# Algorithmic Chain for Lightning Detection and False Event Filtering Based on the MTG Lightning Imager

Pierre Kokou, Philip Willemsen, Mounir Lekouara, Madani Arioua, Andreu Mora, Pieter Van den Braembussche, Emanuele Neri, and Donny M. A. Aminou

***Abstract*— Meteosat Third Generation (MTG) is the next generation of European meteorological geostationary satellites, set to be launched in 2021. Besides ensuring continuity with Meteosat Second Generation imagery mission, the new series will feature new instruments, such as the Lightning Imager (LI), a high-speed optical detector providing near real-time lightning detection capabilities over Europe and Africa. The instrument will register events on pixels, where a lightning pulse generates a transient in the acquired radiance. In parallel, signal variations due to a number of unwanted sources, e.g., acquisition noise or jitter movement, are expected to produce false events. The challenge for on-board and on-ground processing is, thus, to discard as many false events as possible while keeping a majority of the true lightning events. This paper discusses a chain of algorithms that can be used by the LI for the detection of lightning and for the filtering of false events. Some of these algorithms have been developed in the framework of internal research and simulations conducted by the MTG team at the European Space Agency on an in-house LI simulator and therefore will not necessarily reflect the ultimate operational processing chain. The application of the chain on a representative scenario shows that 99.5% of the false events can be eliminated while keeping 83.6% of the true events, before generating the LI higher level data products. Machine learning techniques have also been studied as an alternative for on-ground event processing, and preliminary results indicate promising potential.**

***Index Terms*— Filtering, jitter, lightning, machine learning, Meteosat Third Generation (MTG), Satellite Meteorology, transient detection.**

## I. INTRODUCTION

EUROPE's next fleet of meteorological geostationary satellites, Meteosat Third Generation (MTG), is set to debut from 2021 [1]. The new series, realized through a cooperation between EUMETSAT and the European Space Agency (ESA), will comprise six spacecrafts: four MTG-I (imaging) and two MTG-S (sounding) satellites.

In addition to ensuring continuity with the current Meteosat satellite family, MTG-I satellites will fly the Lightning Imager (LI), an instrument performing full disk observations of lightning from the geostationary orbit (GEO). The LI products will, amongst other purposes, be used to provide near real-time monitoring and short-range forecast of severe weather phenomena.

The LI will complement other systems dedicated to the observation of lightning activity. Several on-ground systems, such as ATDnet [2], make use of a network of sensor stations detecting the electromagnetic field generated by a cloud-to-ground discharge and apply arrival time difference calculations to determine its location.

Optical lightning detection from space, which consists in measuring top-of-atmosphere radiances, started as early as 1995 with the optical transient detector (OTD) [3]. The OTD was the precursor instrument of the Lightning Imaging Sensor (LIS) [4] launched in 1997 and whose mission ended in 2015. The LIS, on which the new generation of GEO lightning sensors is based, is having its original flight spare installed on the International Space Station (ISS) in 2017 [5]. In the future, near global lightning detection coverage from GEO will be ensured through the Geostationary Operational Environmental Satellite R-series Geostationary Lightning Mapper (GLM) [6] centered on the Americas, the Feng-Yun-4 Geostationary Lightning Imager [7] over Asia–Oceania territory and, as discussed in this paper, the MTG LI covering Europe and Africa.

The fundamental working principle behind the LI lightning detection consists in triggering events on pixels where and when a sudden increase of energy is measured over the background radiance image. In addition to lightning pulses, other sources, such as acquisition noise, jitter movements of the instrument line of sight or high-energy particles will induce signal variations that will trigger events. A succession of on-board and on-ground filters is therefore needed to discard the false events while keeping the true lightning information.

In this paper, we present algorithms for the data processing of events generated by the MTG LI. Calibration and geolocation aspects are excluded from the scope of this paper. The ideas reflected here are the result of internal research and simulations conducted by the ESA MTG team on an LI simulator developed in-house. Therefore, the presented algorithms will

Manuscript received June 1, 2017; revised November 10, 2017 and December 20, 2017; accepted February 8, 2018. Date of publication April 20, 2018; date of current version August 27, 2018. This work was supported by the European Space Agency. *(Corresponding author: Pierre Kokou.)*

P. Kokou is with Aurora Technology B.V., European Space Agency, European Space Research and Technology Centre, 2201 AZ Noordwijk, The Netherlands (e-mail: pierre.kokou@esa.int).

P. Willemsen, P. Van den Braembussche, E. Neri, and D. M. A. Aminou are with the European Space Agency, European Space Research and Technology Centre, 2201 AZ Noordwijk, The Netherlands.

M. Lekouara was with the European Space Agency. He is now with EUMETSAT, 64295 Darmstadt, Germany.

M. Arioua was with the European Space Agency, European Space Research and Technology Centre, 2201 AZ Noordwijk, The Netherlands.

A. Mora is with GMV, European Space Agency, European Space Research and Technology Centre, 2201 AZ Noordwijk, The Netherlands.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2018.2808965

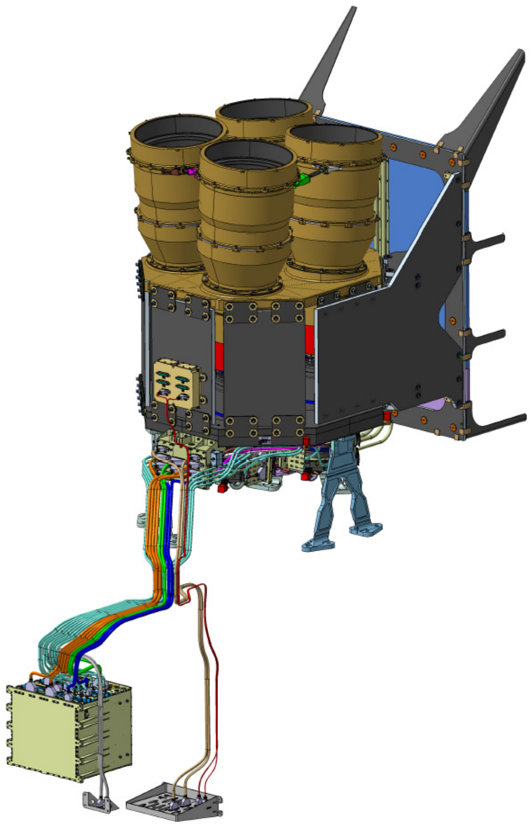

Fig. 1. LI configuration with the four optical heads.

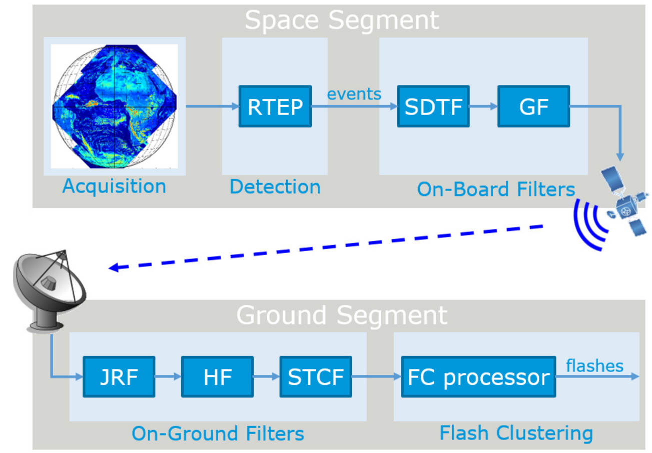

Fig. 2. LI algorithmic chain as proposed in this paper. An image is first acquired and fed to the detection algorithm, based on which events are generated. The events are successively processed by on-board filters, transmitted to the ground segment, and fed to on-ground filters and to a flash clustering processor that will aggregate them to produce the flash products. See Sections III and IV for a detailed description of the different blocks and acronyms.

not necessarily be those which will be used for the operational on-board and on-ground LI data processing chain.

The remainder of this paper is organized as follows. Section II describes the main features of the LI and the proposed algorithmic chain. The process of triggering lightning events from the LI acquisitions is covered in Section III. Section IV discusses the different sources of false events and presents a number of suitable filters for on-board and on-ground processing. A performance analysis is shown on a representative simulation scenario. Section V explains how machine learning concepts, such as decision trees and support vector machines (SVMs), could be used for improved on-ground filtering. Finally, concluding remarks are drawn in Section VI.

## II. MTG LIGHTNING IMAGER

The LI will be mounted on-board the four MTG-I satellites [8]. These three-axis stabilized satellites have a mass of about 3600 kg and are designed to operate for 8.5 years on the geostationary arc. Their main payload will be an imaging instrument, the Flexible Combined Imager, which will provide images of the Earth in 16 spectral channels between 0.44 and 13.3 $\mu$m [9].

The LI will measure radiances at cloud top and provide a near real-time lightning (in-cloud and cloud-to-ground) detection and location service. The instrument comprises four optical heads in an envelope of $718 \times 1200 \times 1456$ mm$^3$, as illustrated in Fig. 1, weighs about 100 kg and has an allocated bandwidth of 30 Mb/s toward the MTG platform. The instrument design and the hardware on which the on-board processing algorithms run are described in detail in [10].

In order to establish the MTG LI specifications, statistical studies have been conducted on the data sets of lightning pulses detected by the OTD, LIS, and Fast On-orbit Recording of Transient Events [11] instruments. A lightning pulse is produced by an electric discharge within or below a cloud. The strongest emission features are produced by a triplet of neutral oxygen lines in the near infrared [12]. The released photons are transported to the cloud surfaces by scattering. The resulting lightning optical signal at the top of atmosphere can be described as a transient phenomenon with an average duration of 0.6 ms and an average footprint of 100 km$^2$ [13], [14].

The LI aims to measure lightning pulses with radiances as low as 7 mW/(m$^2$sr) during night conditions and 17 mW/(m$^2$sr) in day conditions on top of clouds. The instrument has been designed as a high-speed (acquisition frequency of 1000 Hz) event detector operating in a 1.9-nm-wide spectral window centered on 777.4 nm. The spatial sampling is 4.5 km at subsatellite point and the LI will contain 4.7 million pixels, divided into the four detectors, one for each of the separate optical heads.

Because of the high acquisition frequency, the high number of pixels, and the limitations on the data rate to the ground station, the data processing of the LI differs from that of most imagers. Fig. 2 shows a diagram of the on-board and on-ground processing chain, considered in this paper, ignoring calibration and geolocation aspects. At each time frame of 1 ms, the radiance is acquired and lightning events are triggered where the detector reading has exceeded a certain threshold. The information related to the events is then successively processed by on-board filters, transmitted to the ground, and further processed by on-ground filters. The goal of these filters is to eliminate as many false events as

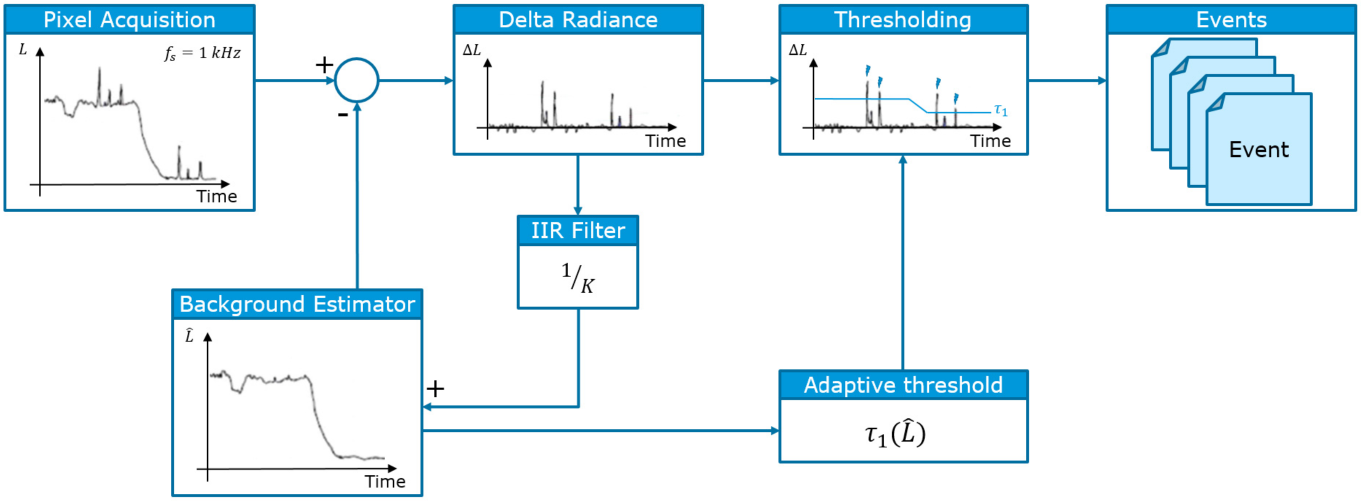

Fig. 3. RTEP. This detection algorithm monitors, on each pixel, the signal acquired at a frequency $f_s$ of 1 kHz. For each integration period, the delta radiance is calculated by subtracting the background estimate from the pixel signal. An event is triggered at each time step for each pixel, where the delta radiance is higher than an adaptive threshold $\tau_1$, which is a function of the background estimate. The delta radiance is also used to update the background estimate via an IIR filter. Adapted from [17].

possible while keeping a high number of true events so as to be able to detect, at the end of the chain, as many lightning pulses as possible. Without these filters, the thresholds used in the detection stage should be raised to limit the number of false events, hence reducing the sensitivity of the instrument to lightning pulses of weak energy.

The pixel-level events that have not been discarded at that stage are then fed into a flash clustering (FC) processor. The principle of this processor consists in aggregating adjacent events occurring during the same time frame into groups, and in further consolidating groups that are in close temporal and spatial proximity to each other to form flashes [15]. The flashes are the primary science data product for most users.

## III. LIGHTNING DETECTION

From the point of view of the instrument, a lightning event will appear as a transient signal, i.e., a sudden increase in the measured energy, over a background slowly varying due to the evolution of sun illumination and cloud movements. The role of the real-time event processor (RTEP) [16] is to read, for each pixel, the signal acquired during the integration frame and to trigger lightning events based on this information. The working principle of this algorithm is illustrated in Fig. 3.

After acquiring the signal $L$ collected on the pixel during the integration frame, the latest estimate of the background $\hat{L}$ is subtracted to determine the *delta radiance* $\Delta L$, which is defined as the signal that could have been caused by a lightning pulse

$$\Delta L[n] = L[n] - \hat{L}[n]$$

where $n$ is the time frame number.

This delta radiance is then compared to an adaptive threshold $\tau_1$, which is a function of the estimated background at that time frame. For every integration period during which the threshold is exceeded, an event is triggered

$$\text{if } \Delta L[n] > \tau_1(\hat{L}[n]) \rightarrow \text{trigger event.}$$

The determination of the proper threshold level is a trade-off between being sufficiently low to trigger true events from weak lightning pulses and being sufficiently high not to trigger too many false events due to noise sources. The adaptation of the threshold to the estimated background ensures a constant expected false event rate due to acquisition noise when the background scene illumination varies. The brighter the background is (e.g., during the day), the higher the noise will be. Consequently, at night, the threshold can be lowered, which allows the LI to be able to detect weaker lightning pulses.

The background estimate is updated after each integration period via an infinite-impulse response (IIR) filter. The update law can be tuned with the parameter $K$ to set the rate at which the estimator should follow variations in the signal:

$$\hat{L}[n+1] = \hat{L}[n] + \frac{\Delta L[n]}{K}.$$

Once triggered, the events are passed to the successive onboard filters. For each event, the information transmitted by the RTEP contains the coordinates of the triggering pixel, the time frame number during which it happened, the detector readings, and the background estimate for a $3 \times 3$-pixel window centered on the triggering pixel.

## IV. FALSE EVENT FILTERING

Other sources than lightning can produce signal variations that will be wrongly recognized as such. Filters can be implemented to discard the false events generated by these noise sources.

The use of efficient filters is critical to have an instrument able to detect lightning pulses of low intensity. Indeed, in order to be sensitive to weak lightning signals, low detection thresholds have to be used in the RTEP. This results in the generation of an increased number of false events, which need to be filtered out.

Since the number of false events can be significantly higher than the number of true events, and since the downlink

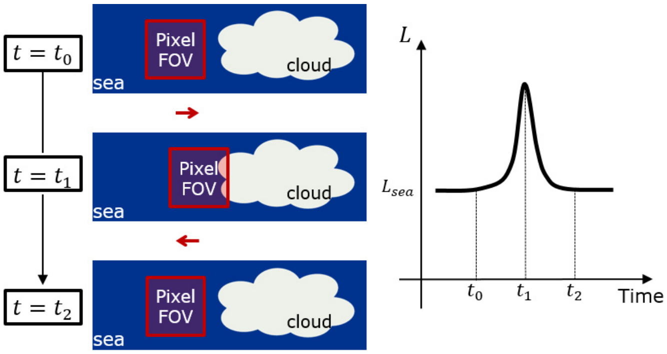

Fig. 4. Jitter movement creating a transient signal. Initially, at time $t_0$, the pixel FOV is only observing a dark scene (i.e., the sea). At time $t_1$, the jitter movement of the line of sight has moved the FOV toward a bright element (i.e., a cloud), increasing the signal acquired on the pixel. At time $t_2$, the FOV has moved in the opposite direction and the signal has decreased. These jitter-induced signal variations can be mistaken for a lightning event.

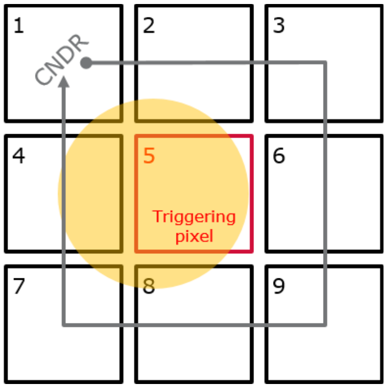

Fig. 5. SDTF. The central pixel (i.e., pixel 5) has triggered an event due to the presence of a lightning pulse (orange circle). Since the pulse is bigger than the pixel's size and not centered on the central pixel, some energy will also be found on the neighboring pixels (i.e., pixels 1–4 and 6–9). The CNDR is a quantity defined to collect the sum of the delta radiances from the neighboring pixels.

bandwidth allocated to the LI is limited, a part of the filtering has to be done on-board the satellite, where the processing power is limited. The additional on-ground filters, applied to the remaining events sent to Earth, will benefit from more powerful hardware equipment to realize more elaborate calculations and further reduce the number of false events before reaching the FC processor.

### *A. Sources of False Events*

In order to design efficient filters for false events, it is important to understand their origin. The two main sources have been identified as the random acquisition noise and the jitter movements of the line of sight.

In this context, acquisition noise refers to random phenomena capable of provoking a sudden increase in the radiance measured over an integration period, such as the shot noise, read-out, or quantization noise. It is worth noting that the shot noise is proportional to the background energy, therefore this type of noise is larger during the day when the observed scene is brighter.

The other main contributor is the jitter movements of the line of sight when observing highly heterogeneous scenes (e.g., coastlines or cloud edges). The LI is mounted on a satellite that can be affected by microvibrations coming from sources external to the instrument, such as the reaction wheels or a cryocooler. The spectrum of the vibrations may contain high frequencies, which would result in the background estimator not being able to follow the induced signal variations. A possible scenario of a false event induced by jitter is illustrated in Fig. 4. The field of view (FOV) of a pixel is initially ($t_0$) centered on a dark background (e.g., sea). At the next time step ($t_1$), the microvibration affecting the line of sight of the instrument shifts the pixel FOV toward a nearby brighter scene element (e.g., cloud), creating an increase in the measured pixel radiance that could be mistaken for a lightning event. This type of false event may occur when a high contrast scene is observed, i.e., under daylight conditions, not during the night when the scene will be uniformly dark.

Other sources will be able to produce transients in the acquired signal, including high-energy particles hitting the detector array, random telegraphic signals (RTSs), stray light, or sun glint. When considering the low detection thresholds used for the LI, simulations have shown that these phenomena are secondary contributors to the production of false events. Specific algorithms to tackle these sources are foreseen but are beyond the scope of this paper.

### *B. On-Board Filters*

The on-board filters are the first filters in the chain (see Fig. 2). They will be fed with the events registered by the RTEP. Their purpose is to reduce the large amount of data coming from the detection step in order to fit within the bandwidth of the communication link to the ground station. The processing power available on-board being limited, these filters must be designed with simple, fast, and yet very discriminating decision criteria.

*1) Single-Detected-Transient Filter:* The rationale of the single-detected-transient filter (SDTF) [10] is that the FOV of the pixels is small enough so that a lightning pulse is likely to illuminate multiple adjacent pixels, as shown in Fig. 5. The *cumulative neighboring delta radiance* (CNDR) is defined as the sum of the delta radiances of the eight neighboring pixels surrounding the pixel having triggered an event and serves, therefore, as a measure of the pulse energy not contained in the central pixel

$$\mathrm{CNDR} = \sum_{i=1, i\neq 5}^{9} \Delta L_i$$

where $i$ indexes the pixels according to the notation introduced in Fig. 5.

The CNDR of each event is then compared to a threshold $\tau_2$, which is a function of the average estimated background of the neighboring pixels

$$\text{if CNDR} < \tau_2\left(\frac{1}{8}\sum_{i=1, i\neq 5}^{9} \hat{L}_i\right) \rightarrow \text{discard event.}$$

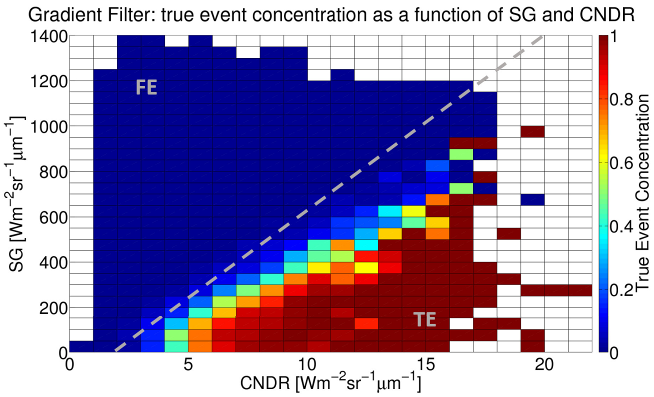

Fig. 6. GF. Distribution of true event concentration according to the CNDR and the SG of events. False events tend to have a high SG and low CNDR. Dashed line: separation curve of the GF, events above this curve are considered false events and are therefore discarded. Data are extracted from the test case simulation described in Section IV-D.

This filter is aimed at eliminating false events due to acquisition noise. Indeed, for these events, the random noise on the central pixel is strong enough to trigger an event. But, since this noise is not spatially correlated, it is unlikely that it would trigger high delta radiances on neighboring pixels during the same integration time frame. These events will therefore have a low CNDR. Most single false events created by high-energy particles and RTS will also be removed by the SDTF.

*2) Gradient Filter:* The gradient filter (GF) is designed to discard false events due to jitter movements affecting the instrument. Jitter-induced false events appear on heterogeneous scenes, such as coastlines or cloud edges (see Fig. 4). The GF uses the *Sobel gradient* (SG) [18] as a tool to measure the degree of heterogeneity of the background scene around the pixel having triggered an event

$$\mathrm{SG}_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \odot \begin{bmatrix} \hat{L}_1 & \hat{L}_2 & \hat{L}_3 \\ \hat{L}_4 & \hat{L}_5 & \hat{L}_6 \\ \hat{L}_7 & \hat{L}_8 & \hat{L}_9 \end{bmatrix}$$

$$\mathrm{SG}_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \odot \begin{bmatrix} \hat{L}_1 & \hat{L}_2 & \hat{L}_3 \\ \hat{L}_4 & \hat{L}_5 & \hat{L}_6 \\ \hat{L}_7 & \hat{L}_8 & \hat{L}_9 \end{bmatrix}$$

$$\mathrm{SG} = \sqrt{\left(\sum_{i=1}^{3}\sum_{j=1}^{3}\mathrm{SG}_{x_{ij}}\right)^2 + \left(\sum_{i=1}^{3}\sum_{j=1}^{3}\mathrm{SG}_{y_{ij}}\right)^2} \tag{1}$$

where $\odot$ denotes element-wise multiplication and the indexes of $\hat{L}$ follow the notation introduced in Fig. 5.

From simulations similar to the one described in Section IV-D, it has been found that sorting events in an SG versus CNDR domain provides a rather neat distinction between true and false events, as shown in Fig. 6. False events tend to have a high SG and a small CNDR. The GF therefore implements a comparison between gradient and CNDR

$$\text{if } \mathrm{SG} > a \cdot \mathrm{CNDR} + b \rightarrow \text{discard event}$$

where $a$ and $b$ are parameters obtained by logistic regression of training data. The use of a higher order polynomial function of CNDR would slightly increase precision but at the expense of a higher computational cost.

This comparison criterion can be represented by a separation curve (dashed line in Fig. 6) in the SG versus CNDR domain. The position of this curve can be dynamically shifted (offset $b$) to regulate the number of events exiting the filter. Since the total number of events that the instrument can send to Earth is constrained by the allocated bandwidth, the separation curve is shifted downward when the number of incoming events increases, in order to eliminate more events. When fewer events are given as input to the filter, the curve is shifted upward. Therefore, the GF allows to maximize the number of true events kept, under the constraint of limited bandwidth.

### *C. On-Ground Filters*

On-ground filters receive the limited number of events that have not been discarded by the on-board filters and must further reduce it to reach the specifications required by the FC processor. Compared with the on-board filters, on-ground processing can benefit from more expensive computations.

*1) Jitter Reconstruction Filter:* The purpose of the jitter reconstruction filter (JRF) is to analyze the events received on-ground to estimate *a posteriori* the jitter movements of the instrument line of sight during the considered period of time. Once the approximation of the jitter is known, it is possible to remove the contribution of this perturbation from the delta radiance of each event and reapply the detection criterion on the remaining signal.

The delta radiance $\Delta L$ of an event can be broken down into a contribution $\Delta L_p$ coming from a lightning pulse (in the case of a true event), a contribution $\Delta L_j$ coming from the jitter movements, and a contribution $\Delta L_n$ from other sources of noise

$$\Delta L = \Delta L_p + \Delta L_j + \Delta L_n.$$

The jitter contribution $\Delta L_j$ can be seen as the result of a small-amplitude movement of the line of sight, projected on North–South ($j_{\mathrm{NS}}$) and East–West ($j_{\mathrm{EW}}$) axes, over an heterogeneous scene surrounding the central pixel

$$\Delta L_j \approx \frac{\partial L_{\mathrm{bkg}}}{\partial \mathrm{NS}} \cdot j_{\mathrm{NS}} + \frac{\partial L_{\mathrm{bkg}}}{\partial \mathrm{EW}} \cdot j_{\mathrm{EW}}$$

where $(\partial L_{\mathrm{bkg}}/\partial \mathrm{NS})$ and $(\partial L_{\mathrm{bkg}}/\partial \mathrm{EW})$ are the derivatives of the background scene seen by the instrument at the central pixel in the North–South and East–West directions. The movements $j_{\mathrm{NS}}$ and $j_{\mathrm{EW}}$ are the part of the jitter that has not been captured in the estimation of the background by the IIR filter. For a particular optical head, these movements of the line of sight are considered identical for all the events occurring at the same time step.

First, the JRF selects a number of events amongst the ones received on the ground, which will be used as *beacons* to be analyzed in order to reconstruct the jitter. A good beacon is defined as an event whose signal $\Delta L$ is primarily composed of the jitter contribution $\Delta L_j$. It is characterized by a large scene gradient (to get a large $\Delta L_j$) and a comparatively small $\Delta L$ (to limit the other contributors $\Delta L_p$ and $\Delta L_n$). The beacon

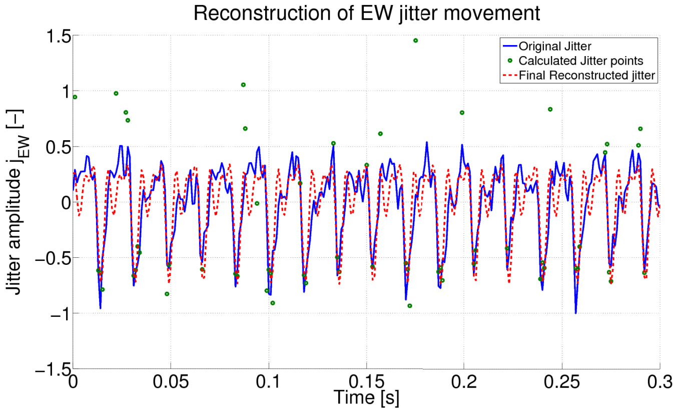

Fig. 7. JRF. East–West normalized amplitudes of the original jitter profile injected in the LI simulator (blue solid line), of the discrete jitter points calculated by solving the linear system (green circles) and of the final reconstructed jitter after the fitting of three harmonics (red dashed line). Data are extracted from the test case simulation described in Section IV-D.

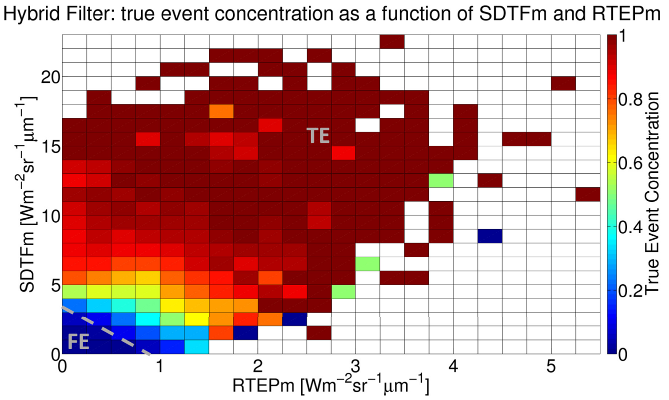

Fig. 8. HF. Distribution of true event concentration according to RTEPm and SDTFm. False events tend to have small margins. Dashed line: separation curve of the HF, events below this curve are considered false events and are therefore discarded. Data are extracted from the test case simulation described in Section IV-D.

selection criterion is therefore to select the events with the smallest ratio of $\Delta L$ over scene gradient.

For each time step where beacons have been identified, the JRF conducts an analysis to estimate the jitter movement during that specific time frame. The analysis assumes that the entire delta radiance of the beacons comes from the jitter contribution and that the background scene derivatives $(\partial L_{\mathrm{bkg}}/\partial \mathrm{NS})$ and $(\partial L_{\mathrm{bkg}}/\partial \mathrm{EW})$ can be well approximated by a simple calculation based on the estimated background information of the nine-pixel window attached to each event. At time frame $n$, if $m$ beacons are available, the JRF solves, in the least-squares sense, the $m$-equation linear system for the two unknowns $j_{\mathrm{NS}}[n]$ and $j_{\mathrm{EW}}[n]$

$$\begin{pmatrix} \Delta L_1 \\ \Delta L_2 \\ \vdots \\ \Delta L_m \end{pmatrix} = \begin{pmatrix} \dfrac{\partial \hat{L}_1}{\partial \mathrm{NS}} \\ \dfrac{\partial \hat{L}_2}{\partial \mathrm{NS}} \\ \vdots \\ \dfrac{\partial \hat{L}_m}{\partial \mathrm{NS}} \end{pmatrix} \cdot j_{\mathrm{NS}}[n] + \begin{pmatrix} \dfrac{\partial \hat{L}_1}{\partial \mathrm{EW}} \\ \dfrac{\partial \hat{L}_2}{\partial \mathrm{EW}} \\ \vdots \\ \dfrac{\partial \hat{L}_m}{\partial \mathrm{EW}} \end{pmatrix} \cdot j_{\mathrm{EW}}[n]$$

where $(\partial \hat{L}/\partial \mathrm{NS})$ and $(\partial \hat{L}/\partial \mathrm{EW})$ are the approximated derivatives of the estimated background in the nine-pixel windows with respect to the North–South and East–West directions.

The next stage in the JRF algorithm consists in fitting a set of harmonics on the discrete jitter points $j_{\mathrm{NS}}$ and $j_{\mathrm{EW}}$. This is based on the assumption that the dominant frequencies (e.g., first harmonics of the cryocooler) of the microvibration spectrum affecting the LI detection are known. After the removal of outliers in the $j_{\mathrm{NS}}$ and $j_{\mathrm{EW}}$ vectors, a least-squares method is applied to find the best phases and amplitudes that fit the set of harmonics on the $j_{\mathrm{NS}}$ and $j_{\mathrm{EW}}$ points. This step increases the precision and extends the estimation of the jitter movements to time steps where no beacons were identified. This process is illustrated in Fig. 7, showing the original jitter, the discrete jitter points obtained by solving the linear systems of equations at different time steps, and the final reconstructed jitter after the harmonic fit.

Knowing an approximation of the jitter movements at every time step $n$, the JRF algorithm can now, for each event received on the ground, estimate the contribution of the jitter to the delta radiance

$$\widehat{\Delta L_j} = \frac{\partial \hat{L}}{\partial \mathrm{NS}} \cdot j_{\mathrm{NS}}[n] + \frac{\partial \hat{L}}{\partial \mathrm{EW}} \cdot j_{\mathrm{EW}}[n].$$

The last step consists of removing the estimated jitter contribution and reapplying the RTEP detection criterion (see Section III) to discard false events due to jitter

$$\text{if } \Delta L - \widehat{\Delta L_j} \leq \tau_1(\hat{L}) \rightarrow \text{discard event.}$$

*2) Hybrid Filter:* The hybrid filter (HF) is aimed at further discarding false events due to acquisition noise. It can be likened to a combination of the RTEP and SDTF criteria. Rather than limiting itself to the binary decision (true or false events), the HF expands these criteria by defining continuous variables indicating the margin by which the events pass the criteria.

For each event, the *RTEP margin*, $\mathrm{RTEP}_m$, and the *SDTF margin*, $\mathrm{SDTF}_m$, are defined as

$$\begin{aligned} \mathrm{RTEP}_m &= \Delta L - \tau_1 \\ \mathrm{SDTF}_m &= \mathrm{CNDR} - \tau_2. \end{aligned}$$

False events due to random acquisition noise tend to pass the RTEP and SDTF criteria with low margins. The HF therefore discards events that feature a small $\mathrm{RTEP}_m$ at the same time as a small $\mathrm{SDTF}_m$. An example of distribution of true event concentration in the $\mathrm{SDTF}_m$ versus $\mathrm{RTEP}_m$ domain is presented in Fig. 8, with an HF separation curve (dashed line) to discard false events.

*3) Spatio-Temporal Coherency Filter:* From the study of physical lightning processes [13], it can be inferred that the LI should not trigger spatially and temporally isolated lightning events. This spatio-temporal coherency feature can be used to further discard isolated false events.

The spatio-temporal coherency filter (STCF) checks that, for each pixel-level event, there exists at least one other event in close spatial and temporal proximity (see illustration in Fig. 9). The performance of this filter is dictated by the spatial and

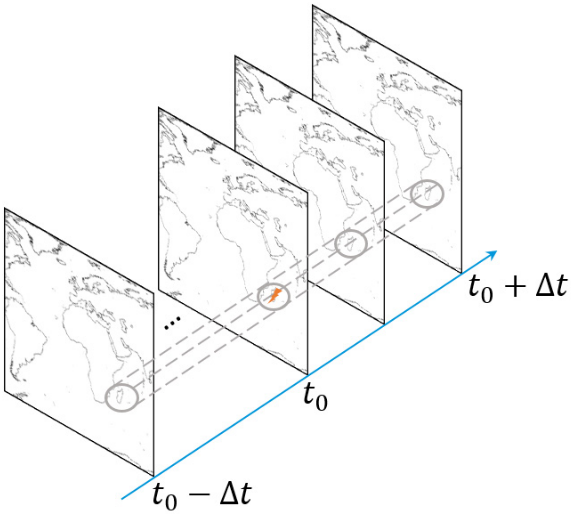

Fig. 9. STCF. An event (orange lightning symbol) occurring at time $t_0$ is considered false event if no other events can be found in close geographical proximity (gray circles) in any of the time frames between $t_0 - \Delta t$ and $t_0 + \Delta t$.

temporal distances used to define the proximity of events. Since this filter is dependent on the number of events given as input, this algorithm must be placed at the end of the chain: the fewer the events that have to be processed by the STCF, the less likely it is that two false events will be in proximity of each other.

This spatio-temporal behavior of the lightning events is used in a more refined way in the FC processor when consolidating the events into group and flash products [15].

### D. Test Case and Results

The algorithmic chain described in this paper has been tested on a simulation. The chosen scenario is based on a background scene generated with a Meteosat Second Generation Spinning Enhanced Visible and Infrared Imager (MSG SEVIRI) image of the October 29, 2011, 12:12 P.M. This has been identified as a representative example of a scene that is apt to produce a high number of false events: the scene is particularly bright and the cloud coverage is made of a multitude of small clouds providing numerous highly contrasted edges (see Fig. 10), which leads to the triggering of a large number of false events due to acquisition noise and jitter. For this scenario, one second of representative LI measurements has been simulated.

The MSG SEVIRI image is processed to produce a top-of-atmosphere background scene spatially and spectrally representative of what the LI would see. On top of this scene, a set of artificial pulses is added, with a random distribution in time and in space, but restricted to cloudy areas (using the corresponding MSG cloud mask product). The generated pulses are circular with a radius of 5 km, have a duration of 0.6 ms, and their energy $L_p$ is set to minimum values as described in the LI specifications [15], using a formula linking it to the background level $L_{\mathrm{bkg}}$ on which the lightning is simulated

$$L_p = \frac{6.7}{1.9}\sqrt{1 + 0.02\,L_{\mathrm{bkg}}} \quad [\mathrm{Wm}^{-2}\mathrm{sr}^{-1}\mu\mathrm{m}^{-1}].$$

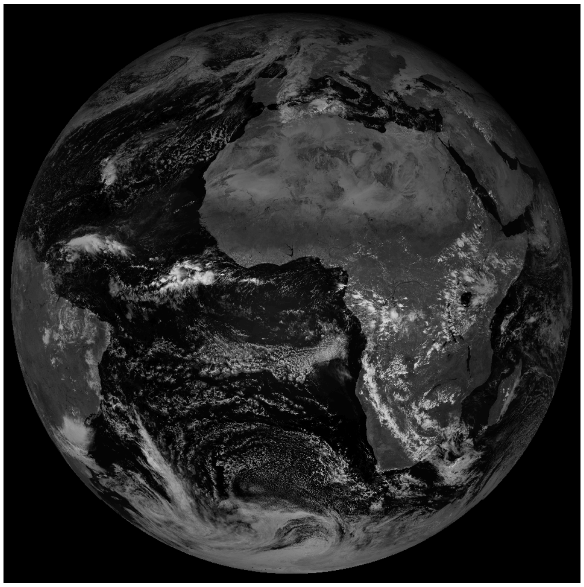

Fig. 10. MSG SEVIRI image (0.8 − $\mu$m channel), from the October 29, 2011, 12:12 P.M., used as background scene for the test case.

TABLE I

RESULTS FROM THE TEST CASE

| | $\#_{TE}$ | $\#_{FE}$ | $\mathrm{TE}_{\mathrm{recall}}$ | $\mathrm{FE}_{\mathrm{reduction}}$ |
|---|---|---|---|---|
| RTEP | 100 | 100 | / | / |
| SDTF | 97.4 | 40.8 | 97.4 | 59.2 |
| GF | 89.0 | 4.2 | 91.4 | 89.7 |
| JRF | 83.9 | 1.5 | 94.3 | 64.3 |
| HF | 83.6 | 0.8 | 99.6 | 46.7 |
| STCF | 83.6 | 0.5 | 100 | 37.5 |

The random distribution of the lightning pulses in space and time is not representative of the typical coherence that exists in lightning storms where pulses are occurring in close spatio-temporal proximity to each other, in the shape of flashes. With the exception of the STCF, this coherency behavior of the lightning is not exploited by the algorithms developed in this paper, which rather focus on the features of each individual event. Hence, the idea behind the random distribution is to obtain a population of true events representative of lightning pulses occurring at various locations in the FOV and with various levels of energy.

The scene and the pulses are then processed by a series of geometric, optical, radiometric, and jitter models simulating the behavior of the LI with the best knowledge of the instrument design as of 2016. The 1000 simulated LI-acquired images are then fed to the algorithmic chain, illustrated in Fig. 2, for events triggering and filtering.

Performance results are presented in Table I. They are expressed in terms of relative number of true events $\#_{\mathrm{TE}}$ and false events $\#_{\mathrm{FE}}$, with respect to the number of events at RTEP output. For each filter, the *true event recall* $\mathrm{TE}_{\mathrm{recall}}$, defined as the proportion of true events that have not been discarded by the filter, and the *false event reduction* $\mathrm{FE}_{\mathrm{reduction}}$, defined as the proportion of false events that have been discarded by the filter, are also shown.

In this particular scenario, it can be seen that the proposed algorithmic chain discards 99.5% of the false events while keeping 83.6% of the true events triggered by the RTEP. Most of the reduction in false events (95.8%) is done by the on-board filters, at the cost of the major part of the true events loss (11%). The more complex on-ground filters further reduce the number of false events by an extra 3.7% while only losing an extra 5.4% of true events. If the events received on-ground are taken as reference, the overall performance of the on-ground filters on this particular scenario is a recall of 93.9% and a false event reduction of 88.1%.

The performance numbers presented here correspond to one particular point on the recall versus false event reduction performance curve for the overall chain. All the described algorithms have parameters that can be tuned in either direction: achieving a higher true event recall performance at the expense of a lower false events reduction, or vice-versa. The algorithmic chain can thus be optimized to find the best compromise between the two objectives.

## V. Machine Learning as Alternative Processing

Alternative ideas for the on-ground processing of events are being investigated, using other techniques from the field of machine learning. The motivation is to make parallel use of all information made available during the execution of the algorithmic chain. Indeed, the chain illustrated in Fig. 2 successively applies filters using only binary decision criteria (i.e., keep or discard the event). The HF was already one step in this direction of parallel usage, combining margin information from RTEP and SDTF. Another goal is to find dependencies of events on other, not yet considered parameters such as the geolocation or the time of the day they occurred.

Separating false events from true events is a typical example of binary classification problems, for which machine learning algorithms have been shown to be extremely powerful. These algorithms are nowadays widely used for the automatic extraction of relevant information from large amounts of data, and increasingly in the field of remote sensing [19].

The followed approach consisted of gathering 200k (100k true and 100k false) simulated events that would have arrived at the ground segment (therefore after application of RTEP, SDTF, and GF) and generated using six different simulation scenarios (various dates, time of day, and jitter profiles). From each event from this large pool, 12 features (i.e., measurable properties) are derived.

1) The identifier of the optical head producing the event.
2) The position of the event computed as a distance to the (0°N, 0°E) point: $(\text{latitude}^2 + \text{longitude}^2)^{1/2}$.
3) Whether the event has occurred during the night or during the day (based on the position and timing of the event).
4) RTEP margin, as defined in Section IV-C2.
5) SDTF margin, as defined in Section IV-C2.
6) GF margin, defined as the distance to the separation curve in the SG versus CNDR domain.
7) Whether the event is eliminated by the JRF or not.
8) The amplitude of the jitter determined by the JRF at the time of the event.
9) The SG calculated on the nine-pixel window.
10) The pixel reading $L$ of the pixel having triggered the event.
11) The delta radiance $\Delta L$ of the pixel having triggered the event.
12) The CNDR.

TABLE II
Results on the Pool of 200k Events

| Algorithms applied | $\text{TE}_{\text{recall}}$ | $\text{FE}_{\text{reduction}}$ |
|---|---|---|
| Nominal chain (without STCF) | 91.0% | 68.3% |
| Decision tree | 98.3% | 70.8% |
| Support Vector Machine | 98.3% | 85.3% |
| Modified chain (without STCF) | 93.7% | 79.7% |

On this large collection of data, various machine learning algorithms can be trained to build a classification model. The choice of a machine learning algorithm is dictated by a trade-off between performance (accuracy in separating the true events from the false events) and interpretability (capacity to understand the underlying classification model, i.e., why the algorithm determines that an event is true or false?). The most powerful machine learning algorithms are generally not easily interpretable [20]. In this particular case, interpretability is critical to verify that the machine learning algorithm is not building a classification model distinguishing true and false events based on characteristics or behaviors linked to unwanted biases introduced in the simulated data.

For this paper, two different machine learning algorithms have been implemented: a decision tree approach, for which the resulting classification model can be interpreted, and a SVM algorithm, known to be more powerful but not easily interpretable. Results of the different approaches are presented in Table II.

The nominal on-ground algorithmic chain (JRF + HF) is applied to the pool of 200k events, which leads to a true event recall of 91.0% and a false event reduction of 68.3%. The STCF is not taken into account here since its working principle is a comparison of all the events generated by a particular simulation scenario to look for spatio-temporal proximity, which is not directly related to the features of individual events.

A custom-made decision tree approach, with a focus on boosting the recall of true events, has also been implemented. The best tree had a performance on the pool of events of 98.3% recall and 70.8% false event reduction. A better performing SVM algorithm has also been tested on the pool of events. By adapting the SVM class biases to improve the true event recall, a recall of 98.3% and a false event reduction of 85.3% could be achieved.

Thanks to the good interpretability of the decision tree, some simple and direct improvements to the previously presented algorithmic chain have been derived and implemented, such as having different HF separation curves for night-time and day-time events. The modified chain led to a true event recall of 93.7% and a false event reduction of 79.7%.

It should be emphasized that these are only preliminary studies which would need to be followed up by more detailed work. The machine learning approach would deliver its highest potential when actual validated data generated by the LI is available. One could imagine using actual events received on-ground from the MTG LI validated with information coming from the ISS LIS or from the GLM, to train machine learning algorithms with actual true lightning events. More investigations on the machine learning approach would still be needed, but the preliminary results presented here already show its potential to design a high-performance on-ground event processor.

## VI. Conclusion

The LI is an Earth observation instrument that will fly on-board MTG Imaging satellites and will perform observations of lightning from a geostationary position. The LI functioning principle consists of acquiring, at a 1000-Hz frequency, monochromatic images to which a detection algorithm is applied to extract lightning events at pixel level.

In order to detect lightning pulses of small energy, the detection thresholds of the instrument need to be set to a low value. Unwanted signal variations due to acquisition noise and jitter movement of the line of sight have been found to be the main sources of false events, with a less favorable situation occurring under daylight conditions. Filters, on-board and on-ground, are needed to discard most of these false events before feeding the data into the FC processor in charge of aggregating events into flash science data products. The benefit of implementing these filters is to allow for the setting of low detection thresholds in order to maximize the detection of weak lightning pulses.

The algorithmic chain proposed in this paper includes filters aimed at eliminating false events due to random acquisition noises (SDTF, HF, and STCF) and jitter movements of the instrument line of sight (GF and JRF). The application of this chain to a test case scenario, representative of a situation producing a high number of false events at the detection stage, showed satisfying results with 83.6% overall recall of true events and 99.5% overall reduction of false events. A machine learning approach has also been investigated to improve the performance of the on-ground part of the processing, using decision tree and SVM algorithms. This approach was shown to be promising but would only realize its full potential if sets of cross-instrument validated true lightning events produced by the LI were available.

## Acknowledgment

The authors would like to thank the teams working in the industrial consortium (Thales Alenia Space, as MTG prime contractor, and Leonardo, as LI prime) and in EUMETSAT for the fruitful collaboration on the on-going Lightning Imager Project.

## References

[1] *Meteosat Third Generation—EUMETSAT*. Accessed: Jan. 12, 2017. [Online]. Available: http://www.eumetsat.int/website/home/Satellites/FutureSatellites/MeteosatThirdGeneration/index.html

[2] S. J. Keogh, E. Hibbett, J. Nash, and J. Eyre, "The Met Office Arrival Time Difference (ATD) system for thunderstorm detection and lightning location," Forecasting Res., Met Office, Exeter, U.K., Tech. Rep. 488, 2006, p. 22.

[3] D. J. Boccippio *et al.*, "The Optical Transient Detector (OTD): Instrument characteristics and cross-sensor validation," *J. Atmos. Ocean. Technol.*, vol. 17, no. 4, pp. 441–458, 2000.

[4] H. Christian *et al.*, "The lightning imaging sensor," in *Proc. NASA Conf. Publication*, 1999, pp. 746–749.

[5] R. J. Blakeslee *et al.*, "Lightning Imaging Sensor (LIS) for the International Space Station (ISS): Mission description and science goals," in *Proc. 15th ICAE*, Norman, OK, USA, Jun. 2014, pp. 1–15.

[6] S. J. Goodman *et al.*, "The GOES-R Geostationary Lightning Mapper (GLM)," *Atmos. Res.*, vols. 125–126, pp. 34–49, May 2013.

[7] J. Yang, Z. Zhang, C. Wei, F. Lu, and Q. Guo, "Introducing the new generation of Chinese geostationary weather satellites, Fengyun-4," *Bull. Amer. Meteorol. Soc.*, vol. 98, no. 8, pp. 1637–1658, 2017.

[8] A. Rodriguez, R. Stuhlmann, S. Tjemkes, D. M. Aminou, H. Stark, and P. Blythe, "Meteosat Third Generation: Mission and system concepts," *Proc. SPIE*, vol. 7453, pp. 74530C-1–74530C-10, Sep. 2009. [Online]. Available: https://www.researchgate.net/profile/Donny_Aminou/publication/258366104_Meteosat_Third_Generation_mission_and_system_concepts/links/553f68190cf2574dcf627ff8/Meteosat-Third-Generation-mission-and-system-concepts.pdf

[9] Y. Durand *et al.*, "The flexible combined imager onboard MTG: From design to calibration," *Proc. SPIE*, vol. 9639, pp. 963903-1–963903-14, Oct. 2015. [Online]. Available: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/9639/1/The-flexible-combined-imager-onboard-MTG–from-design-to/10.1117/12.2196644.full?SSO=1

[10] S. Lorenzini *et al.*, "Lightning imaging from GEO: An innovative sensor for meteosat third generation," in *Proc. 67th IAC*, Guadalajara, Mexico, Sep. 2016, paper IAC-16-B1.3.1.

[11] D. Suszcynsky, T. E. Light, S. Davis, J. L. Green, J. L. L. Guillen, and W. Myre, "Coordinated observations of optical lightning from space using the FORTE photodiode detector and CCD imager," *J. Geophys. Res., Atmos.*, vol. 106, p. 17897–17906, Aug. 2001.

[12] R. E. Orville and R. W. Henderson, "Absolute spectral irradiance measurements of lightning from 375 to 880 nm," *J. Atmos. Sci.*, vol. 41, no. 21, pp. 3180–3187, 1984.

[13] U. Finke, "Characterising the lightning source for the MTG lightning imager mission," EUMETSAT, Darmstadt, Germany, Tech. Rep. EUM/MTG/SOW/05/0028, 2006, pp. 1–46.

[14] U. Finke, "Generation of artificial proxy data for the MTG lightning imager," EUMETSAT, Darmstadt, Germany, Tech. Rep. EUM/CO/09/4600000660, 2010, p. 36.

[15] *Algorithm Theoretical Basis Document (ATBD) for L2 Processing of the MTG Lightning Imager Data*, document EUM/MTG/DOC/11/0155, EUMETSAT, 2014.

[16] H. J. Christian, R. J. Blakeslee, and S. J. Goodman, "The detection of lightning from geostationary orbit," *J. Geophys. Res., Atmos.*, vol. 94, no. D11, pp. 13329–13337, 1989.

[17] D. M. Mach *et al.*, "GOES-R geostationary lightning mapper performance specifications and algorithms," in *Proc. NOAA STAR GOES-R AWG Rev.*, Madison, WI, USA, Jun. 2008. [Online]. Available: https://ntrs.nasa.gov/search.jsp?R=20080037560

[18] K. Pringle, "Visual perception by a computer," in *Automatic Interpretation and Classification of Images*. New York, NY, USA: Academic, 1969, pp. 277–284.

[19] D. J. Lary, A. H. Alavi, A. H. Gandomi, and A. L. Walker, "Machine learning in geosciences and remote sensing," *Geosci. Frontiers*, vol. 7, no. 1, pp. 3–10, 2016.

[20] M. Kuhn and K. Johnson, *Applied Predictive Modeling*. New York, NY, USA: Springer, 2013.

**Pierre Kokou** received the M.Sc. degree in aerospace engineering from the University of Liège, Liège, Belgium, and the M.Sc. degree in aerospace engineering from ISAE-Supaero, Toulouse, France, in 2014.

He is currently with Aurora Technology B.V., European Space Agency, European Space Research and Technology Centre, Noordwijk, The Netherlands, where he is in charge of algorithms and performance for the Lightning Imager.

**Philip Willemsen** received the Diploma and Ph.D. degrees in physics from the Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn, Germany, in 2001 and 2004, respectively.

He was a Systems Engineer and the Project Manager of satellite and space technology demonstration projects with the German Aerospace Center, Bonn, Germany. He joined European Space Agency, Noordwijk, The Netherlands, in 2011. He has led the design and development of systems from concept phases to flight hardware manufacturing and launch. Since 2014, he has been the Principal Instrument Engineer of the Meteosat Third Generation Lightning Imager.

**Mounir Lekouara** received the Engineering Master degree from the École Nationale des Ponts et Chaussées, Marne-la-Vallée, France, and the Ph.D. degree in satellite oceanography from the University of Southampton, Southampton, U.K.

He was with European Space Agency as the Meteosat Third Generation Algorithm and Performance Engineer for the Lightning Imager (LI) and the Flexible Combined Imager (FCI), for five years. Since 2016, he has been a Remote Sensing Scientist for FCI and LI in EUMETSAT, Darmstadt, Germany.

**Madani Arioua** received the master's degree in physical engineering—nanotechnology from the Université Catholique de Louvain, Ottignies-Louvain-la-Neuve, Belgium, in 2014.

He was with European Space Agency, European Space Research and Technology Centre, Noordwijk, The Netherlands, where he was involved in system engineering and algorithm research and development. He joined the private sector, where he is leading research projects within the high-technology industry.

**Andreu Mora** received the B.Sc. degree in telecommunication engineering from the University of the Balearic Islands, Palma, Spain, and the M.Sc. degree in telecommunication engineering from the Polytechnic University of Catalonia, Barcelona, Spain.

He was a Project Engineer with GMV, Madrid, Spain, where he was developing data processors for the Soil Moisture and Ocean Salinity and Swarm satellites. He is currently a seconded GMV Employee with the European Space Agency, European Space Research and Technology Centre, Noordwijk, The Netherlands, where he is in charge of the performance, algorithms, data processing, and analysis tools for Meteosat Third Generation.

**Pieter Van den Braembussche** received the M.Sc. and Ph.D. degrees in electro-mechanical engineering from Katholieke Universiteit Leuven, Leuven, Belgium, in 1992 and 1998, respectively.

In 1998, he joined QinetiQ Space, Kruibeke, Belgium, as a Spacecraft Attitude Control Engineer. In 2007, he joined European Space Agency, European Space Research and Technology Centre, Noordwijk, The Netherlands, where he is currently the Meteosat Third Generation System and Operations Manager.

**Emanuele Neri** received the Laurea degree in electronic engineering from the Sapienza University of Rome, Rome, Italy, and the master's degree in business organization from The Open University in London, London, U.K.

He was with a number of telecom companies in Italy. In 1992, he joined European Space Agency, Noordwijk, The Netherlands, where he has held engineering and management positions in several space programs. He is currently the Meteosat Third Generation-I System and Product Manager.

**Donny M. A. Aminou** received the master's degree in electrical engineering from Université d'Aix-Marseille III, Marseille, France, in 1984, and the Ph.D. degree in physics—optics and high-power lasers from the Institut de Mécanique des Fluides de Marseille, Marseille.

He has over five years of experience in industry. He has been with the European Space Agency (ESA), European Space Research and Technology Centre, Noordwijk, The Netherlands, for over 25 years. His work experience involves satellites and payloads from design to development at both technology and system levels. From 1994 to 2004, he was the SEVIRI System Engineer within the Meteosat Second Generation Project, a series of satellites developed by ESA for EUMETSAT. From 2005 to 2010, he ran the development phases A and B1 of Meteosat Third Generation (MTG). He is currently the MTG Satellite Series Payload Product Manager.