, K. L., 2019. Time evolution of satellite-based optical properties in lightning flashes, and its impact on GLM flash detection. J. Geophys. Res.: Atmos, 125, e2019JD032024. https://doi.org/10.1029/2019JD032024
[FEGS19]	*Sub-flash Comparison of FEGS and GLM Observation from GOES-R Flight Campaign*	Quick, M. G., 2019. Sub-flash Comparison of FEGS and GLM Observation from GOES-R Flight Campaign. Presentation at the 2019 GLM Annual Science Team Meeting, September 10-12, 2019, Huntsville, AL. Available online.
[UPC19]	*ISS-LIS Data Analysis based on LMA Networks over Europe*	Montanyà, J., van der Velde, O., Pineda, N., López, J. ISS-LIS data analysis based on LMA networks in Europe. Scientific report for EUMETSAT. Available via link.

1.4 Terminology

Acronyms and Abbreviations

Acronym/Abbr.	Explanation
ADP	Average Detection Probability
BOL	Beginning Of Life
COM	Calibration and Obscuration Mechanism of the Meteosat Third Generation Flexible Combined Imager
DT	Detected Transient
GLM	Geostationary Lightning Mapper
FAR	False Alarm Rate
FCI	Flexible Combined Imager

Acronym/Abbr.	Explanation
FDE	Flash Detection Efficiency
FEGS	Fly's Eye GLM Simulator (https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20160000254.pdf)
FFAR	Flash False Alarm Rate
FT	False Transients
GLM	Geostationary Lightning Mapper (https://www.goes-r.gov/spacesegment/glm.html)
HYB	Hybrid filter at Level 1b
LIS	Lightning Imaging Sensor (https://ghrc.nsstc.nasa.gov/lightning/overview_lis_instrument.html)
JIT	Jitter-reconstruction filter at Level 1b
MVF	Micro Vibration Filter
OC	Optical Channel
PART	Particle Filter at Level 1b
PRE	Pre-processing Filter at Level 1b
RfD	Request for Deviation
RP	Reference Processor
RTPP	Real Time Pixel Processor
RTS	Random Telegraphic Signal (https://en.wikipedia.org/wiki/Burst_noise) filter at Level 1b
SDTF	Single DT Filter
SSP	Sub-Satellite Point
TT	True Detected Transient

Definitions

Definition/Term	Explanation
DT	It represents an energy excess, with respect to the background scene level, that is detected by the LI at RTPP, i.e., that is above the detection threshold. DTs are the basic element of the LI measurements, processing, and products (see [LIL2ATBD])
Group	Collection of DTs that are clustered over a single LI detection frame (see [LIL2ATBD])
Flash	Collection of groups that are correlated in space and time within a specific spatio-temporal window (see [LIL2ATBD]).

1.5 Document Structure

Section 1	Introduction (this Section)
Section 2	Short description of the EUMETSAT LI end-to-end reference processor, i.e., the simulator with which the pre-flight performances are derived
Section 3	Extensive description of the approach adopted for the definition of the inputs to the simulations, i.e., pulses and flashes to be captured by the LI end-to-end reference processor
Section 4	Extensive description of the simulations settings and Level 2 processing settings used for the computation of the Level 2 pre-flight performances and presentation of the analysis results
Section 5	Discussion of the results
Section 6	Main conclusions

2 LI REFERENCE PROCESSOR AND SIMULATED PERFORMANCES

The LI Reference Processor (hereafter LI RP) is a tool that EUMETSAT has put together by combining:

1. The software (coded in Matlab) employed by industry with the support and management of ESA to undertake the LI Level 0 and Level 1b performance assessment of LI.
2. The EUMETSAT LI Level 2 Matlab prototype in line with [LIL2ATBD].
3. The EUMETSAT Matlab software for generating the inputs for the simulations.

In Figure 1, the reader finds the key elements of the LI RP in the configuration that allows one to produce the end-to-end simulations and performance assessment at Level 0, Level 1b, and Level 2. For the description of the processing steps, from the "Instrument simulator" block up to the "Level 2 prototype processor", one can refer to [LI-9] and/or [LIL2ATBD]. The details on the definition of the "Simulated scene" are provided in Section 3.

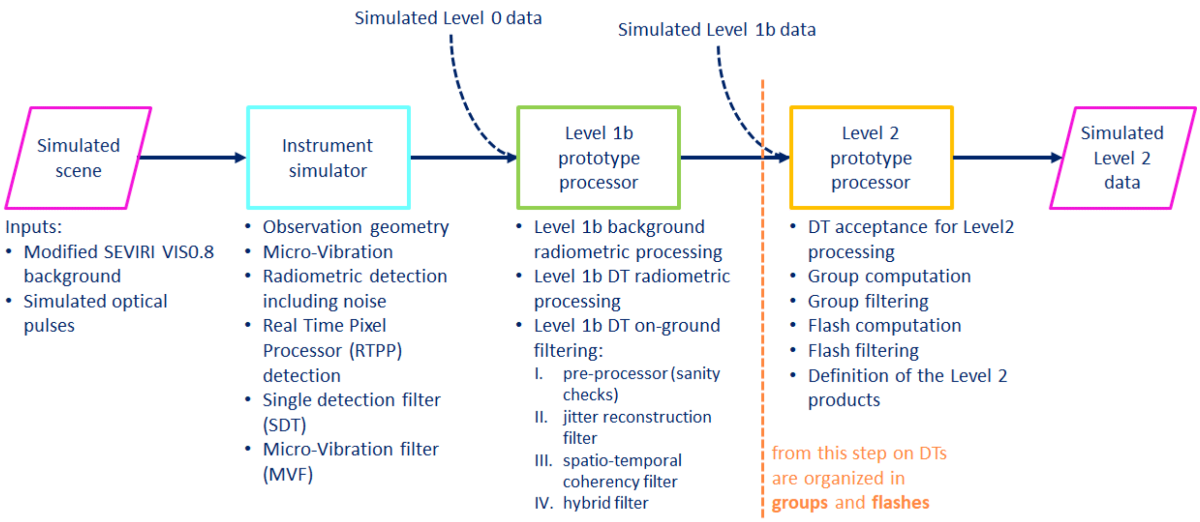


Figure 1. Diagram describing the key processing steps used in the performance assessment simulations.

The LI RP is the best description currently available of the end-to-end detection/filtering chain of the LI System. It is worth stressing that the Level 0 simulator provides one with the up-to-date instrument model description. This is in line with the latest instrument characterization information from industry which is supervised and confirmed by ESA. Together with this, the up-to-date Level 1b filtering prototype is used. Combining the Level 0 and Level 1b software with the up-to-date LI Level 2 processing prototype from EUMETSAT allows one to assess the impact of the up-to-date Level 1b performance on the final Level 2 performances.

2.1 Performance descriptors

The LI Level 2 performances are measured by means of three descriptors:

1. The Average Detection Probability (ADP) quantifies the number of pulses that has been detected with at least one Detected Transient (DT). This quantity is assessed at both Level 1b and Level 2, and it is expressed as a fraction of the total number of input pulses. This definition is in line with the one used for other instruments such as GLM, or LIS[1].

[1] For GLM and LIS this quantity is named pulse Detection Efficiency.

2. The Flash Detection Efficiency (FDE) quantifies the number of flashes that have been detected with at least one DT from one of its pulses at Level 2. It is expressed as a fraction of the total number of input flashes. This definition is in line with the one used for other instruments such as GLM, or LIS.

3. Flash False Alarm Rate (FFAR) is measured as the number of false flashes that are found at Level 2 every second.

In addition to these, the LI detection threshold is also assessed. This is the pulse radiance at which the fraction of detected pulses at Level 2 reaches 50% of the input pulses. The definition adopted here is in line with the one used for GLM (see [FEGS19]).

3 INPUT DATA

In order to assess the pre-flight LI Level 2 performances, different series of simulations (hereafter sessions) were performed. Each one is composed of 10 simulation runs for the computation of average performances associated to each session. The input settings that drive the different sessions are:
1. Background scene.
2. Properties of input pulses and/or flashes.

3.1 Background Scene

Three different Earth illumination conditions have been used as background scenes, namely:
1. Fully illuminated disk at 12:12 UTC on October 10, 2011, hereafter named `day`. The local illumination conditions covered, through the whole LI FOV, with this scenario are from 7am at the west edge to 5:30pm to the east edge.
2. Partially illuminated disk at 18:12 UTC on March 20, 2013, hereafter named `half`. The local illumination conditions covered, through the whole LI FOV, with this scenario are from 1pm at the west edge to 11:30pm to the east edge.
3. Dark disk at 00:12 UTC on March 20, 2013, hereafter named `night`. The local illumination conditions covered, through the whole LI FOV, with this scenario are from 7pm at the west edge to 5:30am to the east edge.

Such scenarios have been selected following the industry approach documented in [LI-9]. They are supposed to represent typical illumination conditions that LI will observe during the day.

3.2 Optical pulses

The approach adopted by industry to evaluate the pre-flight performances is described in detail in [LI-9] and [SRD]. For the benefit of the reader, we present here a short summary:
1. Pulses have a round shape with a fixed diameter of 10 km.
2. Pulses have a fixed duration of 0.6 millisecond.
3. Pulses have spatially uniform radiance whose value is proportional to the background scene over which the pulses are located (i.e., pulses placed on bright clouds always have high radiances).
4. The temporal profile of a pulse is modelled with a step function, whose integral over the 0.6 millisecond duration gives one the total radiance.
5. Pulses are located over the background scene by means of the cloud mask associated to the background. It is important to stress that a cloud mask includes any kind of cloud, from optically thin clouds to thick clouds associated with atmospheric convection.
6. Pulses are treated independently, i.e., they are not included in the sequences of pulses correlated in space and time (i.e., lightning flashes).

The approach adopted by EUMETSAT is presented in detail in the following sections. Here we list three points that highlight difference with respect to industry approach:
1. Pulses are located within flashes (see Section 3.2).
2. Both pulses' and flashes' properties are derived from measurements of real lighting, both in the visible and in the radio (see Section 3.2).
3. Flashes are located (over the background, see Section 3.1) where lightning activity can potentially happen (see Section 3.2.1.1).

3.2.1 Flash modelling

The LI Level 2 performances (see Section 2.1) are formulated for flashes (see Section 1.4 or [LIL2ATBD] for a more detailed definition of a flash). This implies that, in order to simulate realistic LI Level 2 performances, one must simulate the detection of pulses within simulated flashes.

3.2.1.1 Flash location

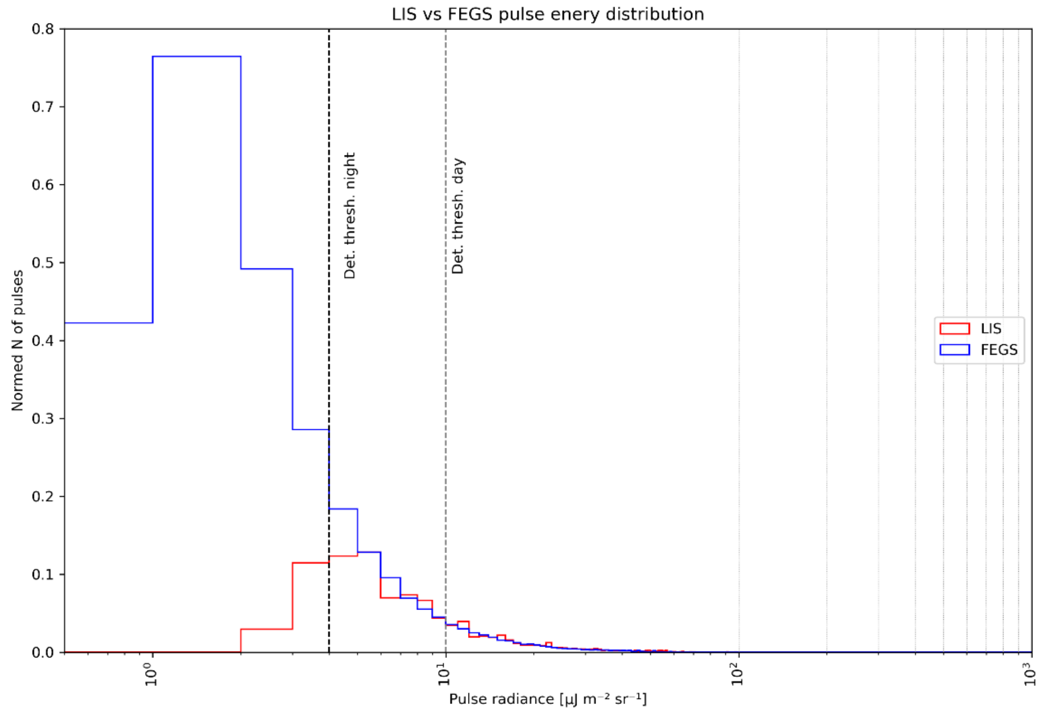
Lightning flashes can appear in both areas with convective precipitation and areas with stratiform precipitation. Thus, in the simulations, flashes are randomly placed over regions with precipitation, these being derived from SEVIRI Multi-Sensor Precipitation Rate Estimate Level 2 product associated to each background employed for the simulations (see Section 3.1). In one case, namely the `night` scenario, the SEVIRI Cloud Mask Level 2 product associated to the background is employed (as done by industry; see [LI-9]). Since the `night` scenario is a uniformly dark background with no possibility of distinguishing cloud-free from cloudy regions in the Visible, the Cloud Mask is employed since it allows one to have a uniform geographical sampling of the LI field of view[2].

3.2.1.2 Number of pulses in flashes

The number of pulses per flash is derived from the distribution of the number of groups per flash from LIS. However, such distribution is biased by the LIS sensitivity. From a comparison between LIS and FEGS distributions of pulse radiances, one can quantify the number of pulses typically missed by LIS to derive a correction factor for the number of pulses per flash from LIS. As a first step, the check of the consistency between LIS and FEGS for energies above the LIS most probable detected radiance was undertaken. In detail, after imposing the match between the two distributions at the bin of the maximum counts for LIS (i.e., the most probable detected radiance at 5 µJ / (m^2 sr)), it is possible to verify the very good agreement between the two distributions above 5 µJ / (m^2 sr) (Figure 2, bottom panel). From the same comparison, one learns that LIS misses a large fraction of the faint pulses that are detected by FEGS. In fact, below 5 µJ / (m^2 sr), FEGS detects about 8 times more pulses than LIS (Figure 2, top panel). Pulses with radiance below 5 µJ / (m^2 sr) constitute 30% of the LIS distribution, meaning that 30% of the LIS number of pulses per flash must be boosted by a factor 8 to match the FEGS number of detections. This means that the total number of pulses per flash must be boosted by a factor of approximately 3[3].

[2] The Cloud Mask is a much looser selection mask with respect to the Muti-Sensor Precipitation Rate Estimate mask, see the difference between the location of pulses in Figure 8 and Figure 12.

[3] $8 \times 0.3 + 0.7 = 3.1$.



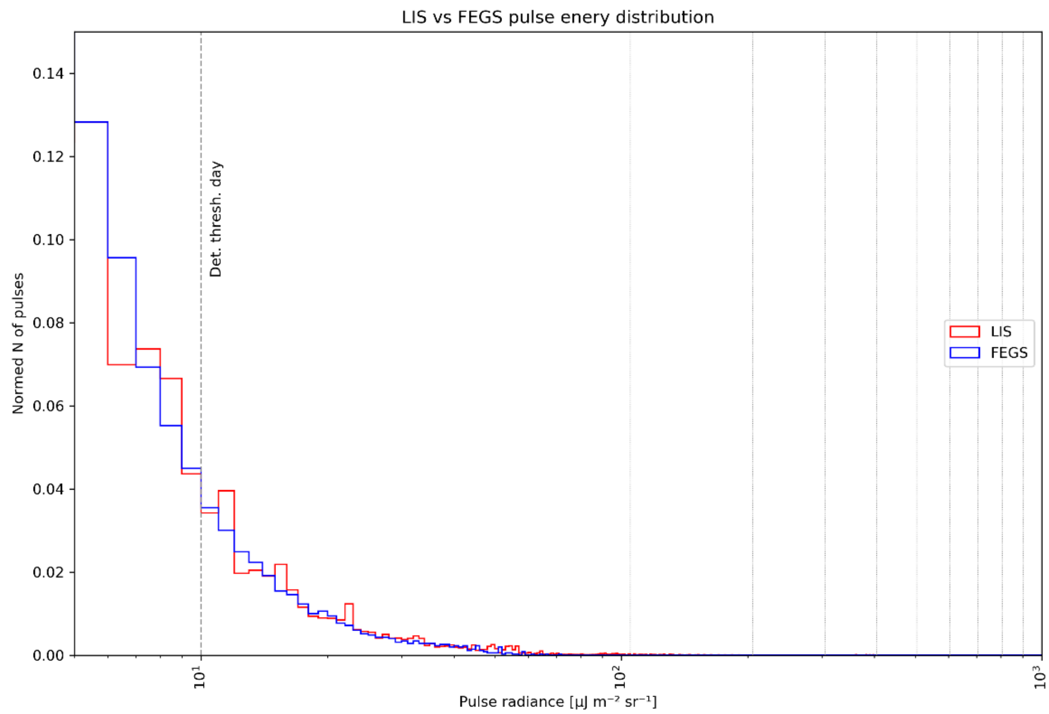


Figure 2. Comparison between the pulse radiance distribution from FEGS and the event radiance distribution from LIS. Top panel: comparison between the LIS original distribution (red solid line) and the FEGS distribution (blue solid line) scaled to match the maximum of the LIS distribution at 5 µJ / (m² sr). Bottom panel: zoom on the high-end tail of the distribution from the bin at which the match was imposed.

3.2.1.3 Time difference between pulses in flashes

When generating sequences of pulses in a flash, the time difference between pulses is derived from the distribution of time differences between groups in flashes of LIS. Despite the use of the boost factor introduced in Section 3.2.1.2, no correction was applied to this distribution since this is already strongly dominated by short time intervals as one would expect. The resulting distribution of the time difference between pulses is presented in Figure 3.

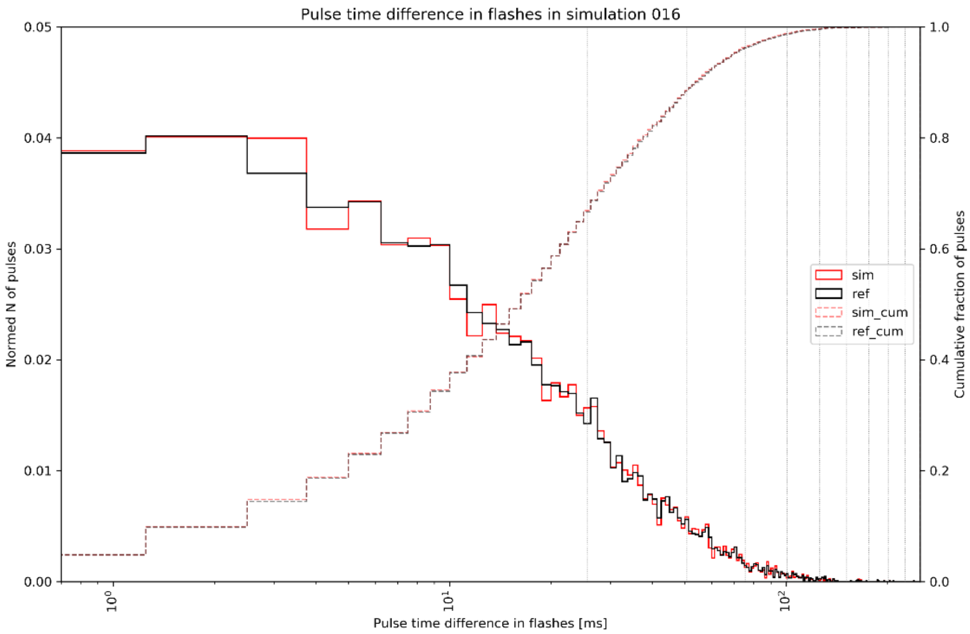


Figure 3. Example of distribution of time interval between pulses in a flash: reference input distribution (black solid line), distribution derived from the input (red solid line), and cumulative distributions from the two distributions (dashed lines with the respective colours).

3.2.1.4 Location of pulses in flashes

When generating the pulses in a flash, a location for each pulse with respect to the flash barycentre is produced. This is done by using a uniform angular distribution around the flash barycentre and a distribution for the distance of the pulse from the barycentre that stems from the flash area derived with the convex hull method on LMA flashes detected over Europe (Joan Montanya private communication). The distribution for the distance from the barycentre is derived with the assumption that the flash area is circular in shape. The typical radius of flashes varies between 10 km and 30 km.

3.2.1.5 Flash duration

From the combination of the flash properties described in Sections 3.2.1.2 and 3.2.1.3, stems the flash duration (see Figure 4). The maximum flash duration that is allowed for the analysis is 2 sec. This flash "truncation" has been imposed since it allows one to have short simulations (with maximum duration of 2 sec), and at the same time a very good representation of the flash duration property. In fact, from LIS statistics one learns that the vast majority of flashes have duration below 1 sec (see "ref" distribution in Figure 4).

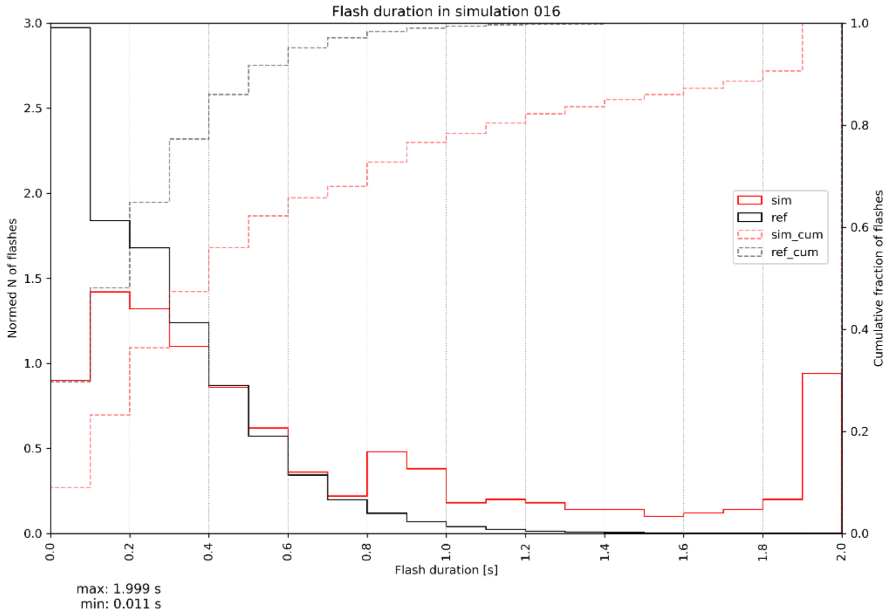


Figure 4. Example of distribution of flash duration: reference LIS distribution (black solid line), distribution derived from the input (red solid line), and cumulative distributions from the two distributions (dashed lines with the same colour format). The peak at 2 secs is due to the flash truncation.

3.2.2 Pulse modelling

Lightning pulses are described by a few properties in the simulations:

1. Beginning time of the pulse.
2. Location of the pulse.
3. Pulse diameter.
4. Pulse radiance.
5. Pulse duration.

Properties 1 and 2 stem from the flash modelling (see Sections 3.2.1.3, with respect to the beginning of the flash, and 3.2.1.4, respectively). The three remaining properties are defined as described in the following sections.

3.2.2.1 Pulse diameter

This property is derived from the distribution of the number of DTs in groups from LIS ($\#_{DT}$). Behind this approach, there is the assumption of the relation: groups-pulses. Knowing that a single LIS DT has an area of about 16 km^2 (A_{DT}) on the Earth surface, the pulse diameter is derived as:

$$d = 2 \cdot \sqrt{\frac{\#_{DT} \cdot A_{DT}}{\pi}} = 8\sqrt{\frac{\#_{DT}}{\pi}} \text{ [km]}.$$

Moreover, the pulse is modelled as a uniform-in-radiance disk. The pulse is then smoothed since this uniform disk is convolved with both the spatial response function of the optics of LI and the pixel response function of the LI detector (see the explanation in [LI-9]). The resulting distribution of the pulse area is presented in Figure 5.

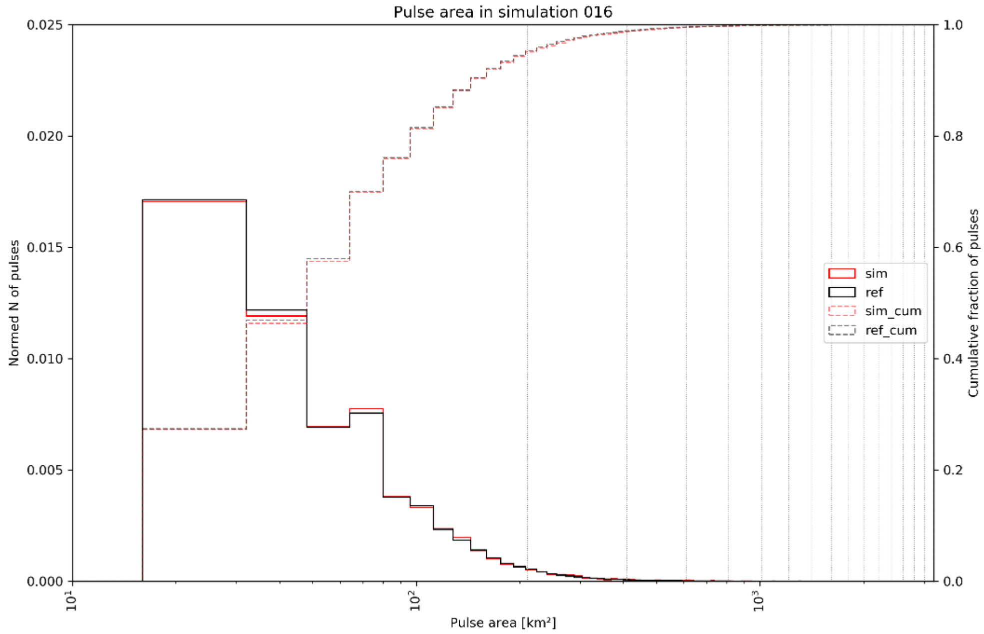


Figure 5. Example of the distribution of pulse area: reference input distribution (black solid line), distribution derived from the input (red solid line), and cumulative distributions from the two distributions (dashed lines with the same colour format).

3.2.2.2 Pulse radiance and pulse duration

These two properties of the pulses are coupled in the pulse modelling. In fact, the temporal profile of a pulse is described with a Maxwell function, with a normalized integral over the pulse duration and a maximum that is reached at one third of the pulse duration. From this model one finds that three pulse properties are related by a simple relation:

$t_P = k \cdot \frac{E_P}{P_P}$, where t_P is the pulse duration, E_P is the pulse total radiance, and P_P is the pulse peak radiance, respectively.

From FEGS measurements, one learns that there is an evident correlation between E_P and P_P. Such correlation has been captured in a 2-D distribution relating these two quantities (see [FEGS19]). This same distribution has been employed to derive pairs $[E_P, P_P]$ for each pulse of the inputs to our simulations, that are then employed to derive t_P through the simple relation expressed above. A check has been done on the distribution of t_P so derived: this is in very good agreement with a family of pulse-duration distributions that can be derived from FEGS (see the 10-10 width, 50-50 width and 10-90 rise time in [FEGS19]). Finally, from FEGS observations one learns that the Maxwell function description for the pulse time profile is very much appropriate. The resulting distribution of the pulse radiance and duration are presented in Figure 6 and Figure 7. In the latter, an example of pulse duration distribution derived from FEGS is reported for a direct comparison.

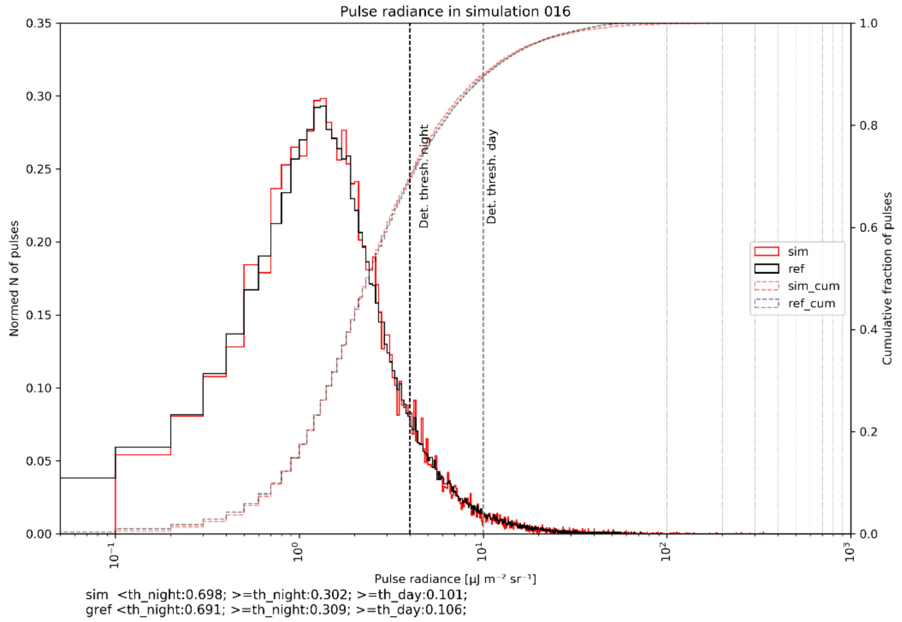


Figure 6. Example of distribution of pulse radiance: reference input distribution (black solid line), distribution derived from the input (red solid line), and cumulative distributions from the two distributions (dashed lines with the same colour format). The minimum detectable energies for night and day are marked with vertical dashed lines. At the bottom, the cumulative fractions with respect to the minimum detectable energy values are reported.

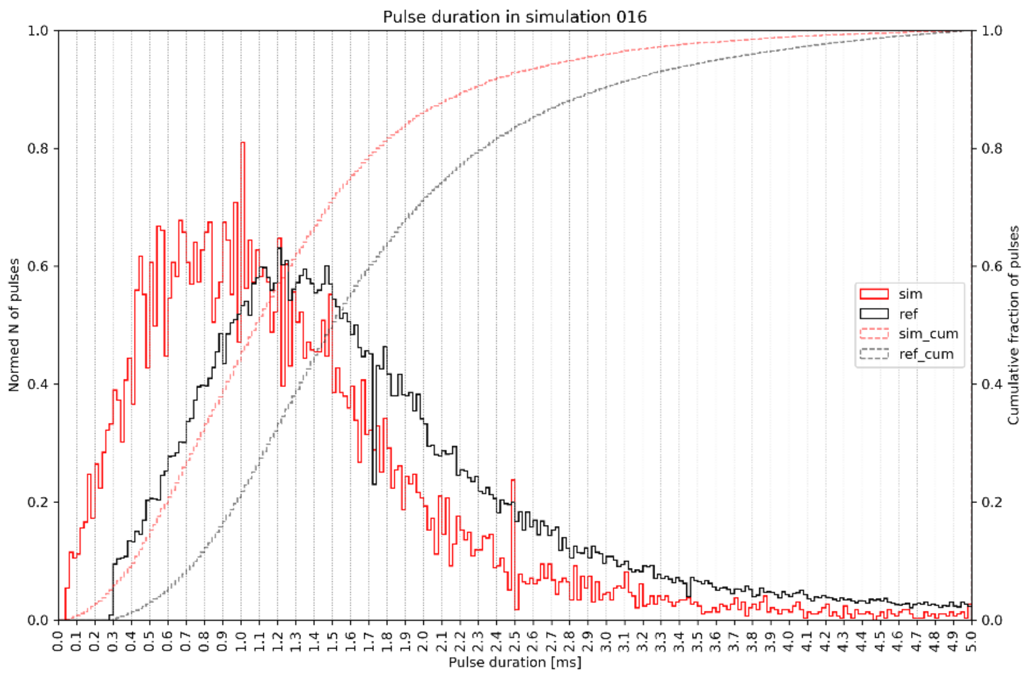


Figure 7. Example of distribution of pulse duration: distribution derived from input pulse total radiance and pulse peak radiance (red solid line), distribution of total pulse duration from FEGS (black solid line), and cumulative distributions from the two distributions (dashed lines with the same colour format).

4 SIMULATIONS AND RESULTS

Sections 4.1, 4.2, and 4.3 present the settings of the sessions and the results derived. When interpreting the results, it is important to consider that both Level 1b and Level 2 processing involves multiple parallel filters with only some of them being used to discern between True Transients (TTs) and False Transients (FTs). In fact, applying individual filters at Level 1b/Level 2 group/Level 2 flash filtering could results in much lower ADP than the final outcome of the filtering done by accounting for the results from multiple filters. In defining the settings of the Level 2 filters, some filters are disregarded. For example, the group size and relative Sobel gradient filter in the group analysis below, see e.g. Figure 9 are not used in the final decision on the group TT/FT classification. With the aim of easing the interpretation of the filtering plots, all parallel filters have grey background in the plots, whereas sequential filters have white background (see Figure 9, Figure 13 and Figure 18). More information about the overall filtering concept and detailed descriptions of individual filters can be found in [LIL2ATBD].

4.1 Session `ID016 day`

4.1.1 Settings

Table 1. Settings for the `ID016 day` session

Session	Settings	
`ID016 day`	Level 0 simulator (see [LI-9])	Standard settings at BOL and no COM component of micro-vibration[4]
	Level 1b prototype (see [LI-9])	Standard settings
	Level 2 filtering (see [LIL2ATBD])	1. Particle filter on Groups 2. Radiance filter, with 0.006 W / (m^2 sr) threshold, for at least half of DTs on Groups 3. Footprint filter at 3 DT on Flashes 4. Distance correlation between groups within 10 km on Flashes 5. Average Sobel gradient normalized to the minimum background with threshold at 10 on Flashes
	Number of flashes	50 per simulation run
	Number of pulses	About 1500 per simulation run
	Flash location	Derived from the Multi-Sensor Precipitation Rate Estimate Level 2 SEVIRI product
	Pulse properties	See Section 3.2.2
	Flash properties	See Section 3.2.1
	Maximum simulation duration (i.e., maximum flash duration)	2 sec

[4] The COM micro-vibration component is generated by the movement of the FCI calibration mechanism and is known to have a minor impact on the overall micro-vibration spectrum (ESA private communication).

	Simulation runs	10

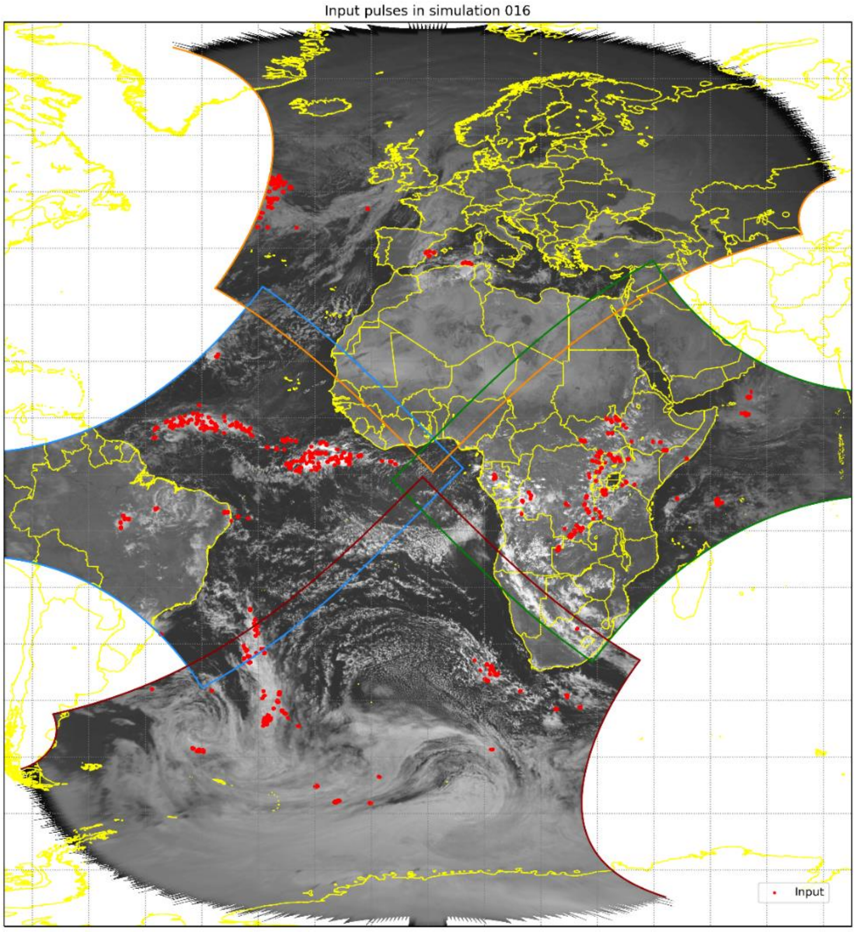


Figure 8. Location of the input pulses (red dots) from 10 different simulations and background scene (grey-scale images) for the session `ID016 day`. Each OC is highlighted with a different colour: OC1 is sky-blue, OC2 is orange, OC3 is green, and OC4 is dark-red. This colour convention, used to differentiate the different OCs, is consistent through the whole document.

4.1.2 Results

In the top panel of Figure 9 the reader finds the Average Detection Probability (ADP) variation through the end-to-end processing. The average-over-10-simulations Level 1b ADP is of 11% while at Level 2 it is 9%. The impact of the Level 2 filtering on the average ADP (i.e., the reduction of the total number of TTs) is very small, i.e., about 2%. At the same time, the reduction of the total number of FTs through the Level 2 filtering steps is of about three orders of magnitude (bottom panel of Figure 9). The average Flash Detection Efficiency (FDE) is 56% for an average Flash False Alarm Rate (FFAR) of 6 flashes per second per OC, i.e., about 24 false flashes per second over the whole LI field of view. It is worth noting that for OC1 and OC4 the average FFAR is higher than for the other OCs.

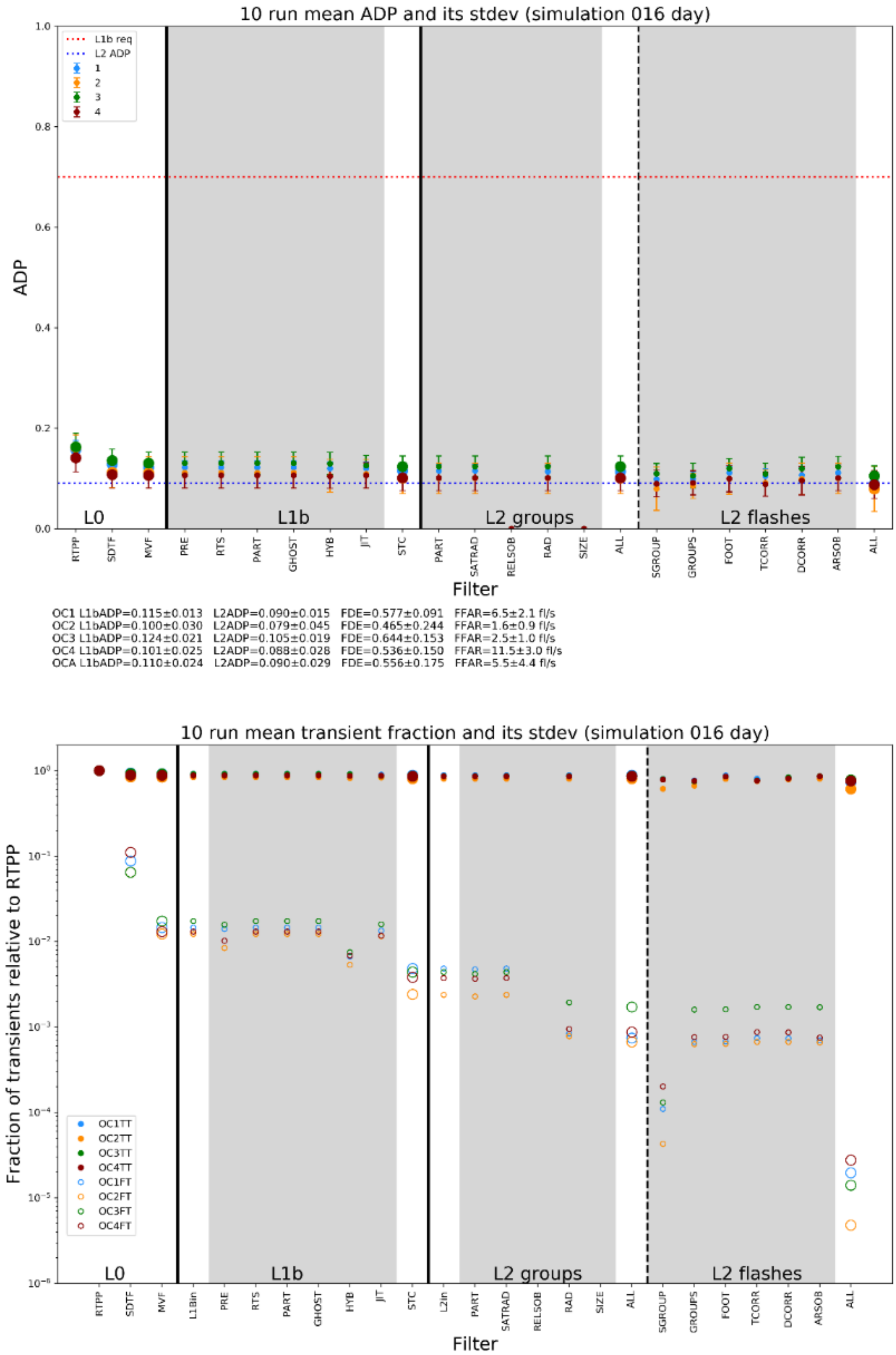
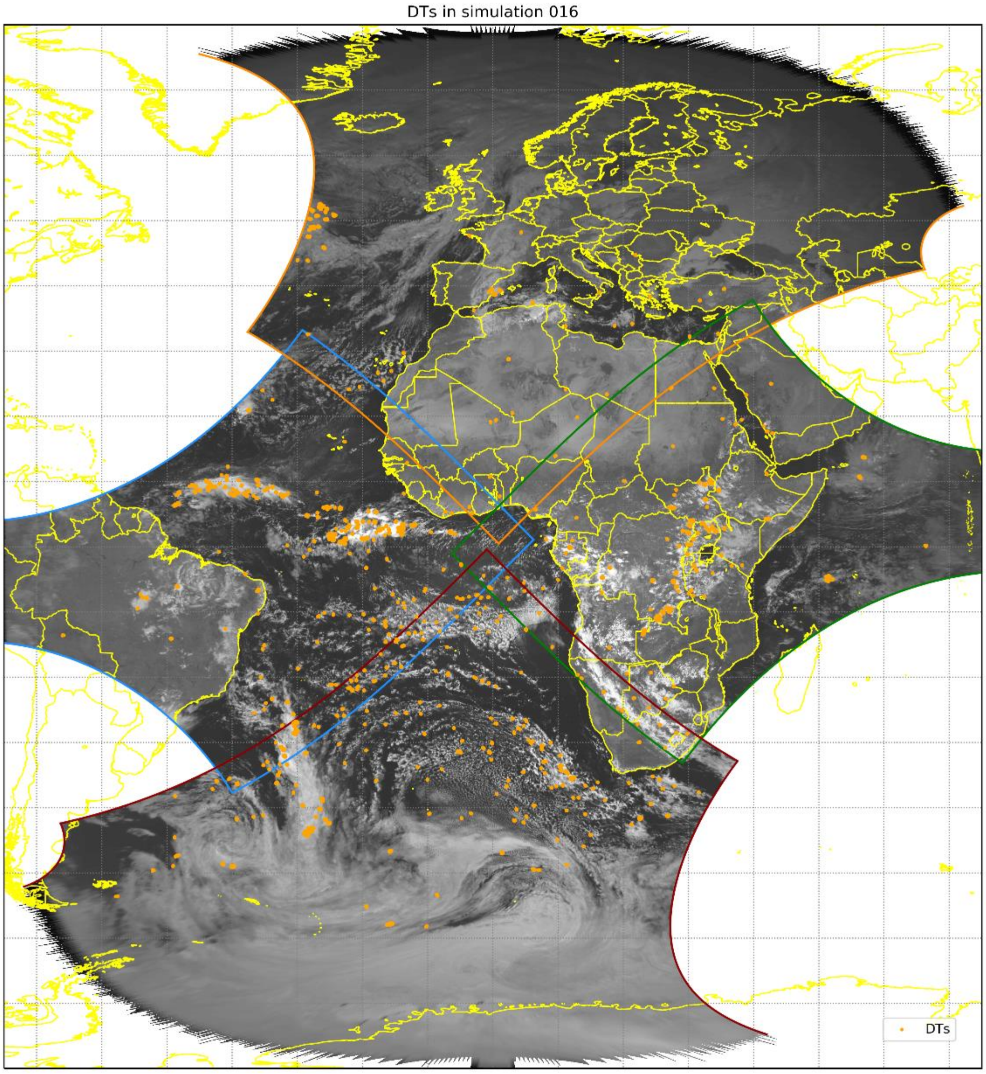


Figure 9. ADP and DT variation for session `ID016 day`. *Top panel – average ADP and standard deviation through the end-to-end filtering. The different colours represent the four OCs, the horizontal red-dashed line represents the Level 1b requirement for the ADP (see [SRD] for the details), the horizontal dark-blue dashed lines represent the average ADP at Level 2. Bottom panel – variation of the number of TTs and FTs (expressed as a fraction with respect to the total number at RTPP; filled circles and open circles, respectively); the colour coding of the OCs is the same used in the top panel. In both plots: the grey-shaded areas represent the results for each analysis step as if it was applied independently. The white areas of the plot represent the key sequential*

steps of the filtering, i.e., Level 0 on-board sequential filtering (RTPP, SDTF and MVF), Level 1b input (L1bin), Level 1b output at STC filter, Level 2 input (L2in), Level 2 final group filtering (ALL) and Level 2 final flash filtering (ALL). The different colour used to represent the results from the different OCs are in line with the colours used to highlight the OCs in Figure 8. In the top panel, the average values for the ADP, FDE and FFAR are presented.

In the top panel of Figure 10, one finds the location of the DTs at Level 2 from all the simulations of session `ID016 day`. In this figure one can individuate the region from which most of the False Transients (FTs) for OC1 and OC4 are emerging: a cloud system in the southern Atlantic Ocean. In the bottom panel of Figure 10, one can see the geographical variation of the average ADP. Despite the poor geographical coverage of the simulated lightning pulses, one can appreciate the performance degradation when moving from the center of the LI FOV towards the edges. The LI daytime Level 2 detection threshold (the energy at which the fraction of detections reaches 50%; see Figure 11) is 14.3 µJ / (m^2 sr).



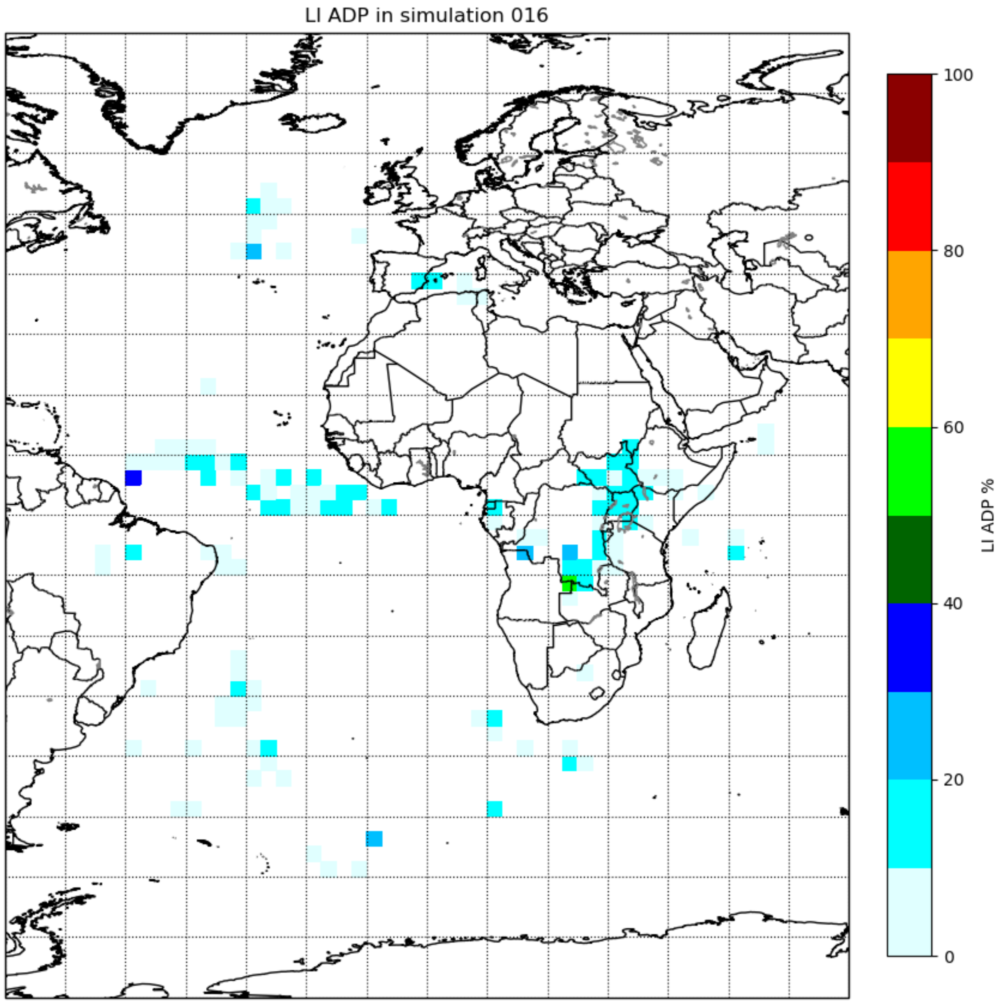


Figure 10. DTs and ADP for session `ID016 day`***. Top panel – Location of the Level 2 DTs (orange dots) from 10 different simulations. The rest of the image is formatted as Figure 8. Bottom panel – average ADP on a geographical grid with bins of 2.5 × 2.5 deg.***

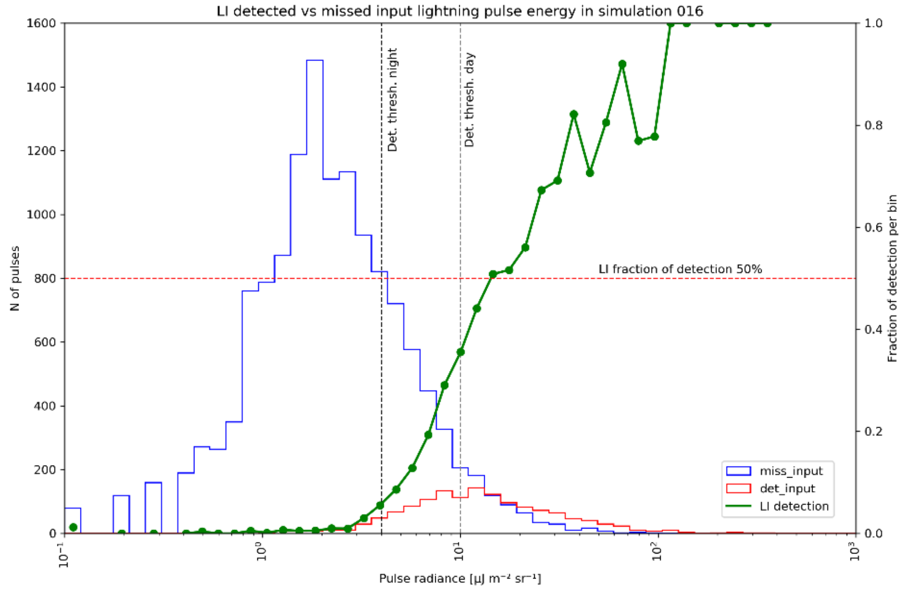


Figure 11. Detection threshold for session `ID016 day`. Histogram of the number of input pulses detected (red) and missed (blue) at Level 2 as a function of the radiance of the pulse (referred to the left y-axis). The green solid line and dots measure the fraction of detected pulses as a function of input radiance (referred to the right y-axis). The detection threshold is defined by the energy at which the fraction of detections reaches 50% (marked by the dashed red line).

4.2 Session `ID017 night`

4.2.1 Settings

Table 2. Settings for the `ID017 night` session

Session	**Settings**	
`ID017 night`	Level 0 simulator (see [LI-9])	Standard settings at BOL and no COM component of micro-vibration
	Level 1b prototype (see [LI-9])	Standard settings
	Level 2 filtering (see [LIL2ATBD])	1. Particle filter on Groups 2. Radiance filter, with 0.002 W / (m^2 sr) threshold, for at least half of DTs on Groups 3. Number of Groups at 3 on Flashes 4. Footprint filter at 3 DT on Flashes
	Number of flashes	50 per simulation run
	Number of pulses	About 1500 per simulation run
	Pulse location	Derived from the SEVIRI Cloud Mask Level 2 product
	Pulse properties	See Section 3.2.2

	Flash properties	See Section 3.2.1
	Maximum simulation duration (i.e., maximum flash duration)	2 sec
	Simulation runs	10

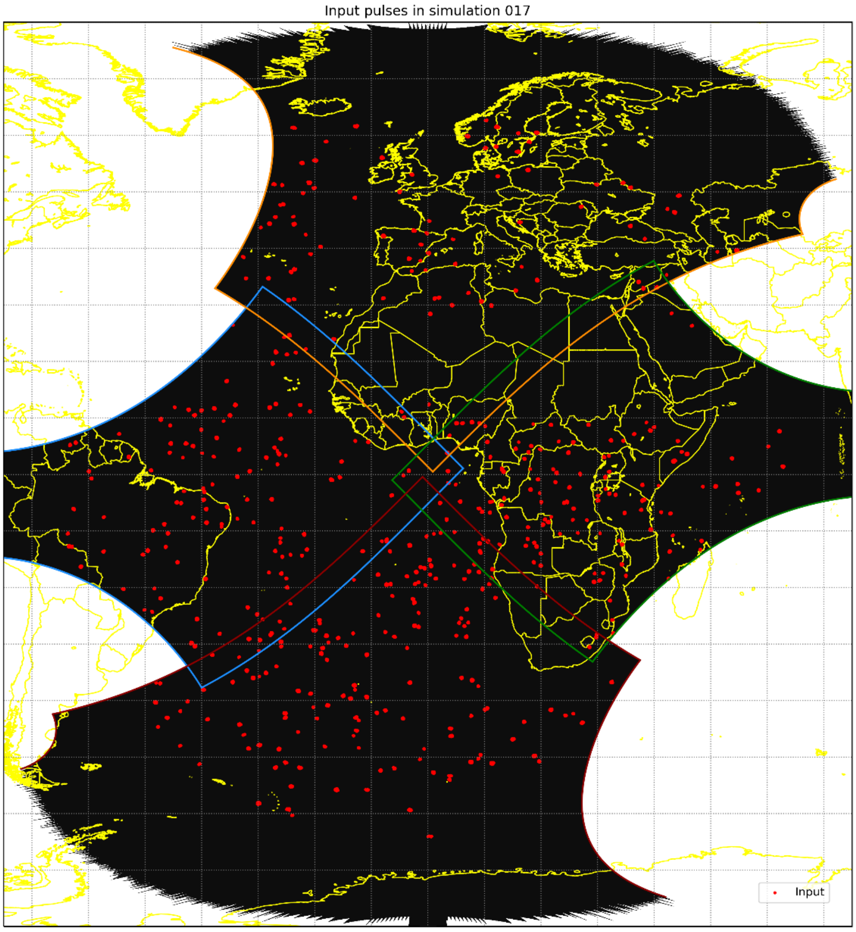
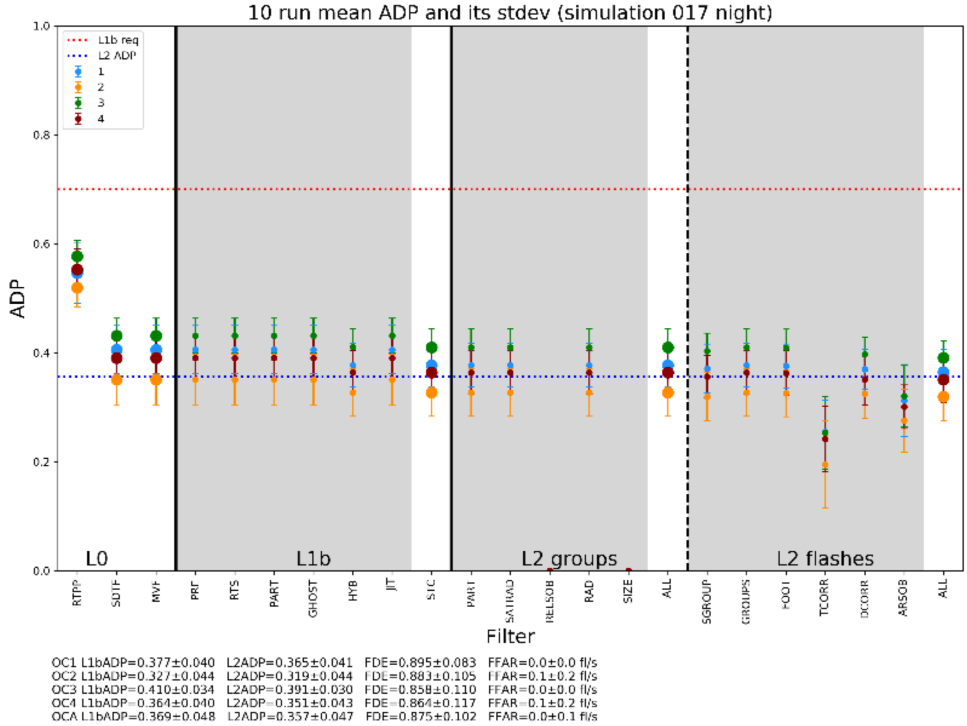


Figure 12. Location of the input pulses in the simulation `ID017 night`. Same format as Figure 8.

4.2.2 Results

The average-over-10-simulations Level 1b ADP is of 37% while at Level 2 it is 36% (Figure 13 top panel). The impact of the Level 2 filtering on the average ADP (i.e., the reduction of the total number of TTs) is very small, only about 1%. At the same time, the reduction of the total number of FTs through the Level 2 filtering steps is still large, i.e. about two orders of magnitude (bottom panel of Figure 13). The average Flash Detection Efficiency (FDE) is around 88%. The average Flash False Alarm Rate (FFAR) is very low, i.e. below 0.1 flashes per second over the whole LI field of view.



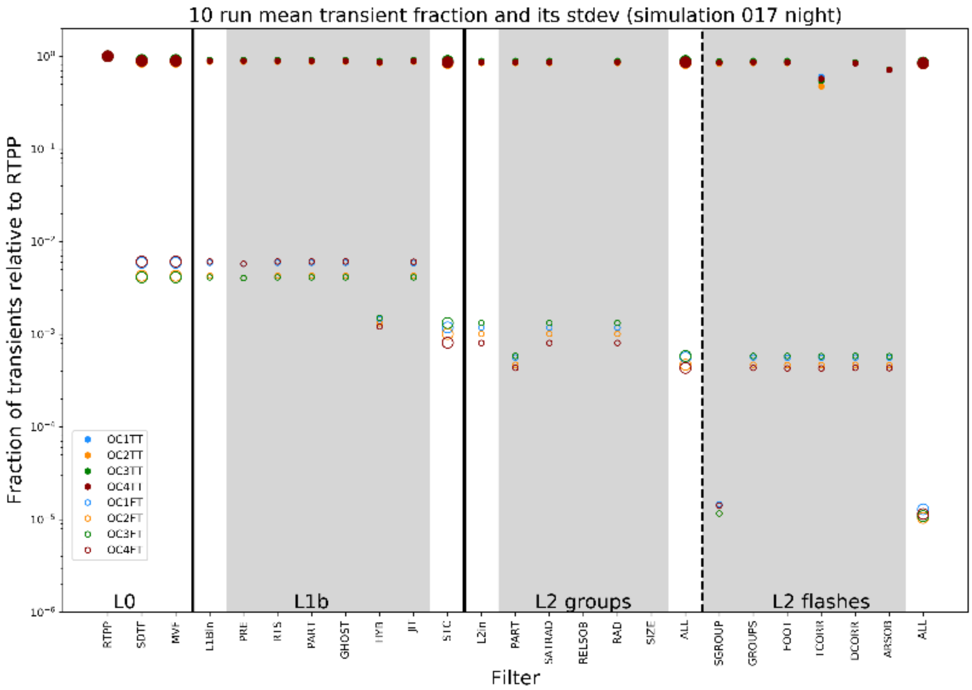
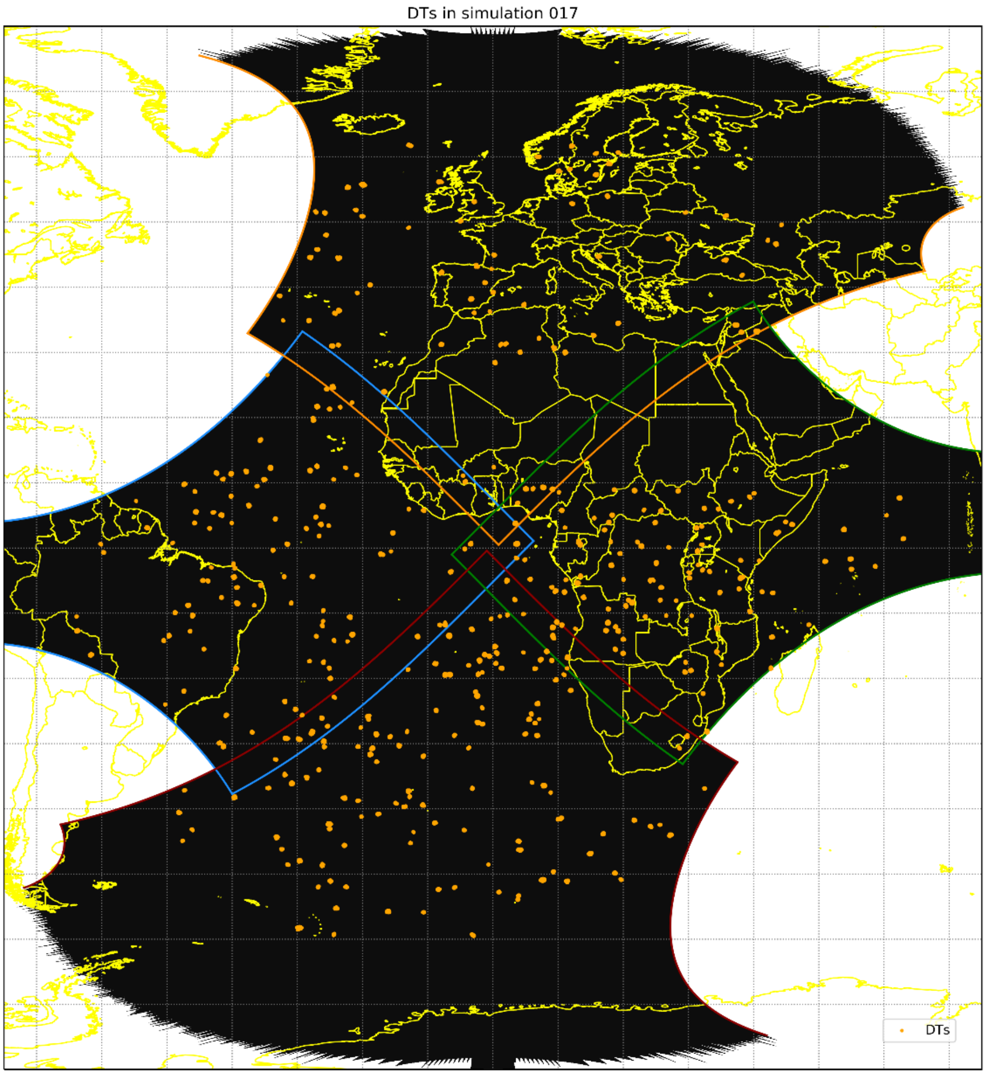


Figure 13. ADP and DT variation for session* `ID017 night`*. Same format as Figure 9.

In the top panel of Figure 14, one finds the location of the DTs at Level 2 from all the simulations of session `ID017 night`. The comparison with input pulses (see Figure 12) shows good agreement between the locations of simulated and detected lightning pulses. No regions with obvious accumulation of FTs can be observed. The LI `night` Level 2 detection threshold is 4 µJ / (m^2 sr) (see Figure 15). In the top panel of Figure 14, one can see the geographical variation of the average ADP. Here, the geographical coverage is much more uniform than in Figure 10, as the input pulses are more randomly scattered. The performance degradation with increasing distance from the centre of the LI FOV (the sub-satellite point, SSP) is obvious. The degradation of ADP and FDE as a function of distance from the SSP is further illustrated in Figure 16. Within 2000 km from the SSP, LI is characterized by ADP of approximately 50% and FDE is larger than 90%. ADP deteriorates faster than FDE with increasing distance from the SSP. At 5000 km (or 45° degrees viewing angle) ADP is about 30% whereas FDE is still about 80%. It is worth pointing out that FDE stays as high as 75% as far as 6000-7000 km from the SSP. At this point the ADP drops to about 20%. Beyond 7000 km, FDE suddenly drops to about 25% with ADP being in the order of 10%.



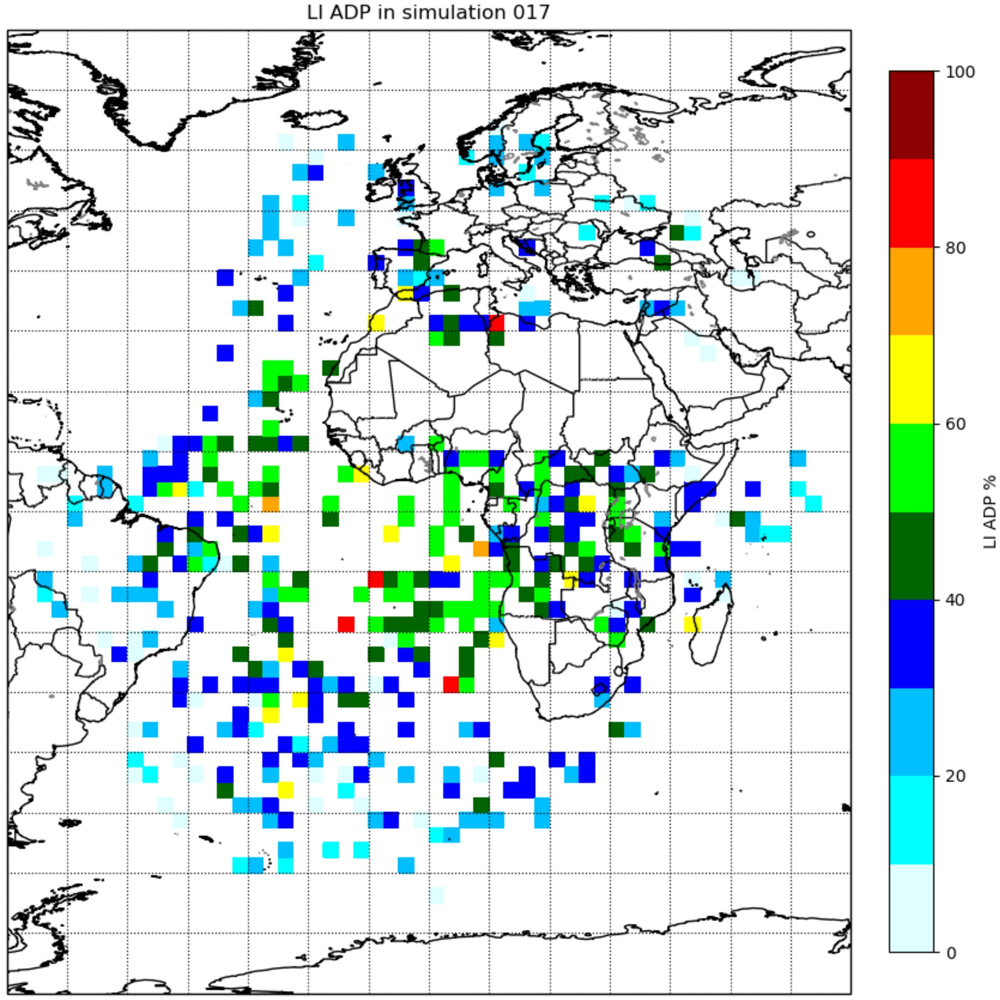


Figure 14. DTs and ADP for session `ID017 night`. *Same format of Figure 10.*

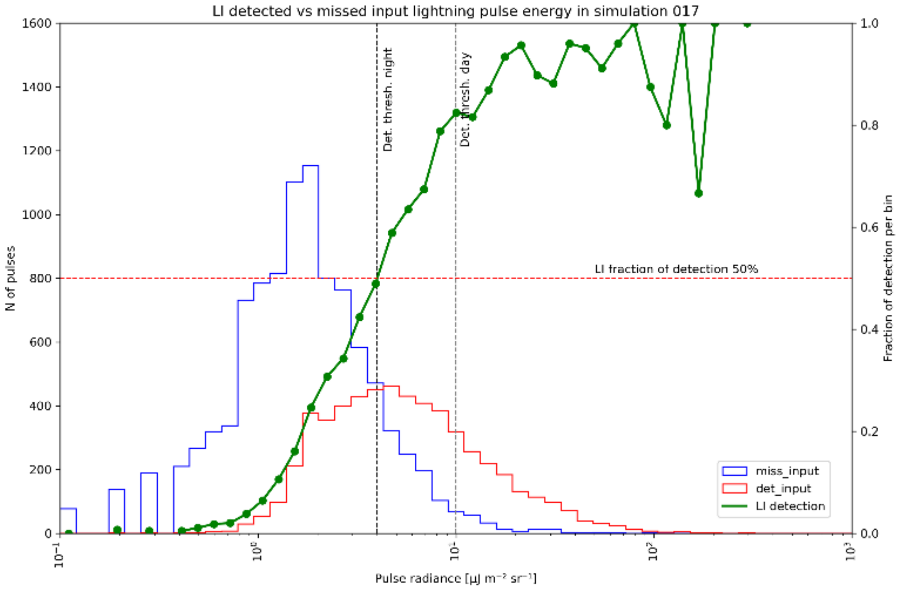
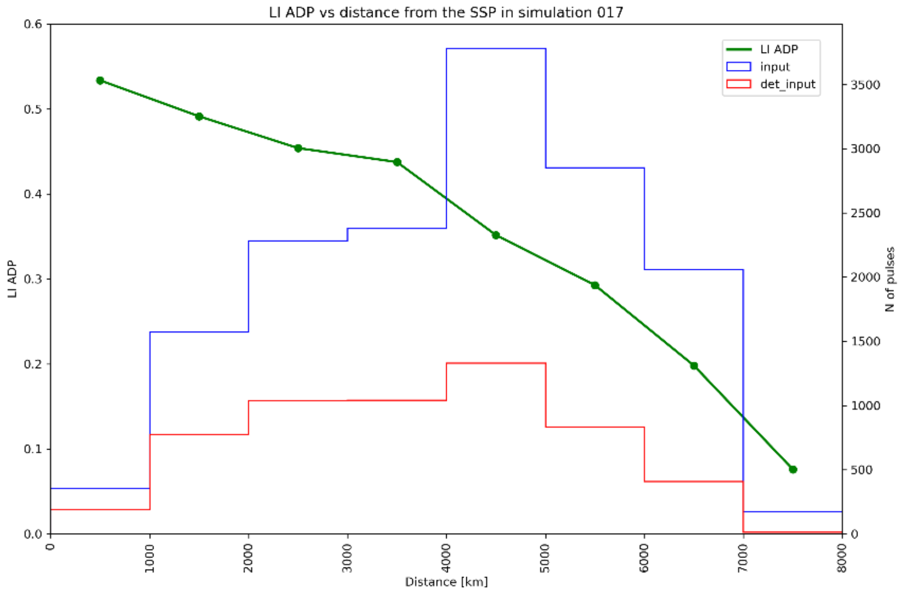


Figure 15. Detection threshold for session `ID017 night`. Same format as Figure 11.



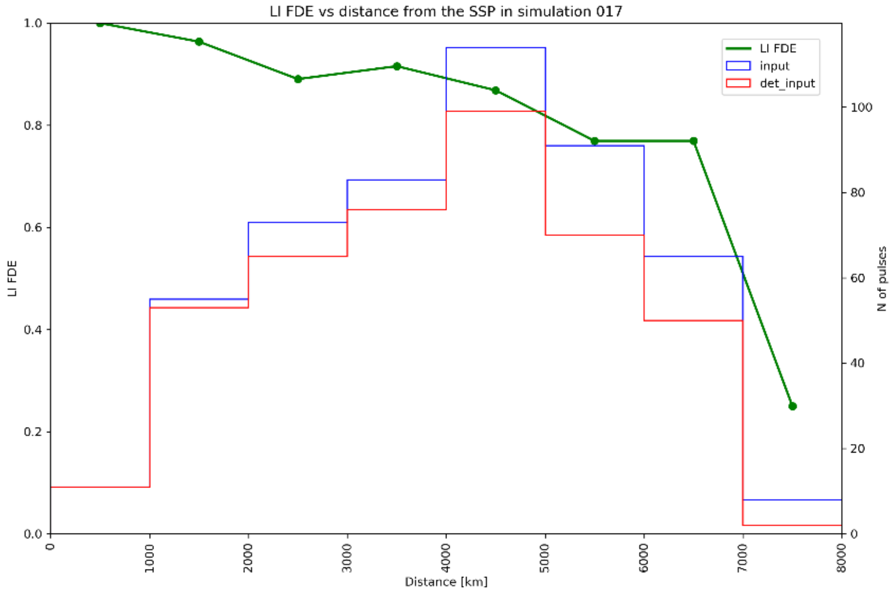


Figure 16. ADP (top panel) and FDE (bottom panel) as a function of distance from the center of the LI FOV (the sub-satellite point, SSP) at Level 2 for session `ID017 night`. The histograms represent the total number of input pulses (blue solid line) and the number of input pulses detected at Level 2 (red solid line) as a function of distance from the SSP (referred to the right y-axis). The green solid line and dots measure the fraction of detected pulses at Level 2 as a function of distance from the SSP (referred to the left y-axis). Note that the distance of 5000 km corresponds to the angular distance of 45 degrees.

4.3 Session `ID018 half`

4.3.1 Settings

Table 3. Settings for the `ID018 half` session

Session	Settings	
`ID018 half`	Level 0 simulator (see [LI-9])	Standard settings at BOL and no COM component of micro-vibration
	Level 1b prototype (see [LI-9])	Standard settings
	Level 2 filtering (see [LIL2ATBD])	OC1 (daylight) 1. Particle filter on Groups 2. Radiance filter, with 0.004 W / (m^2 sr) threshold, for at least half of DTs on Groups 3. Footprint filter at 3 DT on Flashes 4. Distance correlation between groups within 10 km on Flashes 5. Average Sobel gradient normalized to the minimum background with threshold at 10 on Flashes OC2 and OC4 (terminator) 1. Particle filter on Groups 2. Radiance filter, with 0.002 W / (m^2 sr) threshold, for at least half of DTs on Groups

		3. Footprint filter at 3 DT on Flashes 4. Distance correlation between groups within 10 km on Flashes 5. Average Sobel gradient normalized to the minimum background with threshold at 10 on Flashes OC3 (dark) 1. Particle filter on Groups 2. Radiance filter, with 0.002 W / (m^2 sr) threshold, for at least half of DTs on Groups 3. Number of Groups at 3 on Flashes 4. Footprint filter at 3 DT on Flashes
	Number of flashes	50 per simulation run
	Number of pulses	About 1500 per simulation run
	Pulse location	Derived from the Multi-Sensor Precipitation Rate Estimate Level 2 SEVIRI product
	Pulse properties	See Section 3.2.2
	Flash properties	See Section 3.2.1
	Maximum simulation duration (i.e., maximum flash duration)	2 sec
	Simulation runs	10

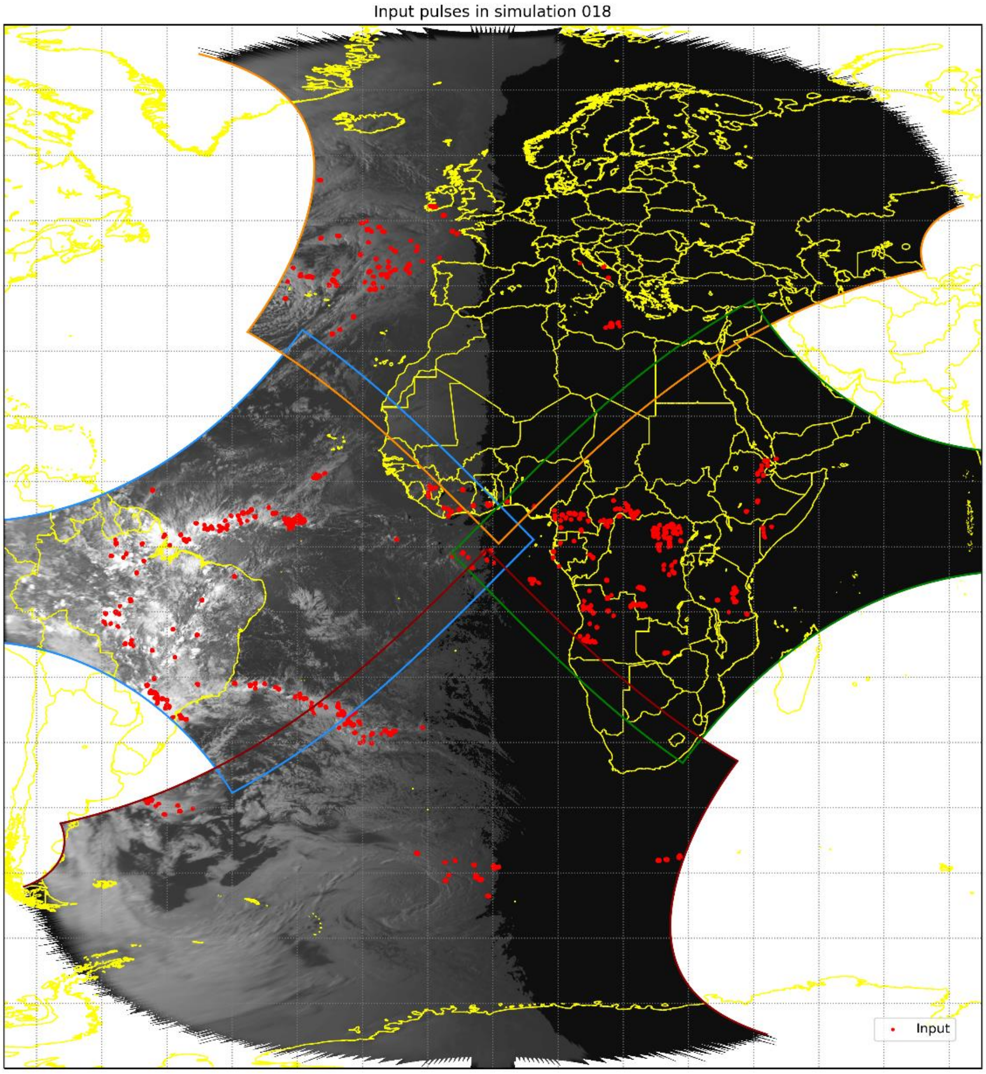


Figure 17. Location of the input pulses in the simulation `ID018 half`. ***Same format as Figure 8.***

4.3.2 Results

The average-over-10-simulations Level 1b ADP is of 27% while at Level 2 it is 23% (Figure 18 top panel). The impact of the Level 2 filtering on the average ADP (i.e., the reduction of the total number of TTs) is about 4%, which is higher than in simulations `ID016 day` and `ID017 night`. The average Flash Detection Efficiency (FDE) is 68% for an average Flash False Alarm Rate (FFAR) of 3-4 flashes per second per OC, i.e., about 15 false flashes per second over the whole LI field of view. Session `ID018 half` is characterized by significant differences between individual OCs, related to different illumination conditions from daylight to darkness (the effect was somewhat mitigated by applying different Level 2 filtering setting for different OCs, see 4.3.1). The average Level 1b ADP varies from 17% (OC1, daylight) to 41% (OC3, night) and Level 2 ADP from 14% (OC1) to 41% (OC3). The impact of Level 2 filtering on the average ADP is smallest on the dark side (less than 1% for OC3) and largest along the terminator (about 8% for OC4 and about 6% for OC2). OC3 (night) has the highest FDE of nearly 90% while OC1 (day) has FDE of 61%. Even lower FDE of 53% is characteristic to OC4 (terminator), possibly the result of much greater reduction of ADP in Level 2 filtering (about 8%) compared to OC1 (about 3%). The average Flash False Alarm Rate (FFAR) is much larger for OCs 1 and 3 (about 6-7 flashes per second) and very low for OCs 2 and 4 (less than 1 flash per second). This is in line with the significantly larger reduction of the total number of FTs through the Level 2 filtering steps for OCs 2 and 4.

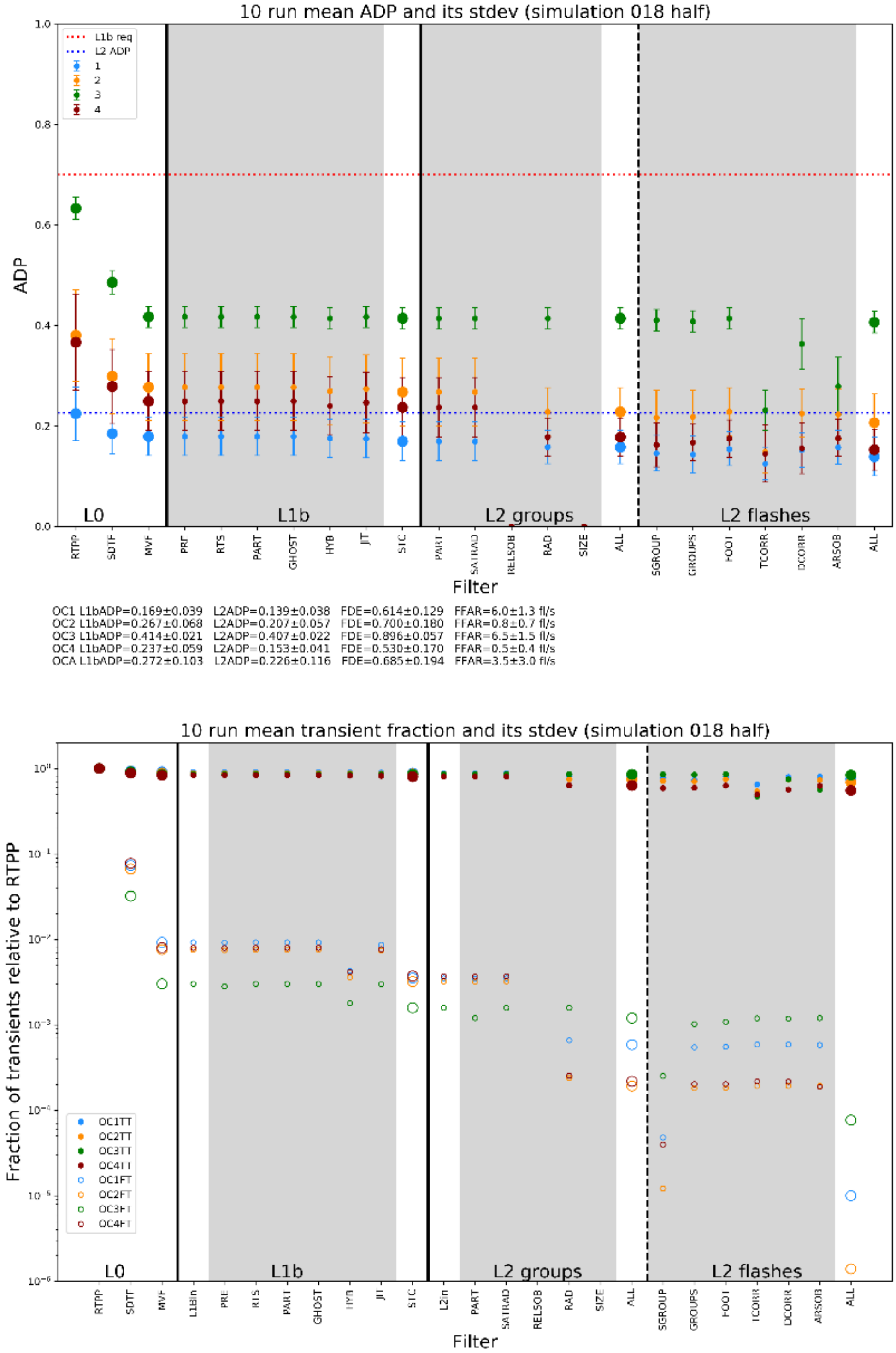
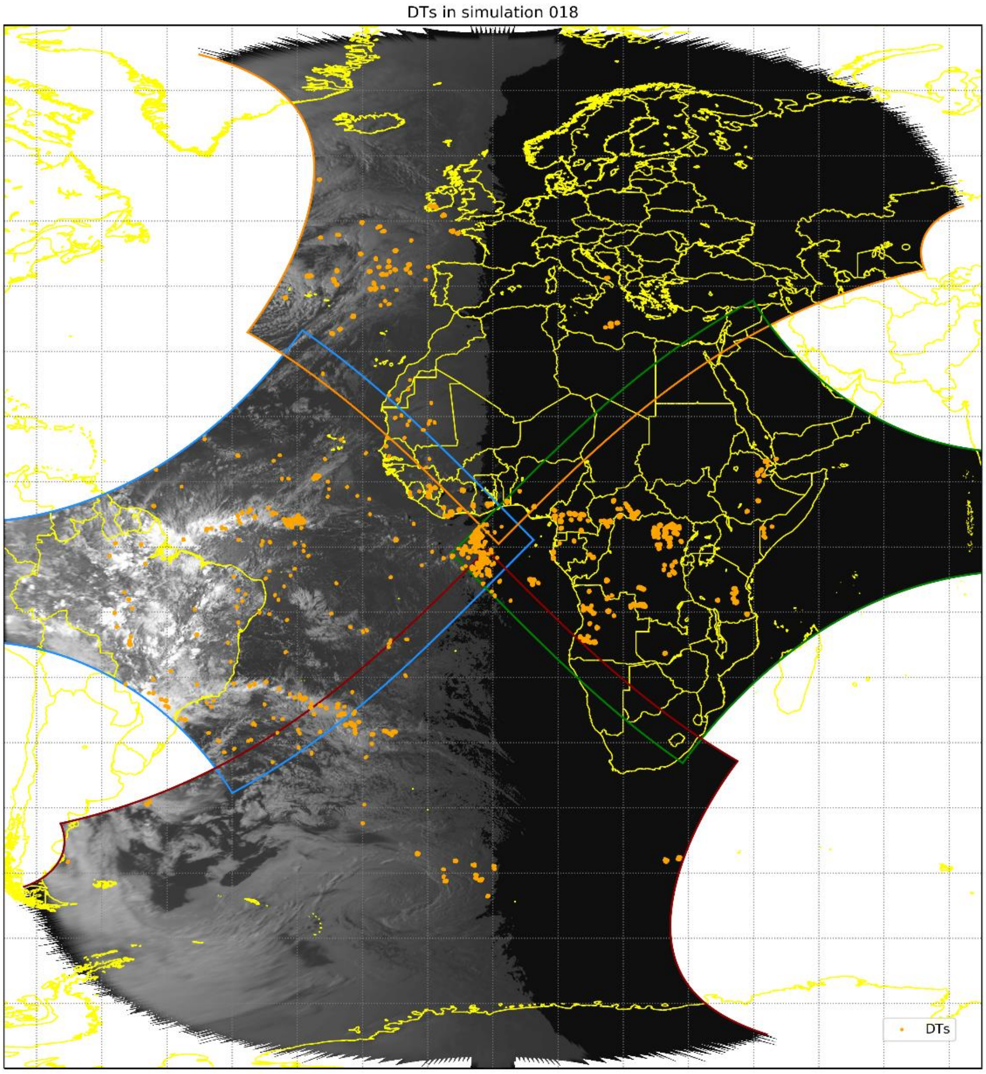


Figure 18. ADP and DT variation for session **`ID018 half`**. *Same format as Figure 9.*

The top panel of Figure 19 reveals a large cluster of FTs related to bright cloud tops over dark ocean surface near the terminator in the westernmost corner of OC3. The bottom panel of Figure 19 clearly illustrates the day-night difference in LI detection with higher ADP values on the dark side to the east of the terminator. The LI half-scene Level 2 detection threshold (the energy at which the fraction of detections reaches 50%, see Figure 20) is 6.5 µJ / (m^2 sr) (see Figure 20).



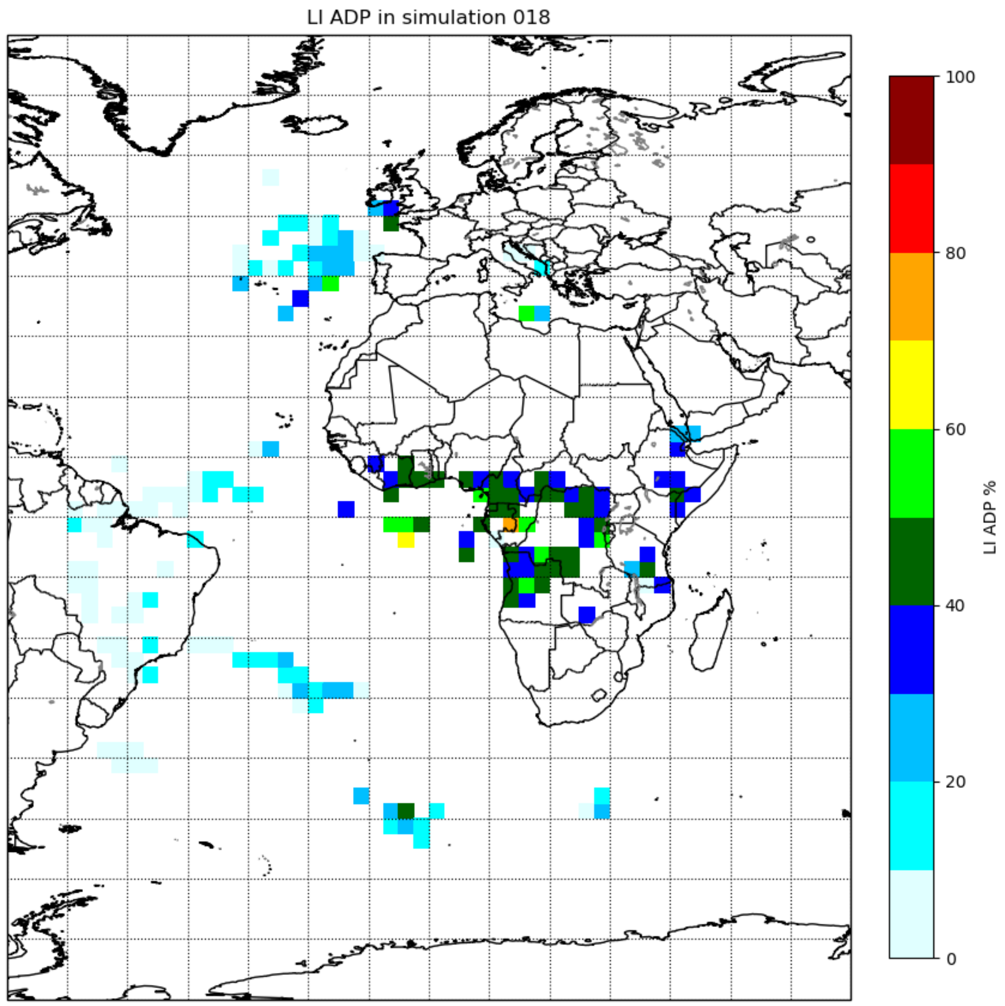


Figure 19. DTs and ADP for session `ID018 half`*. Same format as Figure 10.*

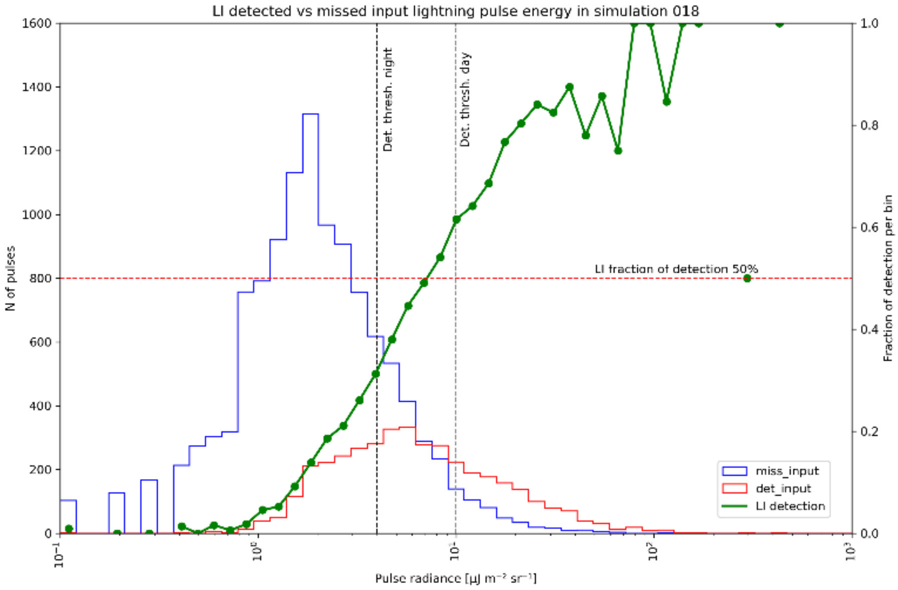


Figure 20. Detection threshold for session `ID018 half`*. Same format as Figure 11.*

5 DISCUSSION

The aim of the analysis presented here is to assess the expected Level 2 performances of LI by using as realistic lightning pulses as possible as detection targets. This is meant to complement the performance assessment undertaken by industry and published in [LI-29]. In the analysis of industry, so called "engineering" pulses with fixed size, fixed duration, and with radiances proportional to background scene are used as inputs to the Level 0 and Level 1b simulator to measure the Level 1b performances (i.e., ADP and FAR, see [LI-9]). The engineering pulses are the pulses that drive the design of the instrument, and are formally used to measure the expected performances based the latest instrument characterization.

It is very important to stress that the design of LI was very much inspired by the LIS instrument. This implies being bonded to LIS sensitivity and performances, i.e., to the knowledge of the lightning phenomena as derived from LIS, in particular for the description of the pulse radiance. This pulse property affects the most the performances of a lightning imager. At the low-end of the pulse radiance distribution of LIS, one finds that the most probable detected radiance is 5 µJ / (m^2 sr) (see Figure 2 upper plot). The minimum required detectable energy of the LI has been set at 4 µJ / (m^2 sr), right below such value. Recent observations done with very sensitive instruments such as the FEGS show that LIS is providing us with a partial description of pulse radiances: below 5 µJ / (m^2 sr), FEGS detected about 8 times more pulses than LIS (see Figure 2 top panel). FEGS is also providing us with the same distribution as LIS above such radiance value (see Figure 2 bottom panel). This means that LIS is capturing the high-end of the pulse radiance distribution known today, and so will LI. In the present analysis, we assess the Level 2 performances of LI with pulses properties that are modelled using FEGS measurements, i.e., the state of the art of the description of the optical properties of lightning pulses observed from space. In fact, in addition to the information on the pulse radiance, FEGS provides one with detailed information about the pulse temporal evolution. Through this, it was possible to relate, in our modelling, the pulse (total) radiance, the pulse peak radiance, and the pulse duration (see Section 3.2.2.2), in a consistent manner. This was possible through the assumption of having a pulse temporal evolution (or temporal profile) described by a Maxwell function. This assumption was also corroborated by means of FEGS measurements[5]. The pulse duration discussed in Section 3.2.2.2 and represented in Figure 7, provides pulses that are systematically shorter than the ones derived from FEGS when evaluating the pulse duration from two times between which the pulse radiance can be identified above the instrument noise. However, by referring to the different distributions of pulse duration available in [FEGS19], one can see that the distribution of pulse duration derived between the 10% levels of the pulse profiles (i.e., 10-10 case in [FEGS19]) is very close to the one employed in our simulations. The remaining property of the pulses used in input, i.e., the pulse radius, has been derived from LIS data since it is known that, due to its small FOV, FEGS is not always capable of capturing the entire optical pulse. On this specific property, it is worth stressing that recent result from FEGS (see [FEGS19]) and [ZHANG19] are in apparent contradiction. In fact, from the exercise of navigating FEGS pulses against matching ground network detection one learns that the typical size of optical pulses is of the order of 20 km in radius, with a spatial variation of the radiance close to a Gaussian profile. In these profiles, the core (i.e., the brightest part of the pulse) has typical size of 10 km of radius. [ZHANG19] put forward a very different scenario in which pulses smaller than the LIS pixel (i.e., 4 km × 4 km) are used to explain the distribution of the

[5] The only exception being long pulses (longer than 2 ms in duration) that, as suggested by FEGS experts, generally tend to have multiple optical peaks. These pulses represent anyway a small fraction of the pulses observed by FEGS.

number of events per group derived from LIS. Our choice of employing the number of events per group of LIS to derive the linear size of a pulse is closer to the picture presented in [FEGS19].

On the flash properties, it is worth stressing that the combination of flash properties derived from LIS (number of pulses per flash and time difference between pulses in flashes) and the European Ebro LMA network (flash size) provide us with a robust overall picture of flash properties. The distribution of the flash duration represented in Figure 4 presents two important differences with respect to the LIS distribution (here employed for a sanity check): *i)* flashes are systematically longer than the LIS flashes, and *ii)* there is a peak in flash duration at 2 sec. The latter can be explained by recalling that a flash truncation is done at 2 sec, this means that the bin at 2 sec is actually representative of the flashes that would have had duration larger than or equal to 2 sec. The former is consistent with the fact that the duration of LIS flashes is known to be systematically shorter than the real flash duration, for example derived by LMA networks (see [UPC19]). Finally, the temporal evolution of the flash properties recently highlighted in [ZHANG19] could play an important role in the LI performance assessment analysis (as proven for GLM). In detail, the fact that flashes longer than a certain time are (in average) composed by pulses with radiances above a certain radiance level, would actually ensure their detection. The modelling of the flashes presented in Section 3.2.1 does not account for such property of flashes. Further investigation into the possibility of including the temporal evolution of flash properties in the input modelling is due.

The combination of the selection of flash locations based on the Multi-Sensor Precipitation Rate Estimate (see Figure 8) and the realistic pulse energy distribution (see above and Figure 6) causes an important reduction of the ADP with respect to the typical values computed by industry (above 70% for any illumination condition, see [LI-29]). From Figure 6, one learns that only 10% of the pulses have a pulse energy above the minimum detectable energy during the day (10 µJ / (m^2 sr)), while about 30% of the pulses are above the minimum detectable energy during the night (4 µJ / (m^2 sr)). The selection criterion for the pulse locations is picking almost exclusively very bright scenes for the `day` scenario; the ADP is actually reflecting the fraction of pulses with energies above the minimum detectable energy for the day (see Figure 9). For the `night` scenario, the selection criterion for the pulse locations is picking only very dark scenes; the ADP is then reflecting the fraction of pulses with energies above the minimum detectable energy for the night (see Figure 13). The ADP for the `half` scenario varies between the `night` performances (for OC3) and the `day` performances (for OC1), respectively (see Figure 18). Slightly higher ADP was observed for OC1, compared to the same OC in the `day` case, namely, 0.14 against 0.09. This is due to the fact that the illumination levels of OC1 in the `half` scenario are not as high as in the `day` scenario, especially in the areas close to the SSP where the LI is generally more sensitive. For the `day` and `night` scenarios, the reduction of ADP from Level 1b to Level 2 is very small, only 1-2%. This is quite close to the target to have a flat ADP curve after Level 1b. This is critical, as flat ADP curve would confirm that no true transients (TTs) are filtered out at Level 2. For the best Level 2 output ADP (and FDE), it is required to keep as many TTs as possible. It is worth noting that for OC2 and OC4 of the `half` scenario, the ADP reduction through the Level 2 is notably higher: 6-8%. This is because within these OCs there is a variation of illumination conditions between day and night while the filtering settings are the same everywhere. The higher ADP reductions suggests that the current approach of OC-based group and flash filtering settings is too coarse and works well only in fully illuminated or dark conditions while it does not address the challenges arising from changing illuminations conditions within the OC. The proposed approach for dealing with varying illumination conditions is to use look-up tables (LUTs). Such tables would contain the values of Level 2 filtering parameters as a function of the Sun Zenith Angle (SZA). LUTs would allow one to tune the Level 2 processing at fine spatial resolution and achieve best performance. In addition to tuning the thresholds of individual filters, it is also important to determine the optimal configuration of the overall filtering procedure at Level 2 groups and flashes. The results indicate that not all the proposed Level 2 filters are needed for effective removal of FTs. Furthermore, some of the filters can be counter-productive by rejecting a lot of TTs, e.g. the flash level time correlation and Sobel gradient filters have detrimental impact on ADP in dark conditions (see Figure 13 and OC3 in Figure 18). At the same time, the flexibility of Level 2 filtering effectively prevents any serious issues, as the problematic filters are easy to exclude from the final TT/FT classification and doing so results in the optimal Level 2 performance with efficient FT removal. It needs to be decided in the future if the problematic filters should be removed or if they can be improved (either by better tuning, or changes in the overall method/code).

The Flash Detection Efficiency (FDE) is still relatively high, approximately 56%, for the `day` scenario (Figure 9 top panel). This clearly indicates that although most of the pulses on very bright clouds are missed (low ADP), it is still often feasible to detect at least two pulses per flash[6]. Most of such pulses are expected to belong to the high end of modelled pulse energy distribution. For the `half` and `night` scenarios, the FDE is 69% and 88%, respectively (see the top panels of Figure 13 and Figure 18). As the background gets darker, weaker pulses are detected more efficiently and the probability of detecting at least two pulses per flash increases. The impact of viewing geometry on ADP/FDE is obvious in all three scenarios with the best LI performance near the SSP and the lowest performance at high viewing angles (Figure 14 bottom panel and Figure 16). In general, the findings are encouraging, showing larger than 75% FDE even as far as 6000-7000 km for the SSP for the `night` scenario. Moreover, most flashes are detected even if ADP is only 20-25%, indicating that the brightest pulses of most flashes are still detected. A notable feature that needs further investigation is the sudden drop in FDE beyond 7000 km from SSP. In principle, it can be either a random outcome of the particular scenario (possibly due to smaller sample size beyond 7000 km) or a persistent feature related to the limitations of LI at very large viewing angles. The almost linear decrease of ADP to below 10% beyond 7000 km supports the latter. It is likely that there is a 'breaking point' ADP value for FDE. If ADP is higher than that then it is likely to detect at least two pulses per flash and FDE stays high. However, if ADP drops below the breaking point value then only the brightest pulses of brightest flashes are still detected and FDE drops suddenly. Further investigation is due, as this can be a significant limitation for the northernmost member states (e.g. central Finland and Scandinavia are ~7500 km from the SSP).

It is important to remember that our current knowledge on the ADP/FDE-viewing angle relationship bases mostly on the `night` scenario where flashes were sufficiently scattered across the visible disk for reliable analysis. The same principle was observed for the `day` and `half` scenarios, but here the results were also affected by other factors such as locations and variable background brightness of different precipitating cloud systems. For a more reliable daytime study, a new test scenario with full-disk uniform cloud cover is proposed. Although such conditions are not realistic, the test scenario would allow to distribute pulses all over the visible disk and produce a more representative ADP/FDE-viewing angle relationship curve.

Another important limitation of the test scenario is that flashes area assumed to be circular. When moving towards the edges of the LI FOV, the apparent shape of the circle gets more and more distorted due to increasing viewing angle, making it harder to detect. In reality, flashes can be significantly elongated, especially in large quasi-linear convective systems where they

[6] Single-group flashes are false

can be even hundreds of kilometres long. The impact of LI viewing angle on the detection of such flashes depends on their location and orientation.

The lightning pulse geometry model also assumes optical emission from flat cloud top surface. This is more or less true for the anvils of large mature convective systems at lower viewing angles. However, near the edges of LI FOV, the sides of thunderstorms are often visible, if not hidden by large anvils or other high level clouds. It is thus likely to detect some optical emission from cloud sides. This will result in more detected lightning energy and lead to higher ADP and FDE at high viewing angles, compared to circular flash model predictions. It is difficult to assess how much this effect will mitigate the drop in FDE near the edges of the LI FOV. The overall improvement might be small as the effect is limited to the areas where cloud sides are visible. Furthermore, it can be accompanied by some unwanted side effects, such as taller thunderstorms hiding smaller storms behind them.

The reduction of the total number of FTs through the Level 2 filtering is of about three orders of magnitude during the day (Figure 9 bottom panel) and two orders of magnitude at night (Figure 13 bottom panel). Most of the FTs that still pass the Level 2 filtering are located in bright bands of fragmented clouds over the oceans for the `day` scenario (Figure 10 top panel). This suggests that most of them originate from the combination of micro-vibration and sharp background gradient. It is worth mentioning that the background was chosen as challenging as possible, i.e. containing a lot of small and fragmented clouds that can trigger FTs through micro-vibration. In the `half` scenario (Figure 19 top panel), significant accumulation of FTs is observed in the area where the otherwise dark OC3 (with night-like filtering settings) touches brighter cloud tops around the terminator. This can be mitigated by determining the values of Level 2 filtering parameters in a finer spatial grid as discussed above. It should be noted that for the LI, the maximum acceptable FFAR is not explicitly stated. The analysed scenarios generally revealed a few False Flashes per second per OC. This seems to be acceptable, given that true lightning is effectively detected.

In general, the ADP/FDE/FFAR results confirm that the Level 2 algorithms perform well with realistic lightning pulses. This is an important step forward as the Level 2 group and flash level filters had been previously tested only with large and bright test pulses that are easier to detect and distinguish from FTs. Having demonstrated no major issues with processing weaker realistic lightning pulses on bright background allows us to focus on finer tuning in the framework of the existing algorithms.

The results of the present analysis allow for some important comparisons of the expected performance of the LI against the performance of the GLM. The LI FDE is similar to GLM: LI detected 56 ± 18% of flashes in the day scenario and 61% ± 13% for OC1 of the half scenario; GLM detected 61% (range 51-85%) of FEGS flashes between 6am and 6pm local time observations [FEGS19]. GLM pulse detection efficiency relative to FEGS is 29% while LI detected 9 ± 3% of pulses for the day scenario and 14 ± 4% for OC1 of the half scenario. The LI day scenario detection threshold is 14.3 µJ / (m^2 sr), while GLM's is 10 µJ / (m^2 sr). These differences are likely related to the limitations of the FEGS dataset that contains only 6 daytime storms in the US in the time period 6am – 6pm local time. This implies a bias towards relatively low GLM viewing angles compared to the present analysis where the LI test pulses were distributed randomly and as far as to the edges of the FOV. Higher viewing angles make pulse detection more difficult and lead to a relative bias towards lower LI ADP and FDE values (see Figure 16), compared to GLM in the US. It is easy to demonstrate that limiting the viewing angle to values comparable to the GLM case studies in the US results in very similar FDE values. For example, a recent assessment of GLM performances against the Kennedy Space Flight Center (Florida) LMA revealed that GLM FDE was characterized by strong storm-to-storm variability with an average of 74% over 24 hours [ZHANG19]. The study area is located

~3000 km from the SSP. In the similar distance from the SSP, the LI detected approximately 60% of flashes during the day, 90% at night (Figure 16), i.e., about 75% as an average over 24 hours. Another important factor to consider is the selection criterion for test pulse locations that is picking almost exclusively very bright scenes for the day scenario, which makes faint pulses difficult to detect. Despite a somewhat lower ADP, LI FDE is still comparable to GLM flash detection efficiency. This indicates that despite GLM detecting more pulses per flash, the LI still detects enough pulses to identify a large fraction of the input flashes. Regarding the relatively high LI FFAR of up to 24 flashes per second, it has to be considered that GLM operated with quite high FFAR (up to 20%) for some time. Finer tuning of the LI Level 2 filters, taking into account the local variations in the illumination condition, offers a way to remove more FTs and reduce the FFAR. Moreover, the daytime scenario of the current analysis represents the worst case in terms of FT generation, i.e. most of the time the amount of sharp cloud edges in the LI FOV will be smaller, leading to less micro-vibration related FTs.

6 CONCLUSIONS

The assessment of the LI Level 2 expected performances has been performed by means of:

a. The realistic LI instrument model and level 1b performance based on measured data, thanks to industry and ESA.
b. The up-to-date EUMETSAT LI Reference Processor, which is the most realistic representation currently available of the LI detection and filtering end-to-end chain.
c. The use of realistic modelling of pulses and flashes as input to the simulations. The inputs have been defined exploiting state-of-the-art data characterizing optical pulses from lightning detected from space (FEGS and LIS) and from the ground (Ebro LMA network). In particular, the use of the pulse radiance from FEGS allowed us to account for the dominating component of lightning activity with pulse radiance below the minimum detectable energy of LI at night.
d. Three different illumination scenarios: full day (named `day`), full night (named `night`) and half-illuminated Earth (named `half`). These are a rough representation of the different illumination conditions that LI will be observing over 24h.
e. The up-to-date Level 2 filtering settings, which account for the change of illumination conditions between the different OCs for the three scenarios in d.

Each illumination scenario is associated to a session, which involves 10 simulation runs for the computation of average key performance indicators: Average Detection Probability (also known as pulse Detection Efficiency), Flash Detection Efficiency, Flash False Alarm Rate and detection threshold. All three sessions had 500 input flashes and approximately 15 thousand input pulses. The flashes were randomly placed in the regions with precipitation, derived from SEVIRI Multi-Sensor Precipitation Rate Estimate Level 2 product for the `day` and `half`. In one case, namely the `night` scenario, the SEVIRI Cloud Mask Level 2 product was employed. The LI Level 2 expected performances are provided, in a compact fashion, in Table 4.

Table 4. The main MTG LI expected performance characteristics, averaged over the whole FOV and 10 simulation runs of each session.

Session	Level 2 ADP	Level 2 FDE	Level 2 FFAR	Level 2 detection threshold
`016 day`	0.09 ± 0.03	0.56 ± 0.18	6 ± 4 1/(sec OC)	14.3 µJ / (sr m^2)
`017 night`	0.36 ± 0.05	0.88 ± 0.10	0 ± 0 1/(sec OC)	4 µJ / (sr m^2)
`018 half`	0.23 ± 0.12	0.69 ± 0.19	4 ± 3 1/(sec OC)	6.5 µJ / (sr m^2)

Based on the results, the following main conclusions can be drawn:

- Despite of most input lightning pulses having radiances below the LI detection threshold, LI is capable of detecting 56-88% of all input flashes at Level 2 (at least 2 pulses must be detected to detect a flash).
- The reduction of the total number of FTs through the Level 2 filtering is of about three orders of magnitude during the day and two orders of magnitude at night with the final Level 2 FFAR being in the order of a few false flashes per second per OC in the illuminated conditions and close to 0 in darkness.
- The LI Level 2 detection threshold (the energy at which the fraction of detected pulses at Level 2 reaches 50% of the input pulses) varies from 4 µJ / (sr m^2) at night to 14.3 µJ / (sr m^2) during the day.
- For the night scenario, LI ADP decreases almost linearly with increasing distance from the sub-satellite point (SSP) from about 55% around the SSP to about 10% near the edges of the FOV (7000-8000 km from the SSP). At the same time, LI FDE stays above 75% within 7000 km from the SSP and only drops to about 25% in the very edges of the FOV.

- A fairly conservative estimate of the LI Level 2 performances is in line with key GLM performance indicators measured over the US.