Evaluation of the LI Level-2 algorithm strategies and parallax correction methods

EUM/CO/13/4600001289/JKG

Final Report

Ullrich Finke

University of Applied Sciences and Arts
Hannover

15th September 2014

Contents

1 Introduction **4**

2 Data Clustering Algorithms **5**
2.1 Theoretical Formulation and Classification . . . 5
2.1.1 Cluster Analysis Principles . . . 5
2.1.2 Cluster Analysis for Lightning Data . . . 5
2.2 Evaluation Methodology . . . 6
2.2.1 Bulk Algorithm Description . . . 7
2.2.2 Lightning Test Data Source . . . 8
2.3 Group Clustering . . . 8
2.3.1 Distance Functions and Linkage Criteria . . . 8
2.3.2 Evaluation . . . 9
2.3.3 Discussion . . . 11
2.4 Flash Clustering . . . 12
2.4.1 Distance Function and Linkage Criteria . . . 12
2.4.2 Evaluation . . . 14
2.4.3 Discussion . . . 16
2.4.4 Clustering of Proxy Data Sets . . . 16
2.5 Summary of Data Clustering . . . 17

3 Accumulated Products **18**
3.1 Theoretical Evaluation . . . 18
3.1.1 Product Definition . . . 18
3.1.2 Mathematical Formulation . . . 19
3.1.3 Discussion of the Product Definition . . . 20
3.1.4 Individual Product Discussion . . . 22
3.2 Investigation of Test Cases . . . 22
3.2.1 Accumulated Products Generation . . . 22
3.2.2 Detailed Intercomparison of the Accumulated Products . . . 23
3.2.3 Accumulated Products for Proxy Data . . . 25
3.3 Summary of Accumulated Products . . . 27
3.3.1 General Summary and Conclusion . . . 27
3.3.2 Characterization of the Products . . . 27
3.3.3 Prospected Use of Accumulated Products . . . 28

4 Parallax Correction **29**
4.1 Mathematical Formulation . . . 29
4.1.1 Geostationary Projection . . . 29
4.1.2 Parallax Error . . . 30
4.2 Correction of the Parallax Error . . . 31
4.2.1 Correction Strategy for Level-2 Products . . . 33
4.2.2 Discussion of Altitude Functions . . . 33
4.3 Summary of Parallax Correction . . . 36

5 Summary and Conclusion **37**

References **39**

List of Figures **40**

Abbreviations **41**

1 Introduction

The Lightning Imager (LI) on the Meteosat Third Generation (MTG) is dedicated to detect lightning with high temporal and spatial resolution. This will be accomplished by processing the incoming optical radiation in order to detect the lightning signal on the bright cloud background. The primary data output are the 'events' corresponding to triggered pixel events of the detector matrix which where caused by lightning.

These data are processed in a second processing level (L2) into a set of products: (i) a hierarchy of events-groups-flash which organizes the point data by their structure and connectivity, and (ii) accumulated products which provide integrated information on lightning distribution.

The specification of these products and the principal algorithms of product generation are described in the Algorithm Theoretical Basis Document (ATBD) for the Lightning Imager (ATBD-LI, 2012). The aim of the present study is to investigate and to evaluate the proposed algorithms of L2-processing. In the ATBD the specifications are kept flexible and contain different variants for the algorithms and adjustable parameters. These variants and parameters are also investigated in this study.

The specific topics of the study are the:

- evaluation of the L2 group and flash clustering algorithms,
- evaluation of the L2 accumulated product algorithms,
- investigation of the parallax correction method options.

Each of these topics is first theoretically formulated and discussed. Afterwards, the algorithms are investigated in simple implementations using test data. These lightning test data are prepared from the NASA LIS observation and from proxy data generated an the basis of lightning ground observations.

This report is organized in chapters for each of the three topics and contains a summary with conclusions and recommendations.

2 Data Clustering Algorithms

The MTG Lightning Imager (LI) detects the optical radiation from lightning by means of a pixel matrix. The level-1 output data is the stream of detected pixel 'events' which were generated by the lightning discharges.

The purpose of the level-2 data clustering is to obtain the natural structure of the detected lightning phenomenon. It consists in organizing the pixel events into a hierarchy of 'groups' and 'flashes', where a group corresponds to a single optical pulse from an electrical discharge and a flash represents a lightning flash consisting of several discharges. The basic algorithm strategy for this group and flash building is partially inherited from the NASA OTD/LIS approach (Christian et al., 1989, 1999) and is described in the ATBD-LI (2012). This chapter discusses and evaluates these algorithm strategies.

2.1 Theoretical Formulation and Classification

2.1.1 Cluster Analysis Principles

The operations which are to be performed by the level-2 group and flash building algorithms are special cases of the cluster analysis. In general, a cluster analysis is used to organize a set of objects in several clusters by minimizing a certain cost function which is calculated from the relative position of the objects to each other. More specifically, a distance function quantifies the proximity between the objects or relative to the established cluster and a linkage rule defines whether an object is to be included in a cluster.

Distance Function

The distance function $d(i,k)$ is a measure which quantifies the 'proximity' of the objects i and k. This function is usually also a metric, i.e satisfies the triangle inequality, symmetry and zero self distance condition. The distance can be a continuous or discrete function in the object parameter space. Examples are Euclidean distance, lattice distances, distances on trees.

Linkage Rule

The linkage rule defines the criterion for including an object into a cluster. Usually it is formulated by threshold values for the distance. This can be the mutual distance between objects, the distance to the cluster center (mean, median), or the distance to the nearest or the farthest neighbor in the cluster.

2.1.2 Cluster Analysis for Lightning Data

The lightning event data which are to be clustered have specific, distinguishing properties which result from the lightning characteristics and from the detection process.

Characteristics of Lightning Data

The optical radiation, which is detected by the instrument, originates from the luminous lightning channel in the cloud. This original radiation is transported by multiple scattering on the cloud particles toward the cloud surfaces. As the result of this diffusive process the initially short and narrow signal is broadened in time to about 600 μs pulse width and in space to a 10 km pattern (see Christian et al., 1989). This optical pulse radiates from the cloud upper surface and is detected by the MTG-LI.

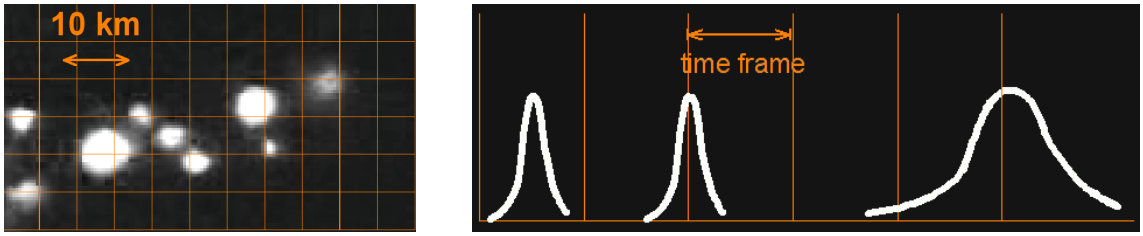


Figure 2.1: Optical signature of lightning observed on the cloud top. Left: Spatial pattern with typical size of 10 km. Left: Time series scheme with typical width of 0.6 ms.

Detection Process

The detection procedure contains an integration of the signal over time and space intervals. This results in a discretization of the optical radiation pulse $B(x, y, t)$ in time and space:

$$b_i(x_i, y_i, t_i) = \int_{x_i}^{x_i+\Delta x} \int_{y_i}^{y_i+\Delta y} \int_{t_i}^{t_i+\Delta t} B(x, y, t)\, \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}t \tag{2.1}$$

This expression is simplified here, e.g. the instrument function is neglected (set to 1).

The original signal B has finite extent in time and space. Therefore, the signal will be covered by more than one integration interval, i.e. it is split into several events. The relation of the typical spatial size and temporal width to the detection discretization parameters determines the degree of splitting, i.e. the number of events and also the energy distribution to each event.

The discretized radiation data are tested against a detection threshold. The result is the set of the detected pixel 'events' with the discrete parameters: time, location x- and y-value, and amplitude.

$$e_i = \{t_i, x_i, y_i, b_i\} \tag{2.2}$$

For convenience the event data also contain the corresponding geographical coordinates (longitude, latitude) and a calibrated radiance value.

These characteristics determine the choice of the clustering method – in particular the distance function and linkage rule. The distance function is chosen as lattice distance on a discrete grid for the events and as Euclidean distance for groups. For the time distance the simple linear difference is used. The continuity of the pulse suggests to apply the 'nearest neighbor' linkage rule.

2.2 Evaluation Methodology

The algorithm strategies for the clustering into groups and flashes are described in the ATBD. These strategies formulate the basic principles and possible variants of algorithm steps. They contain parameters which can be varied and were intentionally kept flexibly for later adaptation to instrument specifics and optimization.

These strategies have to be critically discussed and evaluated in this report. The parameters which have to be investigated in the context of the data clustering are the space-time criteria for group and flash clustering. The evaluation methodology consists of 2 steps: 1. theoretical discussion and 2. statistical investigation of the algorithm parameters variation.

For the statistical analysis the following tasks were performed:

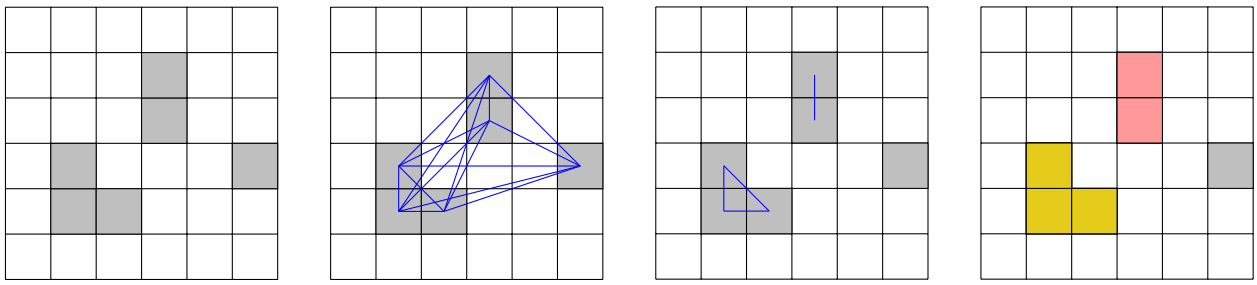
- develop a simple modular and flexible clustering algorithm for group and flash building
- prepare data for the test runs
- perform and interpret the test runs

2.2.1 Bulk Algorithm Description

A conceptually very simple algorithm was constructed for the evaluation of group and flash clustering. The basic idea is to establish adjacency for each possible event pair and merge the adjacent pairs into clusters. The algorithm is a bulk algorithm, i.e. processes all data in one call. This is advantageous for the study purpose. Any processing in data chunks or windows complicates the program code with the organization of loops.

The steps of the algorithm are (see Fig. 2.2)

1. Calculate the pairwise distance for all events $d(i,k)$
2. Apply the linkage criterion for the event pairs
3. Iterate through the pairs, assign cluster numbers, merge clusters if applicable:
   - if both events are still unnumbered: assign to both a new available group number
   - if one of the events already has a number: both get this number
   - if both events already have but different numbers: both get the higher number
   - isolated events get own group numbers
4. Calculate the parameter field values for the clusters



(0) pixel events (1) pair distances (2) adjacent events (3) groups

Figure 2.2: Scheme of the simple clustering algorithm steps. For the pixel events (0) all pairwise distances are calculated (1). The linkage rules selects the adjacent pairs (2). Pairs are merged into clusters (3).

Notes

- The separation of distance calculation (2) and linkage criterion application (3) makes it possible to introduce and combine various distance functions and linkage rules in a modular way.
- The distances calculation (2) is of complexity $\sim O(N^2)$, in practice a pre-filtering in time reduces the complexity to $\sim O(N)$

The algorithm is described here for documentation of the used evaluation tools. The Matlab-code is provided in the software package to this study. It is not intended to be a prototype of the future operational algorithm, which has to be a streaming algorithm as specified in the level-2 processing specification. The real data will arrive in chunks from the satellite, and the algorithm must run infinitely.

2.2.2 Lightning Test Data Source

The data for the evaluation procedure were derived form the optical lighting detector LIS on the TRMM satellite which is in a low (now 400 km) orbit with 35° inclination for more than 10 years now (Christian et al., 1999).

LIS-data are well suited for the evaluation procedures by several reasons: They are optical data in the same wavelength and the detection procedure is very close to the prospected LI detection. The event data contain realistic time-space structure and the existing group-flash structure build by the OTD/LIS algorithms can be compared against the new algorithm results. Additionally, simulated proxy data can be generated for the investigation of special situations. This was not done here, since the clustering method is relatively simple.

The short 90 s observation time for LIS is not a limitation in the present context, since flash duration is much shorter. The motion of the low orbit spacecraft results in slight spatial shift of 15 m between time frames of 1.9 ms and of 1 km during typical interstroke times of 150 ms which can be neglected.

LIS event data were gathered with a time frame of 1.9 ms but some jitter appears in the data. This time frame differs from the expected time frame for the future LI instrument. This fact has to be kept in mind, but is not a limitation for the principal evaluation in this study.

The large amount of over 10 years LIS-data allow for the processing of long series with statistical results of high confidence.

2.3 Group Clustering

The aim of the group clustering is to restore the detected optical pulse which was split into several events by the detection procedure.

Group parameters The group parameters are time, center longitude and latitude, total footprint, total radiance, number of child events. These parameters are calculated from the composing events. Additionally, the group holds a reference to the events (children) belonging to the group. In turn, each event also refers to its parent group.

The degree of splitting (resp. the number of children events) depends on the ratio of pulse size to detection interval size. Spatial splitting is obvious since the pulse size is usually larger than the detector pixel footprint. Temporal splitting happens also frequently since even for pulse width t_{pulse} smaller detection integration frame width t_{window}, a high probability of splitting may appear. For an ideal rectangular pulse the probability of being split is $P = t_{pulse}/t_{window}$. For typical values of $t_{pulse} \approx 0.6$ ms and $t_{window} = 1$ ms this is 60% probability for splitting.

2.3.1 Distance Functions and Linkage Criteria

For the group clustering the distance of events have to be considered on the discrete detection grid. The event data have discrete space and time values. Here, for simplicity a grid spacing of 1 is assumed, which can be achieved in practice by normalizing.

Adjacency on a discrete lattice is defined as all cells surrounding a given cell which attach a side (von Neumann neighborhood) or attach a side or vertex (Moore neighborhood). For the square lattice the 'chessboard' distance function is used here defined for 2 pixel events as

$$d(e_i, e_k) = \max(|x_i - x_k|, |y_i - y_k|, |t_i - t_k|) \tag{2.3}$$

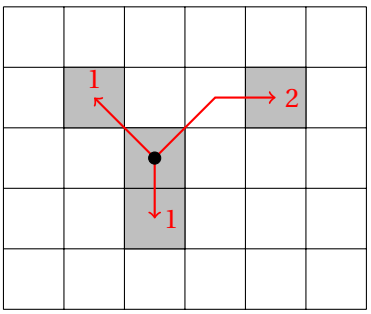
It corresponds to the distance for the moves of the king on a chessboard (see Fig. 2.3). The nearest neighbor adjacency is given by all events at distance $d(e_i, e_k) = 1$. Note the inclusion of the time difference, i.e.

all events which attach by a side or corner and which are adjacent or simultaneous in time are considered being adjacent.

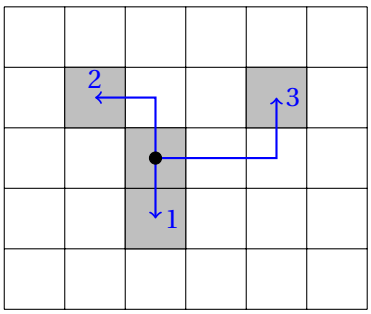
An alternative distance function on the lattice could be the 'manhattan distance', where distance is determined by the horizontal and vertical moves.

$$D(e_i, e_k) = |x_i - x_k| + |y_i - y_k| + |t_i - t_k| \tag{2.4}$$

For the clustering of lightning events to groups the chessboard distance is chosen. This is supported by the fact, that the pulses have usually a convex shape and the events are detected by a threshold. It can always be assumed, that neighboring pixels are also filled with energy from the pulse but below the threshold. Hence the inclusion of corner events will not wrongly attach pixels.



chessboard distance



manhattan distance

Figure 2.3: Lattice distances chessboard (left) and manhattan (right). The arrows denote the path from the central event (black dot) to the other events, the numbers indicate the distance. All events with distance 1 are included in the group cluster.

As described above the linkage rule is nearest neighbor in space and time, i.e. all events belong to the same cluster for which

$$d(e_i, e_k) = 1 \tag{2.5}$$

2.3.2 Evaluation

The ATBD specifies for groups the criterion of spatial adjacency and adjacency in time as an option. This is different to the OTD/LIS processing, where groups only in the same time frame were considered. As discussed above from physical reasons the adjacency in space and time should be both equally applied, since the events are produced by splitting of the optical pulse during the detection process in space and time intervals. For the MTG-LI the length of the detection time frame is set as a variable parameter. The smaller the detection time frame the higher is the probability of splitting in time.

In this evaluation the consequences of the time adjacency option will be analyzed for test data and compared against the LIS-processing. Additionally the variation of the distance and time threshold is investigated.

Group Clustering by Time Adjacency

In this analysis LIS event data for different days were processed into groups by the methods described above. The statistics of group parameters were compared against the old LIS-processing.

Fig. 2.4 (left part) shows the frequency of time differences between groups in the LIS-processing, i.e. without time adjacency. The first bin of the histogram contains groups with the same time stamp. These groups are partially 'broken' groups, i.e. groups produced by the same lightning pulse but too far apart

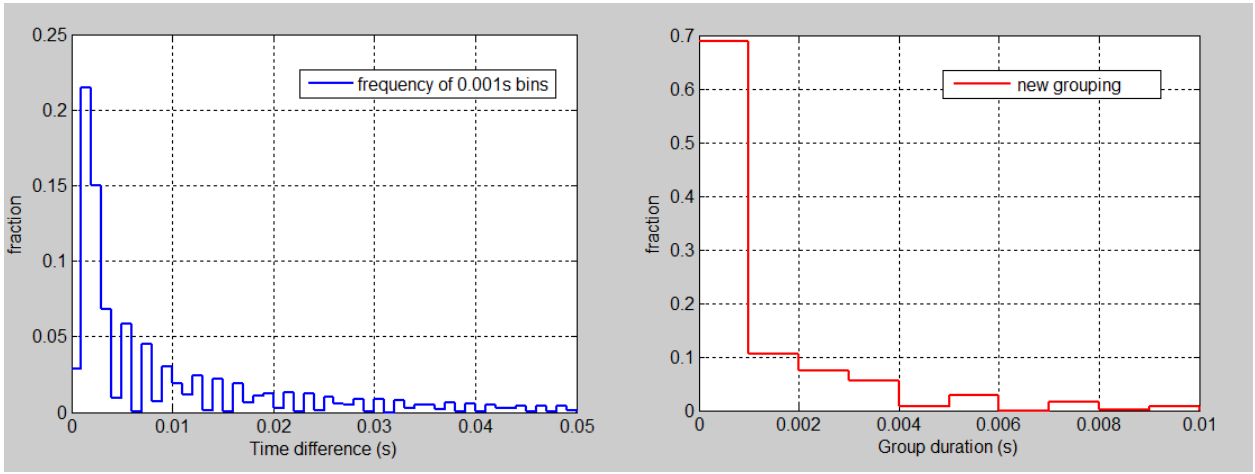


Figure 2.4: Left: Histogram of time differences between groups for LIS-clustering. Many groups are adjacent in time (≈40%). Right: Histogram of group duration in new clustering. 70% of the groups extend over 1 or 2 time frames.

to fall in the same cluster. The partition of these broken groups is <3%. The next 2 bins contain groups which follow each other in time without time gap. These are interpreted as split groups, their partition is ≈40% of all groups. These groups will be joined by the new clustering procedure.

The processing with the new criteria, which include time adjacency reduces the number of groups by ≈38%. Fig. 2.5 compares in the left part the difference between group starting times for LIS and for the new processing. In the new data the large portion of groups with small time differences has disappeared. The enhanced frequency of group time differences up to 0.025 s corresponds to the multistrokes in a lightning flash. The right part of Fig. 2.5 shows the histogram of child counts, i.e. the number of events per group. This distribution changed moderately: the new processing produces less small groups and more larger groups.

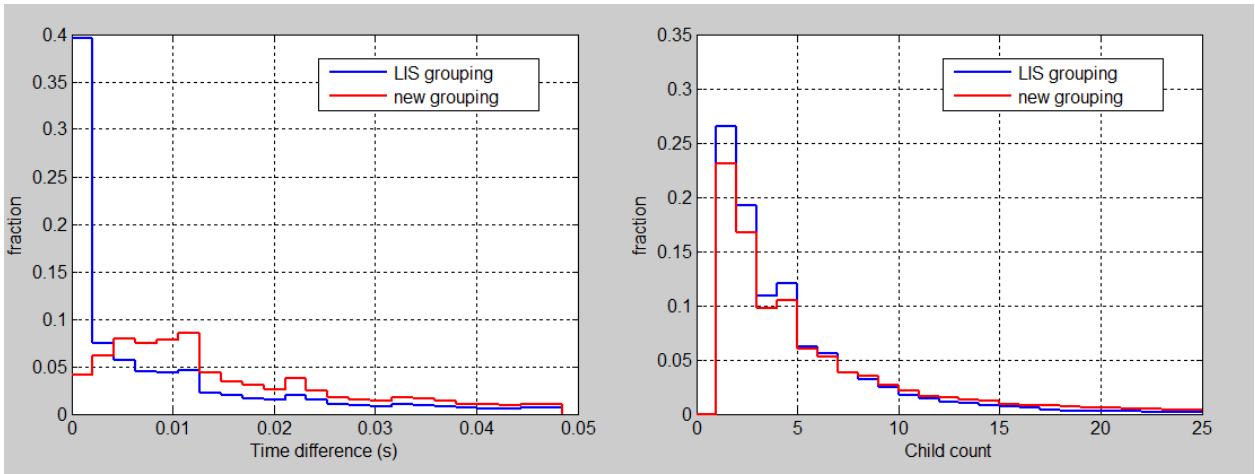


Figure 2.5: Comparison of group statistics for LIS and for the new processing. Left: Distribution of time differences between group start times has now only a small fraction of very close groups. Right: the child count distribution changes towards larger groups.

Due to the applied time-adjacency linkage rule groups can now extend over one or several time frames. As a consequence new group parameters can be introduced and other parameters have to be interpreted differently. First, a time duration parameter can be specified. Most of the new groups are still of short duration as the right part of Fig. 2.4 shows. Additionally, a center time for the group can be calculated accordingly to the center location.

Some of the existing group parameters change their interpretation. So, the footprint for a multi-frame group describes the maximum extent (the silhouette) of the group in space, while for a one-frame group it is just the total of the event footprints.

Variation of Linkage Thresholds

The values for space and time linkage threshold were varied in discrete steps. The resulting change in group number is shown in the Table 2.1

distance threshold (pixel)	time threshold (frame)	change of group numbers	remark
1	0	160%	LIS-grouping
1	1	100%	new grouping
2	1	97%	broken groups
2	2	85%	interrupted groups

Table 2.1: Percental changes in the number of groups due to changes in the linkage criteria threshold. The first row correspond to LIS processing, i.e. groups are limited to single time frames.

The data show the significant reduction of the number of groups between LIS and the proposed new MTG-LI processing. A further increase of the neighborhood changes the number of group only slightly. This corresponds to the merging of groups with portions of lower radiation energy, which are partially below the detection threshold. An increased neighborhood can rejoin these groups.

2.3.3 Discussion

The theoretical reasoning and the results of the test runs show the general validity of the group clustering strategy. Corresponding to the data characteristics most appropriate is the use of the chessboard distance function. This is a discrete lattice distance, which gives distance 1 to all 8 neighbor cells on a square lattice. The linkage criterion for clustering is adjacency of events in time and space.

The consequent application of the time and space adjacency changes the group parameters: New parameters describe the temporal extension: start time and duration, also a center time can be introduced. Additionally, existing parameters change such as the footprint area which contains now the maximum area covered by the group and differs from the total of children footprints. The child count parameter has to be interpreted with care. In one-frame groups it can be interpreted as proportional to spatial size. For mulit-frame groups the random time splitting enhances the number of children artificially. Here, the parameters of total radiation and footprint area and duration should be preferably used for analysis of group characteristics.

Assembling groups by their full temporal extent implies also different results compared to the OTD/LIS processing. In OTD/LIS all groups are single time frame groups. In the new processing the number of groups will be lower, since split groups are joined now. Consequently, the statistical distributions of other parameters such as child count, total radiance, footprint area will change to larger values. The statistics from MTG-LI and LIS can be easily compared after a reprocessing of the LIS data by the proposed algorithm to produce groups with full time extent as demonstrated in the above results.

The study of the variation of the linkage threshold values shows that a further increase of the distance and time thresholds does not increase the group number significantly. Hence, the chosen distance of 1 pixel in space grid and 1 time frame is the recommended setting for these parameters.

Broken Groups

Due to the inhomogeneous scattering processes optical pulses may contain darker regions. Depending on the detector threshold setting fainter pixels can be undetected and the whole group becomes broken

into several groups. In principle this problem can be addressed by enlarging the linkage criterion, i.e. accepting also distances larger 1, and perhaps introduce a radiance weighting. However this makes the algorithm more complex and adds degrees of freedom which cannot fixed by the data.
Moreover, the group clustering is not intended to reproduce the real optical pulse, but rather to restore the optical pulse as it was detected by the instrument. The above analysis of LIS event data shows that broken groups are not very frequent.

Noise

The algorithm strategies still do not take into account the various influences from noise, since by definition the input level-1b data is free of noise. However, the effects of 'remaining' noise can be estimated. In contrast to the events which are organized in groups the noise appears either without any structure as random small points in time and space. Or other types of noise show a characteristic fingerprint (e.g. 'lollypops', particle tracks in LIS), which makes it recognizable.

Non-lightning noise pixels might be falsely classified as 1 event groups, thereby enhancing the number of small groups. Another possible effect of remaining noise can be the random attachment of noise pixels at the groups' borders, thereby enlarging the groups.

The influence of noise to the group clustering is still to be investigated using the results of the noise studies which are currently in progress.

2.4 Flash Clustering

A lightning flash comprises the whole electrical discharging activity between charged cloud regions and also to the ground which is realized by hot ionized air channels in the result of an electrical breakdown. It usually consists of several single discharge processes of various kinds. The whole flash is located in a certain cloud region but can also propagate by changing the electrical environment of the cloud. While the single breakdown is a very short process of just a few microseconds, the whole flash can last up to one second and more.

In ground based lightning location systems usually the striking point of the return stroke to the ground is of special interest and is the primarily detected flash component due to the instrument design. A lightning flash then comprises multiple return strokes which use the same lightning channel and usually have the same ground striking point and are separated typically by tens of milliseconds. For many applications in now-casting, cloud physics and chemistry and also in climatology the complete electrical discharge process in the cloud – the total lightning – is of great value. In the case of optical detection from space all luminous components of lightning contribute to the optical signature which appears after scattering on the cloud surface. Hence, the optical radiation signal contains a large part of the total lightning processes.

The purpose of the flash clustering algorithm is to collect the groups, which correspond to single optical pulses, into flashes.

2.4.1 Distance Function and Linkage Criteria

The flash, which has to be constructed, is not split into groups by the detection process but rather by the physics of the electrical discharge process and charge distribution in the cloud. Therefore, the distance function has not to be defined on the detection lattice. Instead, the real physical distances can be used which can be related to the typical dimensions of lightning flashes. The spatial distance is measured in kilometer on the ground projection, temporal distance can be measured in milliseconds.

While the defining concept for group building was immediate adjacency, for the flash components the space and time criteria are more arbitrarily defined. Still the proximity in time and space defines the linkage rule, however a wider range of threshold values is possible.

The ATBD specifies the basic strategy of the flash building algorithm. This strategy and the provided distance function and the linkage rule are oriented on the LIS and GLM processing steps. Both will be evaluated in this chapter by theoretical arguments and statistical analysis of test data.

Theoretical Considerations

Distance Function in flash clustering is continuous and related to the real space distance. The ATBD proposes the weighted Euclidean distance (WED), which forms the L_2-norm of the distances along the 2 space coordinates and the time coordinate after proper normalization.

$$d(e_i, e_k) = \sqrt{(x_i - x_k)^2 + (y_i - y_k)^2 + a^2 \cdot (t_i - t_k)^2} \tag{2.6}$$

Thus, in the WED the time and space coordinates are treated uniformly. This makes the approach appealing. On the other hand, there is no physical reason for this equal treatment of space and time for the lightning flash. As a consequence of the summation of the terms in (2.6) an increasing spatial distance might be compensated by a decreasing time distance. This is not the expected behavior for signals from lightning. Rather the opposite would be more obvious: for a propagating flash an increase in space corresponds to an increase in time.

Instead of this (unnecessarily) combined space-time distance a simple independent distance function is proposed here. With spatial distance $\Delta r_{ik} = \sqrt{(x_i - x_k)^2 + (y_i - y_k)^2}$ and temporal distance $\Delta t_{ik} = |t_i - t_k|$ a L_1-norm type of maximum distance function is defined by

$$d(e_i, e_k) = \max(\Delta r_{ik}, b \cdot \Delta t_{ik}) \tag{2.7}$$

where b normalizes time to the desired space distance.

It is expected however, that in practice the different flavors of distance calculation will have no strong effect on the flash clustering results.

Linkage rule in flash clustering should be also the 'nearest neighbor' rule, i.e. distance between the outer borders have to be shorter than a certain threshold. The distance to the center point of the flash is not an appropriate criterion, since the cluster center is only established by the building process and will change during the processing time. This may lead to ambiguous results for flashes. Moreover, the pure size of a flash should not be an excluding criterion. It is well known that flashes can extend over large cloud regions. A more appropriate, natural criterion is the existence of a gap in time and space between discharged regions and consequently for flash size and duration.

The physical nature of the lightning process also suggests a nearest neighbor linkage rule: The spatial extent of the group is a result of an extended line source, usually there is no center point for the lightning discharge channel. Hence the outer border marks the physical extent of the electrically activated region and should be used for the linking of the pulses belonging to a flash.

Nearest distance to the outer border means distance between events of the groups. Hence, in this approach the flashes can be calculated by the event data alone. Both group and flash clustering is independently calculated from the event data. The group clustering is not necessarily a prerequisite for the flash.

For the proposed distance function the 'nearest neighbor' linkage for events of the groups formulates as:

$$\Delta r_{ik} \leq D_r \quad \wedge \quad \Delta t_{ik} \leq D_t \tag{2.8}$$

i.e. both time and space distances must be independently smaller than the threshold.

Figure 2.6 illustrates the described strategy for flash building.

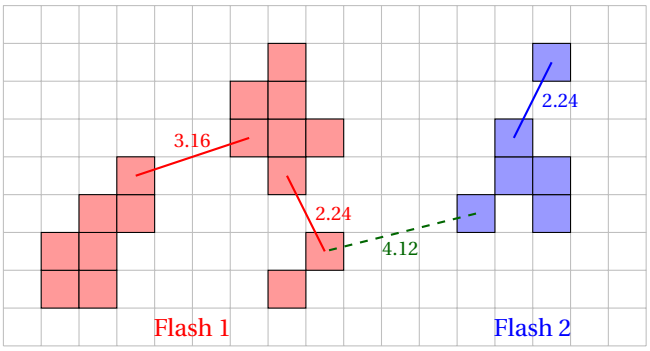


Figure 2.6: Scheme of flash clustering: Several groups are associated to flashes by the euclidean distances of their nearest events. The distance between the nearest events of the blue and red flashes is too large (broken line). The time criterion is not depicted here.

2.4.2 Evaluation

The flash building algorithm was investigated with respect to variations in the threshold parameters. As before, LIS event data were used which were processed into groups as described in the previous section. For the groups and their contained events the clustering into flashes is performed. The same algorithm was used now with the distance function and linkage rule for flashes given in the previous section 2.4.1.

The relevant variable parameters are distance limit D_r and time limit D_t. The principal ranges for these parameters are on the lower bound the discretization interval, i.e. the pixel size, and on the upper bound the cloud size. Reasonable values for both parameters are in the range of several kilometers and few hundreds of milliseconds.

The ATBD for the GLM specify for these threshold $D_t = 330$ ms and $D_r = 15$ km.

From statistics of lightning distribution useful ranges for these parameters can be estimated. The graph of the distribution of time differences between groups (Fig. 2.7) shows strong decrease until ≈0.4 s, with more uniform distribution for larger time differences. This suggests a flash clustering time threshold at about 0.3-0.5 s .

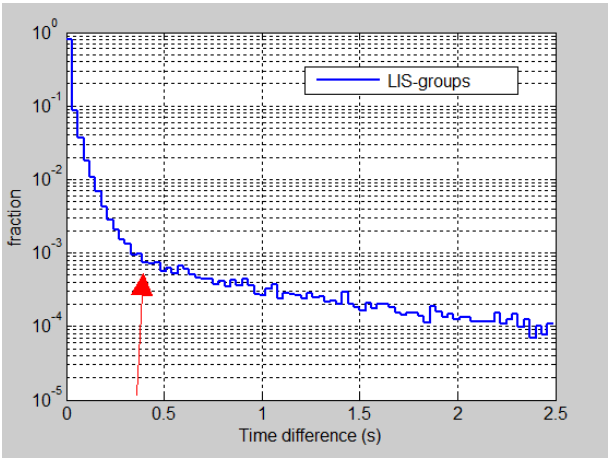


Figure 2.7: Frequency distribution of time differences between groups. The strong decrease smooths at time differences of 0.4 s. This gives an indication for the time threshold for flash clustering.

The ATBD for LIS and GLM specify a stop criterion to prevent the algorithm from building a flash infinitely. This situation is principally possible but did not occur in practice (Mach, personal communication).

Flash Examples

The capability of the flash clustering algorithm in building and also discriminating flashes is demonstrated in Fig. 2.8-2.9. The examples show flashes which are in the same area but separated in time (Fig. 2.8). The opposite situation of flashes overlapping in time, but distant from each other is shown in Fig. 2.9.

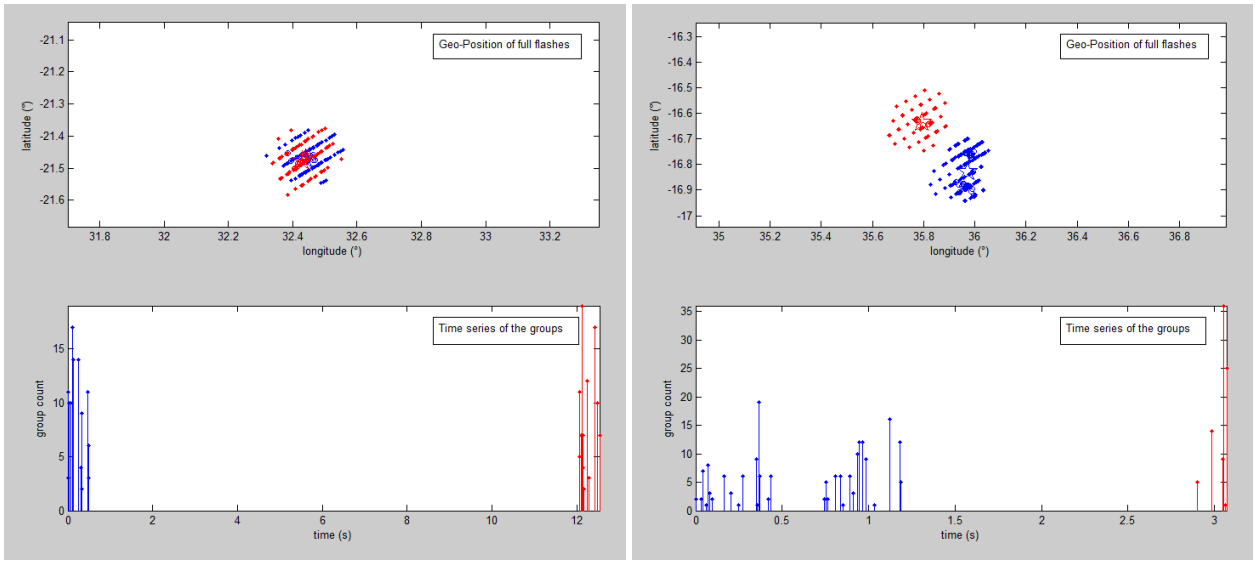


Figure 2.8: Flashes at close spatial distance but with large time distance are clearly distinguished.

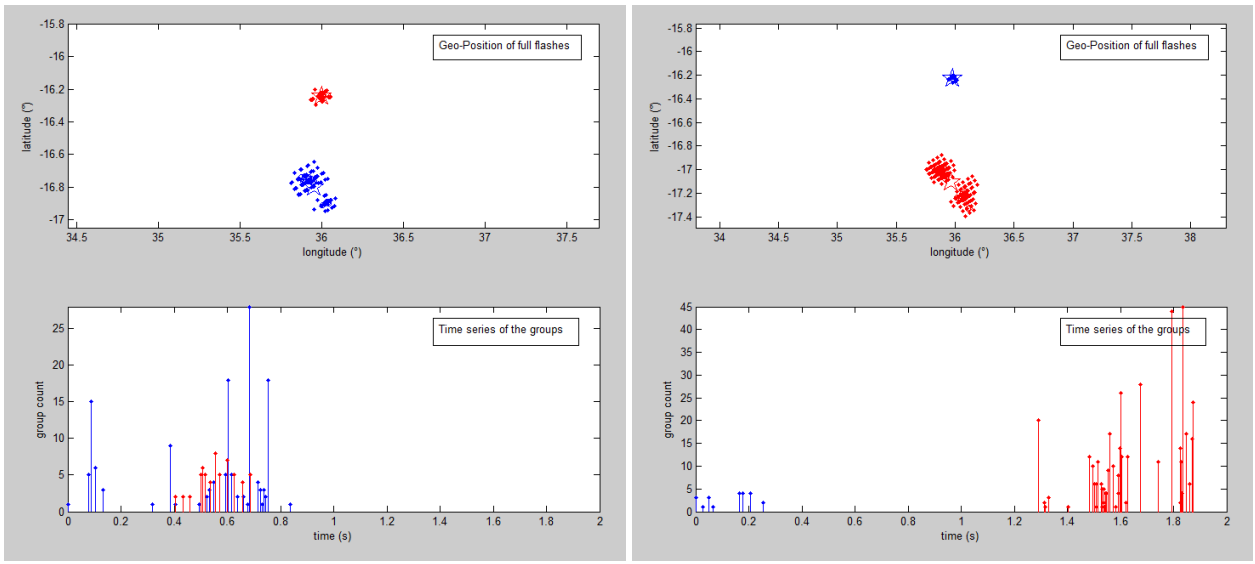


Figure 2.9: Flashes overlapping in time (left), but at different positions. Both pictures show also flashes with distinguishable groups, which were close enough to be associated to a common cluster.

Parameter Variation

The output of the flash clustering was investigated for varying space and time threshold parameters. The results are given in Tab. 2.2 and on Fig.2.10

distance threshold (km)	time threshold (s)	change of flash numbers
5.5	0.33	100%
11.0	0.33	96%
16.5	0.33	93%
5.5	0.17	133%
5.5	0.33	100%
5.5	0.66	95%
5.5	0.99	91%

Table 2.2: Percental change in the number of flashes due to changes in the linkage criteria threshold for spatial and time distance between the nearest events of the groups. The data are referred to the values $D_r = 5.5\,\text{km}$ and $D_t = 0.33\,\text{s}$. Except for very short time, there are only small changes.

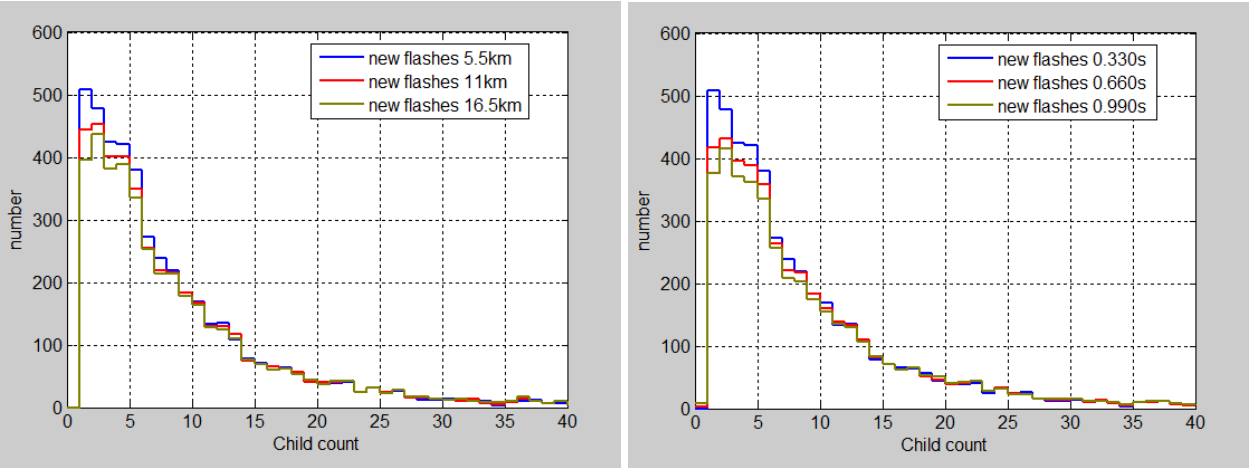


Figure 2.10: Distribution of group number in flashes for variation of the thresholds for distance and time. Larger threshold values decrease the number of groups only slightly.

2.4.3 Discussion

The results of the flash clustering are robust against variation of the algorithm parameters space and time threshold. For space threshold $5.5\,\text{km} < D_r < 16\,\text{km}$ and time thresholds $0.33\,\text{s} < D_t < 0.99\,\text{s}$ the number of resulting flashes reduces only slightly. Consequently, these parameters can be tuned during commissioning phase possibly with validation to ground based data, in order to produce reasonable flash stroke numbers, durations and sizes.

In the ATBD the terms 'fit-first' and 'fit-all' were used for different strategies of group association to flashes. Particularly the LIS processing used a 'fit-first' approach, while the GLM will apply the 'fit-all' method. The presented algorithm for MTG-LI with nearest linkage rule for the events in the groups is a 'fit-all' method by design.

2.4.4 Clustering of Proxy Data Sets

The group and flash clustering was applied also to proxy lightning event data, which were derived from ground based lightning location data. An example is shown on Fig.2.11, where positions of lightning flashes were detected by a ground based LLS over Scandinavia (NORDLIS). From these position data a set of events was simulated and processed by the clustering algorithm. The output flashes are in good agreement with the input positions.

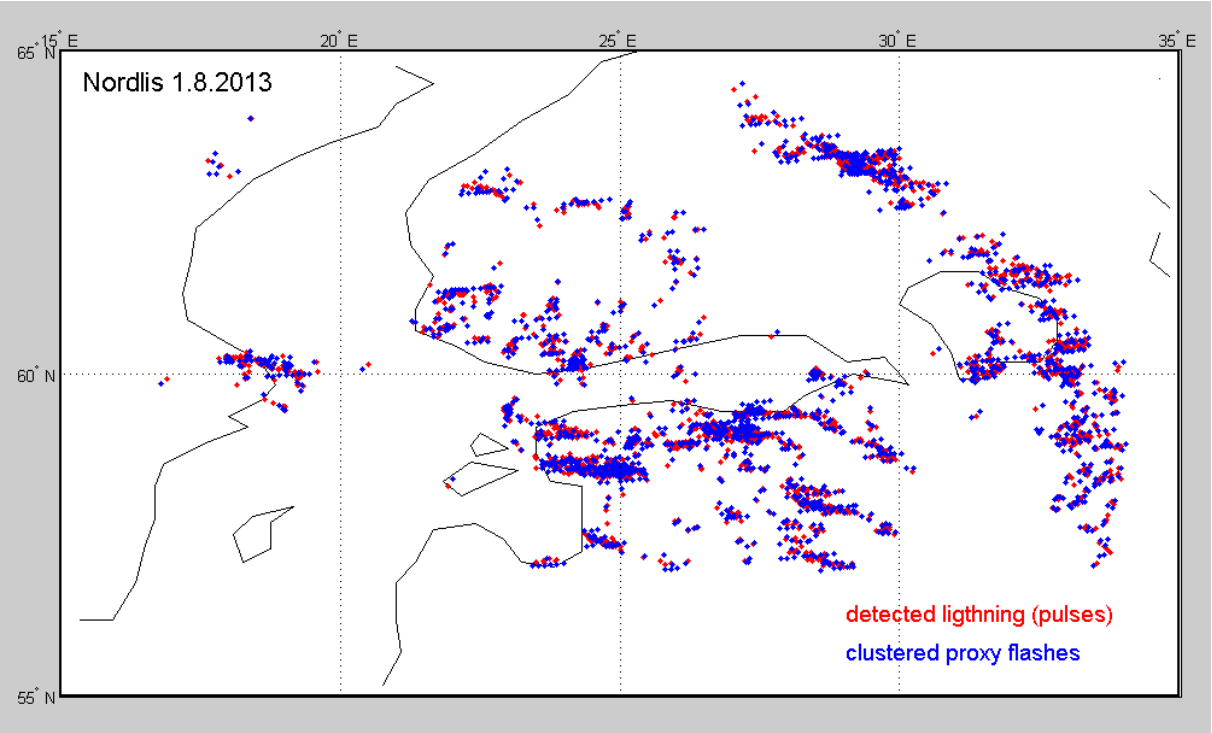


Figure 2.11: Flash clustering for proxy data. Lightning positions were detected by a ground based NORDLIS LLS (red dots) and used to generate proxy event data. The clustering of these events yield the displayed flash positions (blue dots).

2.5 Summary of Data Clustering

For the lightning data clustering algorithms a common cluster analysis approach was formulated with the central concept of a distance function quantifying the closeness and a linkage rule of 'nearest neighbor' adjacency deciding about inclusion into the cluster. For group clustering the adjacency of a lattice distance 1 in space and time defines the cluster. In flash clustering a real space and time distance limit is used and was found robust with respect to parameter variation. The algorithm was realized in simple flexible software modules for testing with various test data using detected LIS data and also simulated proxy data.

Both group and flash clustering procedures base only on event data. Hence, both clustering procedures can be performed independently, since in the presented flash clustering algorithm no group information is necessary. For optimized processing performance however, a pre-filtering of the groups may be advantageous.

Recommendations

- In group clustering the linkage rule adjacency should be applied both in space and time to allow for an interpretation of groups as optical pulses from single discharges.
- New data fields for groups can be introduced, which contain the time duration information. The calculation and interpretation of the group's footprint area parameter should be also adapted.
- In flash clustering the linkage criterion should test time and space distance limits separately. The weighted Euclidean distance (WED) gives no advantages and is physically hard to justify. An independent criteria for time and for space seems to be more appropriate.

3 Accumulated Products

3.1 Theoretical Evaluation

3.1.1 Product Definition

The event-group-flash hierarchy of clustered lightning data represents the inner structure of the lightning flash in form of point data which are irregularly distributed in space and time. The data are highly concentrated in time (< 1 s), but have relatively large size in space (> 8 km). Further information of the data is contained in parameter fields such as radiance, spatial extent, duration.

Besides these point data an additional product group is of interest with a larger time scale of several seconds and a lower spatial scale of few kilometers. This product should accumulate the lightning point data information for the time interval and present it on a grid which can be easily combined with the cloud imager (FCI) products.

General Description

As specified in the ATBD these accumulated products cumulate flash related quantities over intervals of 30 s in time and 2 km in space at the sub satellite point (SSP).

The 30 s accumulation time is intermediate between the LI detection frame (~ 1 ms) and the satellite repetition time of few minutes. This is still a high resolution in time to be useful in nowcasting for the detection of the lightning activity development in thunderstorms. In space it is a finer scale than the lightning detection i.e. adapted to the FCI-grid. The accumulated products are also dedicated to be processed further by the user.

The accumulated products base on the flash data which where established by the clustering procedure, and the calculation uses the events constituting the flash. This guarantees the use of the full available information for the accumulated products.

Definitions of the 3 Accumulated Products

The ATBD defines 3 accumulated products which result from the data accumulation over 30 second time intervals for each cell of the size 2 km×2 km at SSP by the following methods:

1. **Flash index**: count for the cell the number of coverings by a flash (named hereinafter **'flash area'**)
2. **Flash number**: add for the cell the portions on flashes
3. **Flash radiance**: add for the cell the radiance contributed by flashes

These definitions imply the following integral characteristics:

1. Flash area: integrates to the total area of the flashes (in cell units)
2. Flash number: integrates to the total number of flashes
3. Flash radiance: integrates to the total radiance of the flashes (in radiance units)

3.1.2 Mathematical Formulation

The accumulation is performed in time and space. The time integration period is very large compared to the detection time frame. Hence the single flash data appear as singular discrete data, which can be described by a delta-function in time. Thus integration over time results in a sum of the discrete contributions and is not explicitly shown in the following.

Accumulation to Flash Radiance

The spatial radiance distribution function for a flash is $b_m(\vec{r})$

Total energy of the flash is integration over whole space

$$B_m = \int_G b_m(\vec{r})\,\mathrm{d}\vec{r} \tag{3.1}$$

a series of lightning flashes is given by a sum of distribution functions

$$b(\vec{r}) = \sum_m b_m(\vec{r}) \tag{3.2}$$

The accumulation in a cell g_{ik} is the integration

$$\beta_{ik} = \int_{g_{ik}} b(\vec{r})\,\mathrm{d}\vec{r} = \sum_m \int_{g_{ik}} b_m(\vec{r})\,\mathrm{d}\vec{r} \tag{3.3}$$

sum over all cells is then the total energy

Accumulation to Flash Area

The spatial extent distribution for a flash is a boxcar function $a_n(\vec{r})$ being 1 inside the area and 0 outside

Total area of the flash is integration over whole space

$$A_m = \int_G a_m(\vec{r})\,\mathrm{d}\vec{r} \tag{3.4}$$

A series of lightning flashes is given by a sum of distribution functions

$$a(\vec{r}) = \sum_m a_m(\vec{r}) \tag{3.5}$$

The accumulation in a cell g_{ik} is the integration

$$\alpha_{ik} = \sum_m \int_{g_{ik}} a_m(\vec{r})\,\mathrm{d}\vec{r} = \sum_m \delta_{m \in g_{ik}} \tag{3.6}$$

this is a count over flashes which cover the cell. The sum over all cells is then the total of covered cells' area which equals the flashes' area.

Accumulation to Flash Number

The number distribution function for a flash is the normalized area function:

$$n_m(\vec{r}, t) = \frac{a_m(\vec{r})}{A_m} \tag{3.7}$$

with the box function $a_m(\vec{r})$ being 1 inside the area and 0 outside.

Total integration over all time and space yields

$$\int_G n_m(\vec{r})\,\mathrm{d}\vec{r} = \int_G \frac{a_n(\vec{r})}{A_n}\,\mathrm{d}\vec{r} = 1 \tag{3.8}$$

i.e. 1 count per flash.

A series of lightning flashes is given by a sum of distribution functions

$$n(\vec{r}) = \sum_m n_m(\vec{r}) \tag{3.9}$$

The accumulation in a cell g_{ik} and time interval τ is the integration

$$\nu_{ik} = \sum_m \int_{g_{ik}} \frac{a_n(\vec{r})}{A_n}\,\mathrm{d}\vec{r} = \sum_m \frac{\delta_{m\in g_{ik}}}{A_m} \tag{3.10}$$

this is a count over flashes which cover the cells, but each divided by flash area. The sum over all cells is then the total flash number.

3.1.3 Discussion of the Product Definition

Relation to Usual Lightning Density Calculation

The specified accumulated products differ from usual accumulation of lightning data which result e.g in a flash density per km^2. These flash densities are a typical output from regional and global ground based lightning location systems (LLS). They are usually calculated for larger accumulation intervals: typically for cells of 20 km×20 km in space and for time intervals of 1 hour, 1 year or similar intervals depending on the application. The source data resolution and accuracy (1 km, 1 ms) is much smaller compared with these integration intervals. Therefore, the accumulation procedure consists in counting the number of flashes hitting the corresponding cell during the integration time. Dividing the accumulated number by the area yields the density.

In contrast, as described above, satellite borne data from optical detection have a larger spatial footprint due to scattering in the cloud. This spatial extent has to be preserved in the accumulated product. The spatial footprint is larger than the spatial integration size, i.e. each flash contributes to several cells in the accumulation product. This makes it more difficult to interpret the data value in a single cell of the accumulated product.

Additionally, since the LI accumulated products are calculated for the satellite view projection, they are not delivered as densities, i.e. are not normalized by the area. This density calculation can be done after mapping to earth projection coordinates.

Resolution in Space and Time

The LI accumulated products will be used as basis for other derived products. They can be accumulated further, e.g. in time to generate hourly, daily, monthly or yearly distribution pictures. Or they can be

integrated in space for the generation of time series for certain regions. In order to preserve the necessary flexibility and data consistency for the these prospected use cases the product definition was kept elementary simply. The product definition guarantees also consistency with the other level-2 point data.

The accumulation time of 30 s is short for moderate storms with lightning flash rates of $< 20\,\mathrm{min}^{-1}$. In combination with the small spatial detection scale of 10 km the accumulated products will be very sparsely filled with data. Moreover, due to the low number of contributing flashes in each cells the values will have a low dynamic range and vary strongly between the few discrete values. This has to be considered when displaying the products.

Since the data gathering and accumulation is performed on the GEO observation grid, the ground projected product will have increasing distortions with nadir angle.

Resampling

The ATBD specifies a resampling step which maps the lightning data from the LI-detector source grid to the FCI target-grid. This mapping consists of a rotation and a sub-sampling step during which the data totals are preserved.

The ATBD formulates the rule that the data of the source cell has to be distributed equally to all target cells which are at least to 50% covered by the source cell. The option to modify the threshold value of 50% is given. This rule needs to be completed by a treatment for target cells which are hit by the vertices of the source cells (see central cell in Fig. 3.1). These target cells are covered partially by 4 source cells and the 50% threshold may not be reached at all. This would leave the target cell empty.

For incompletely covered cells in principle an interpolation procedure would be possible. However any smoothing interpolation would change the data structure which has to be avoided.

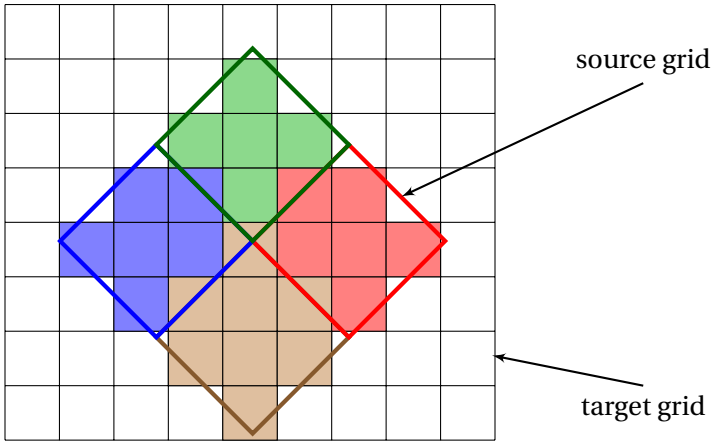


Figure 3.1: Mapping of the lightning data from LI-grid (source) to the FCI-grid (target) by the nearest neighbor interpolation.

Instead, a nearest neighbor subsampling rule is suggested (see Fig. 3.1): Each target cell gets its data from the closest source cell. (This rule applies also for empty source cells.) This nearest neighbor subsampling can be realized by a lookup table, since the source and target grids are unchangeable. The data from the source cell are uniformly distributed to the target cells.

In the general case the number of target cells which are associated to a source cell is not constant but may vary (see Fig. 3.1). This affects the data distribution to the target cells since the total of the data in both grids has to be conserved. This effect depends on the relation of the grid spacing of both grids, on the offset and rotation angle between the grids. It should be studied and minimized for the real grids in the instruments.

3.1.4 Individual Product Discussion

Flash Area

The flash area product contains the spatial extent of the lightning flash. The cell values are integer numbers with increment 1. On a short time scale this product is easy to interpret and to display in the context of storm development. It simply marks the cloud area with lightning. For a given cell the value indicates how frequent the cell was covered by a part of a lightning flash. For larger accumulation times the interpretation is more difficult.

Flash Number

The flash number product is related to the usual flash number density which can be calculated in units of flashes/km^2. The spatial integration over a certain region yields the total flash number in that region. As a consequence of the applied normalization a flash with large footprint area gives a small contribution to each covered cell. This characteristic is counterintuitive and complicates the interpretation of single cell values and also the display on small time scales. In contrast, for larger integration time this product approaches the flash number and is proportional to the lightning density.

Flash Radiance

The flash radiance product accumulates the energy for each cell. It is closely related to the energy in the optical range and also to the channel temperature. It is therefore in principle easy to interpret on any time scale. However, the radiation of the events is the quantity with largest uncertainty. It depends on processing details such as background subtraction method and detection threshold. It is affected by the scattering process and depends on cloud microphysics, on channel orientation and altitude. A straightforward relation of the cell values to the lightning emission energy or even the temperature seems to be very uncertain.

3.2 Investigation of Test Cases

This section compares the products against each other for lightning test data.

3.2.1 Accumulated Products Generation

For the investigation of test cases the accumulated products were generated by software modules operating on dedicated test data.

Software Modules

For the calculation of the accumulated products a sequence of software modules was created which performs the following steps. Input data are in the flash-event structure, where the events are given with discrete positions on the detection grid. Then for each flash the event data are accumulated for the 3 products on the detector grid for the 30 s time interval. The result is a grid of accumulated data. It follows the subsampling to the FCI grid, which is not realized here, since it has the same impact on all 3 products.

The products output are given on a grid in GEO satellite projection together with a georeference grid. This enables a convenient mapping to other projections for the various applications.

The software modules were coded in the Matlab language using the Mapping Toolbox for convenience. Since large grids have to be handled the software utilizes the sparse matrix data type. The Matlab-code is provided in the software package to this study.

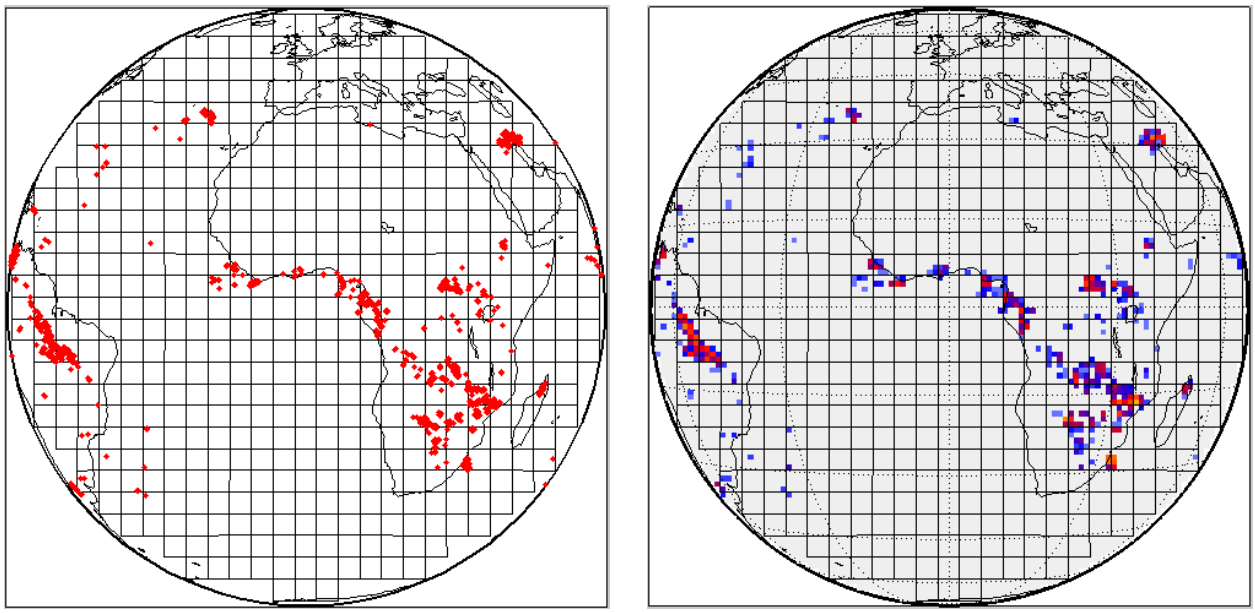


Figure 3.2: Generation of accumulated products. Left: Positions of flashes in the satellite field of view. Right: Accumulated flash area data in a grid. Note, that for display legibility the grid size is much enlarged here.

Test Data Preparation

For the test data two kinds of sources were used: LIS-data and proxy data from ground based lightning observations.

LIS data have the advantage of a realistic flash-group-event structure and realistic distribution of radiance values, temporal and spatial structure. Disadvantageous is the short observation period of 90 s, which is close to the time integration interval of the accumulated products. Hence a longer observation of an individual storm is not possible.

LIS-data are limited to the latitude range of ±35°. To overcome this limitation the data can be shifted artificially to other locations and mapped to the GEO FOV with same resolution of the detection grid.

Proxy data are generated basing on real lightning position data from ground based LLS. The method of generation of proxy data is described in Finke (2010). The position data (time and space) are seed for an artificial photon cloud which simulates the result of scattering in the cloud. These photons are input in a simulated detection procedure which outputs the stream of events. The detected events are subsequently clustered into groups and flashes as described in the previous chapter.

Demonstration of the Processing Modules

The Fig. 3.2-3.4 illustrate the processing steps of the accumulation product generation. LIS data from a whole day where used and processed as described above. For sake of better display the shown grid is chosen much larger (100 km footprint) than in reality.

3.2.2 Detailed Intercomparison of the Accumulated Products

In order to analyze the specific properties of the 3 products a special scenario with data from LIS was selected for a storm situation over Central Africa with several thunderstorm cells in the direction of the TRMM-satellite motion. By this setting a longer series of products could be generated for the region. It

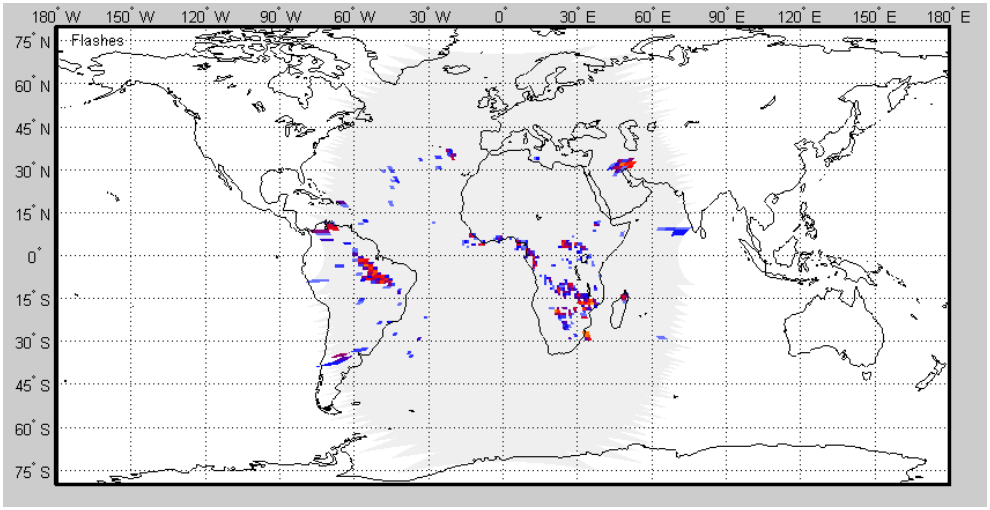


Figure 3.3: Accumulated flash area product mapped to cylindrical projection. Note the distortions at large nadir angles.

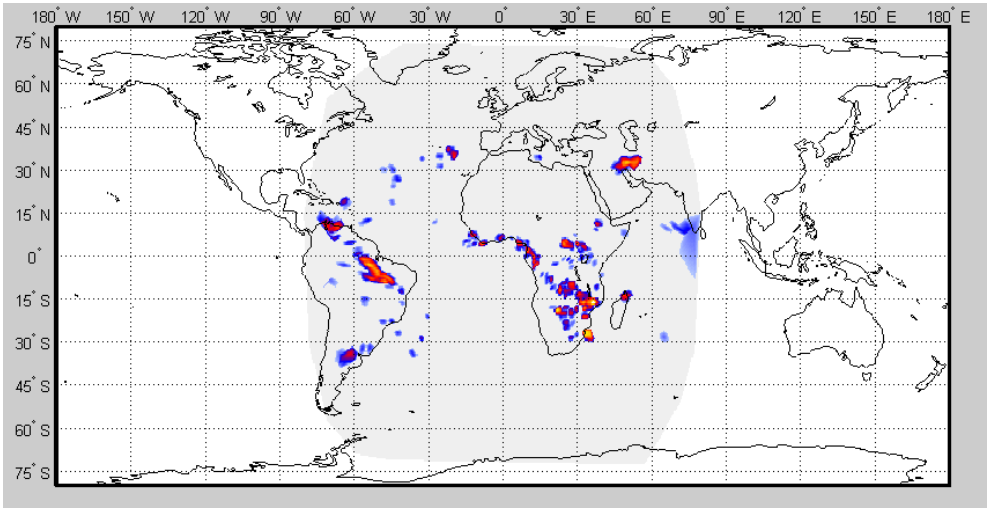


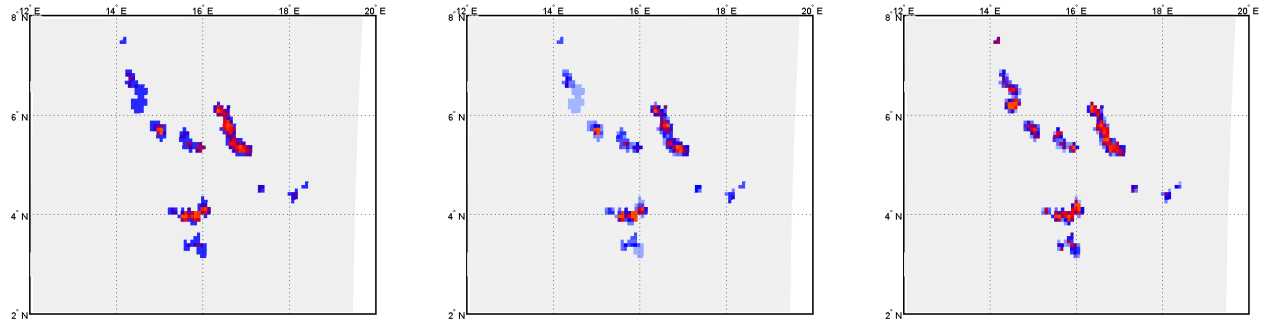
Figure 3.4: Accumulated flash area product transformed and interpolated to a regular longitude-latitude-grid with 0.5°cell size.

must be kept in mind, that the observation time of LIS is only 90 s over any location, hence the storms seem to disappear after short time. Using the described proxy data method longer series can be generated, however for the detailed intercomparison of the products the realistic lightning characteristics in LIS were preferred.

Comparison in 30 s Intervals

For a selected time and region the 30 s products were intercompared for several 30 s pictures. Figure 3.5 shows the 3 products for one of these 30 s intervals. The coloring of the products was chosen to make the appearance as similar as possible. The flash area and flash number are scaled linearly, while the flash radiance product was logarithmically scaled for better visualization.

The comparison yields the following results. The overall distribution of lightning is displayed well, all products show generally the same pattern. Flash area and flash radiance have a similar appearance. Notable is the lower dynamic range for the flash area. Looking at the fine structure it is found that the flash number shows for large flashes small values distributed over the flash footprint. This is evident e.g.



Flash areas Flash number Flash radiance

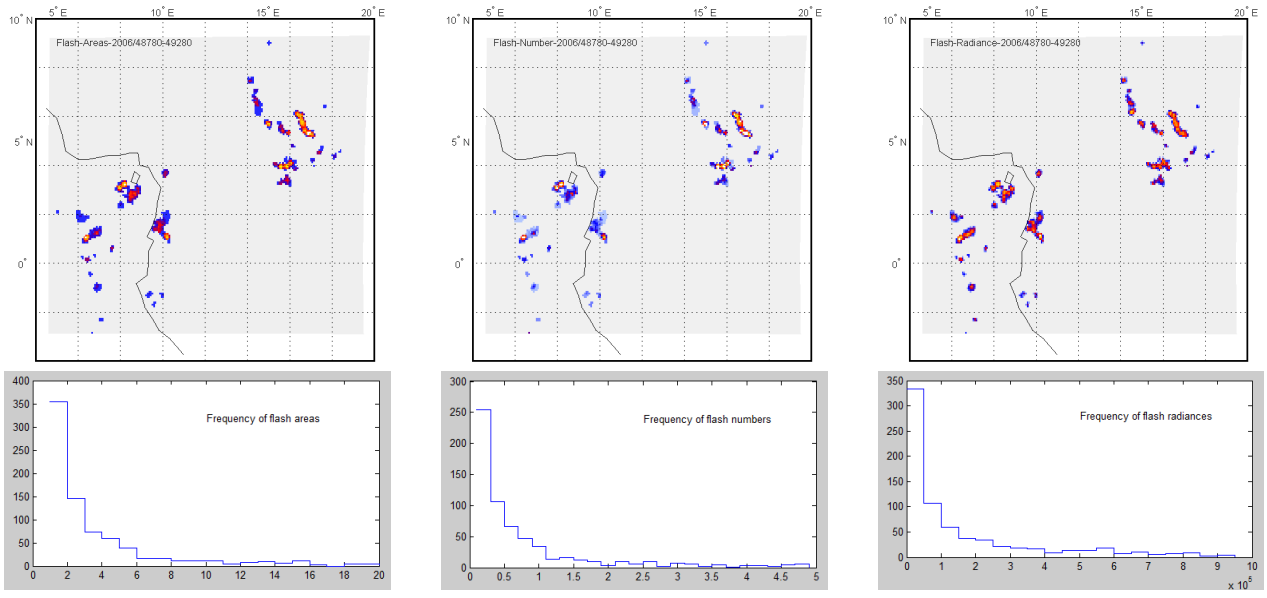
Figure 3.5: Accumulated products for a 30 s time period of storms in Central Africa with high lightning flash rate as detected by LIS.

in the southern parts of the structures at (16°E, 3.5°N) and (14.5°E, 6.5°N). The flash radiance can show very large values in single cells caused by single very bright events or groups.

Comparison in a 300 s Accumulation Period

Figure 3.6 shows the accumulated products for the same storm situation in the same area accumulated over a longer period of 300 s. The features described for the 30 s intervals are confirmed by the other lightning pattern in these pictures. Multi-stroke flashes can be distinguished by the large values in all 3 products. Flashes with large radiation appear prominently in the flash radiance product. Large area flashes produce characteristic faint structures at the outer parts of lightning signatures.

The frequency distribution of the data values for all the 3 products is shown by the histograms in the lower part of Fig. 3.6. The histograms were found to be very similar.



Flash areas Flash number Flash radiance

Figure 3.6: Accumulated products for a 300 s time period and a larger region of the same storm situation as in Fig. 3.5.

3.2.3 Accumulated Products for Proxy Data

In the following a demonstration of the accumulated product 'flash area' is given for a set set of LI proxy data which were derived from ground based lightning position data. The same case as in section 2.4.4 is

considered. The lightning position data were processed into LI proxy data as described in Finke (2010). These proxy data were then clustered into groups and flashes. Finally, from the events and flashes the accumulated flash area product was calculated. The result is shown on Fig. 3.7 for two accumulation intervals.

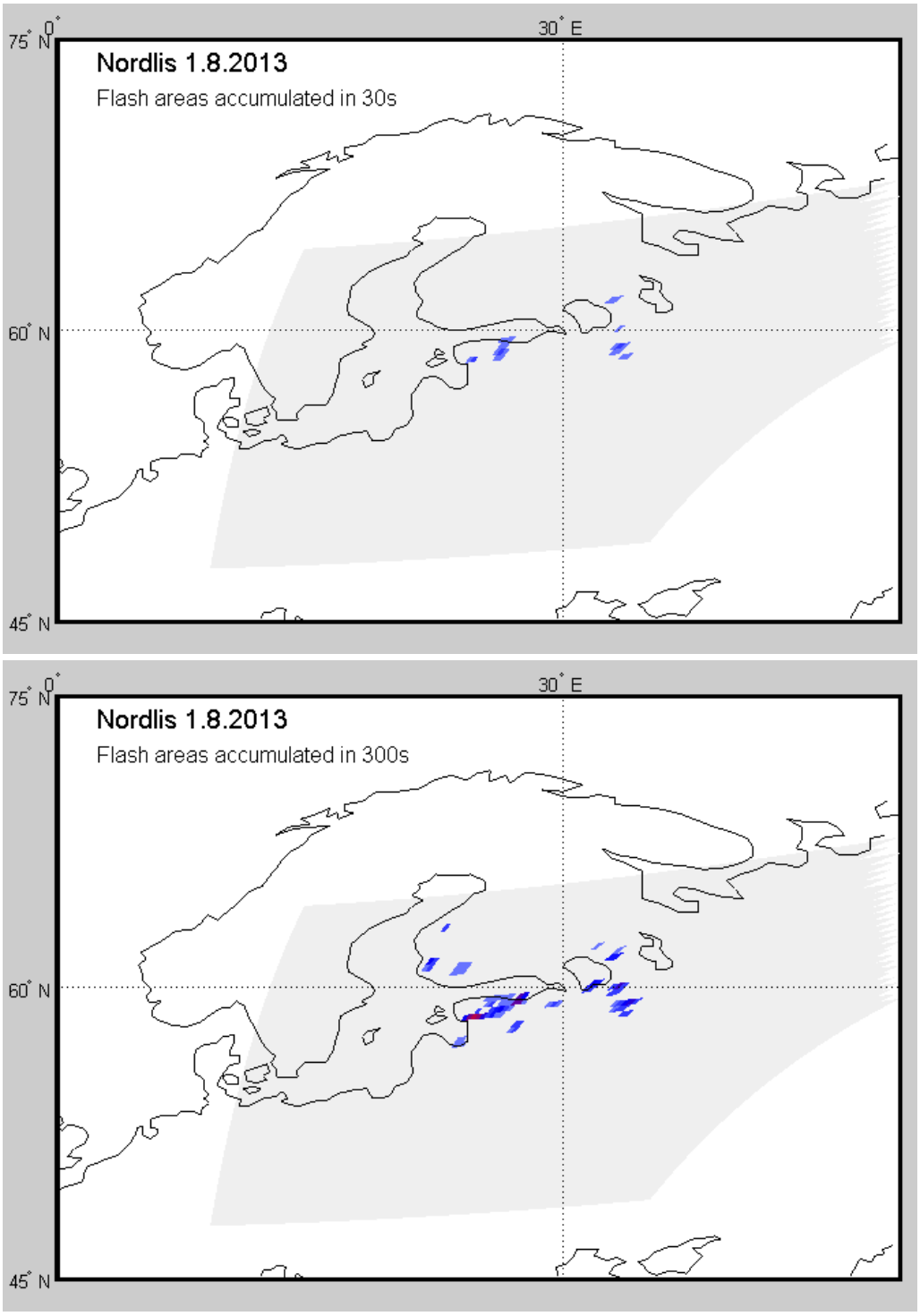


Figure 3.7: Accumulated flash area product for proxy data derived from Nordlis lightning observation on the 1st August 2013. The data were accumulated for 30 s (upper part) and for 300 s (lower part).

3.3 Summary of Accumulated Products

3.3.1 General Summary and Conclusion

The 3 accumulated products represent a useful data product type which bridges the time between satellite scans. They provide a summarized lightning information on an intermediate time scale which is easy to handle for the user. Due to the elementary structure of the accumulated products a processing towards further derived products such as densities or regional statistics is straightforward.

Methods and Parameters

The accumulated products use the established flash as the basic object since it is most robustly detected and less dependent on instrument characteristics. The time and space sampling is adapted to FCI images with 30 s and 2 km at SSP.

Subsampling to the FCI-IR grid is recommended to be performed by a nearest-neighbor subsampling since it does not smooth the data. The totals of the data must be preserved on the target grid.

Accumulation time of 30 s is relatively short, however necessary for the intended use for nowcasting applications. As a consequence the accumulated product matrices are sparsely filled and the data have low dynamics. E.g. for storms with moderate flash rate ($< 20\,\mathrm{min}^{-1}$) only few low values appear in the 30 s products. A longer accumulation is easy to realize and valuable for storm observation and tracking, for grid oriented lightning analysis and studies, and for climatological applications.

Dependency on Detection Conditions

The products will have different robustness against noise. The sensitivity to noise will be low for the flash number, intermediate for the flash area and high for the flash radiance.

Due to the detection geometry on the geostationary orbit all 3 products will exhibit distortions and large footprints at large nadir angles.

3.3.2 Characterization of the Products

Flash area is a clear and simple indicator of lightning activity, it marks the extent of the area covered by lightning. It will be very useful for nowcasting applications but probably less important for climatological accumulation.

Flash numbers is weighted by the flash area, therefore large flashes give a low contribution to each single cell. This characteristic may be less intuitive in nowcasting interpretation. However, the flash number product can be easily calculated into flash number density for comparison with other LLS and climatologies.

Flash radiance contains the total radiance per cell and is thus naturally linked to the flashes' deposited energy (heat) which is an important parameter in lightning physics and chemistry. The practical use however, might be hampered by the uncertain relation of detected radiance to the emitted channel radiance. This product also depends on detection process parameters such as threshold adaptation, background processing and noise removal.

3.3.3 Prospected Use of Accumulated Products

The accumulated products will be the basis for many statistical and climatological derived quantities, where direct use of the flash point data is not needed. These derived products could be

- **Density** of flash number and flash radiance: making it comparable to other lightning location systems.
- **Further Accumulation** for defined time and space intervals, e.g.
  - Time series of the lightning products for certain regions.
  - Daily, monthly, annual flash numbers for certain cells or regions, lightning climatologies
- **Combination** with other data from satellites or ground stations

4 Parallax Correction

The satellite based observation of lightning on the top surface of clouds is affected by the parallax effect, which results in a dislocation of the ground mapped lightning position – the parallax error. The ATBD describes the general approach to parallax correction of the level-2 products. This approach is investigated in this chapter.

4.1 Mathematical Formulation

In the following the formulas for the satellite projection and the parallax effect will be given. For sake of simplicity the 'vertical perspective projection' formulas for a spherical Earth (see e.g. Snyder, 1987) will be used. For more detailed studies and operational purposes the EUMETSAT formulas (e.g. CGMS, 1999) for the Earth ellipsoid are to be applied.

4.1.1 Geostationary Projection

The geostationary projection is a special case of the 'vertical perspective projection' corresponding to the view from a satellite above the equator with an orbit period equal to the (sidereal) rotation period of the Earth. The distance of the geostationary orbit from the center of the Earth is 42 164 km. For the Earth radius R of 6 378 km (at the equator) the geostationary satellite altitude H is 35 786 km (see Fig. 4.1).

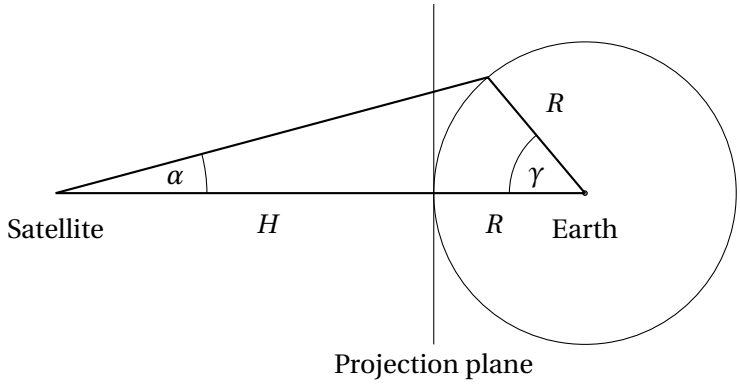


Figure 4.1: Vertical perspective view from satellite at an altitude H for spherical Earth with radius R. The angles are: α – nadir angle and γ – angular distance from SSP.

The driving parameter of this projection is the ratio of satellite altitude H and Earth radius R. In Snyder (1987) the projection parameter P is defined as the ratio between distance to sphere center and sphere radius by

$$P = \frac{R+H}{R} = 1 + \frac{H}{R} \qquad = 6.61 \quad \text{(for GEO)} \tag{4.1}$$

The projection has rotational symmetry around the axis of view. Any point on the Earth surface with an angular distance γ to the sub-satellite position (SSP) is projected to a nadir angle α by the relation

$$\tan\alpha = \frac{\sin\gamma}{P - \cos\gamma} \tag{4.2}$$

The inverse relation is

$$\sin\gamma = \left(P - \sqrt{1-(P^2-1)\cdot\tan^2\alpha}\right)\cdot\frac{\tan\alpha}{1+\tan^2\alpha} \tag{4.3}$$

The visibility limit depends on the satellite altitude and is given by

$$\cos\gamma > \frac{1}{P} \tag{4.4}$$

for the GEO this gives $\gamma = 81.3°$ and a corresponding nadir angle of $\alpha = 8.7°$

Geographic Coordinates

Positions on the Earth surface are given usually in the geographic coordinates longitude λ and latitude ϕ. For Meteosat the SSP is at latitude $\phi_0 = 0$ and at longitude $\lambda_0 = 0$. Therefore, the SSP distance angle γ calculates for any (λ,ϕ) position by

$$\cos\gamma = \cos\lambda\cdot\cos\phi \tag{4.5}$$

The mapping to the projection plane gives x- and y-distances in that plane. Alternatively, here a horizontal and a vertical viewing angle from satellite is defined by $\tan\xi = x/H$ and $\tan\eta = y/H$. These angles ξ and η can be directly related to the index positions in the detection pixel matrix.

For any ground position (λ,ϕ) these viewing angles are given by

$$\tan\xi = \frac{\cos\phi\cdot\sin\lambda}{P-\cos\lambda\cdot\cos\phi} \tag{4.6}$$

$$\tan\eta = \frac{\sin\phi}{P-\cos\lambda\cdot\cos\phi} \tag{4.7}$$

The inverse formulas are

$$\tan\lambda = \frac{\tan\xi}{\sqrt{\tan^2\xi+\tan^2\eta}}\cdot\tan\gamma \tag{4.8}$$

$$\sin\phi = \frac{\tan\eta}{\sqrt{\tan^2\xi+\tan^2\eta}}\cdot\sin\gamma \tag{4.9}$$

where the angular distance angle γ follows from the Eq. (4.3) with the nadir angle α substituted by $\tan^2\alpha = \tan^2\xi+\tan^2\eta$.

4.1.2 Parallax Error

A parallax error appears when a signal originating from a cloud with altitude h is wrongly treated as arriving from the Earth surface (Fig. 4.2). Mathematically this corresponds to the use of the wrong Earth radius: The signal was emitted at $R+h$, while the projection parameter P in Eq. (4.4) was calculated with R in the denominator. In the result a distance angle γ is calculated, while the true distance angle of the emission point was γ_0.

$$\gamma = \gamma_0 + \Delta\gamma \tag{4.10}$$

The parallax error $\Delta\gamma$ is the difference between the mapped angle γ and the true angle γ_0. It is positive, i.e. the position appears with an enlarged angular distance from SSP.

The correct inversion formula (4.3) should use the parameter $P_0(h)$

$$P_0(h) = \frac{R+H}{R+h} = \frac{P}{1+h/R} \quad \approx P\cdot\left(1-\frac{h}{R}\right) \tag{4.11}$$

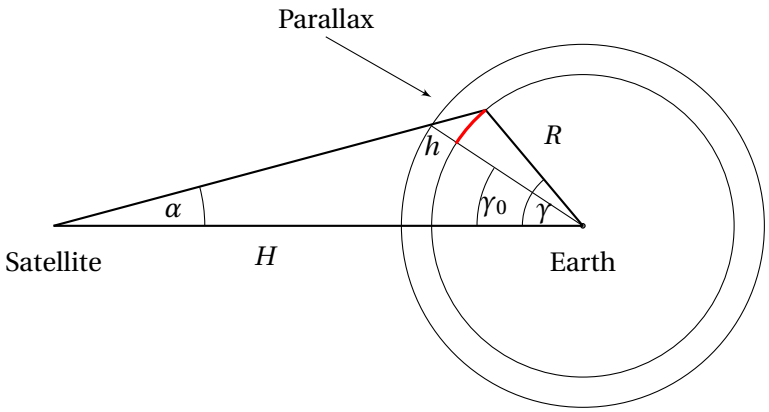


Figure 4.2: Parallax appears for signals originating from elevated cloud layers with altitude h. The incorrect back projection yields γ, while the true angle was γ_0.

The parallax error can be expressed by the above projection formulas. In practical cases for thunderstorm cloud heights the ratio h/R is always small. Therefore, the expression for the parallax error can be linearized to

$$\Delta\gamma(\gamma) \approx \frac{P \cdot \sin\gamma}{P \cdot \cos\gamma - 1} \cdot \frac{h}{R} \tag{4.12}$$

Substituting the angle γ by the Eq. (4.5) yields the parallax error as function of longitude and latitude

$$\Delta\gamma(\lambda,\phi) = \frac{P \cdot \sqrt{1-\cos^2\lambda \cdot \cos^2\phi}}{P \cdot \cos\lambda \cdot \cos\phi - 1} \cdot \frac{h}{R} \tag{4.13}$$

and in the same linear approximation follow the longitude and latitude parallax errors:

$$\Delta\lambda = \lambda \cdot \frac{\Delta\gamma}{\gamma}, \qquad \Delta\phi = \phi \cdot \frac{\Delta\gamma}{\gamma} \tag{4.14}$$

The parallax error can be expressed as a distance value multiplying the angular error by the Earth radius $\Delta s = \Delta\gamma \cdot R$. Example data for various cloud top altitudes are shown on Fig. 4.3. The geographic distribution of the parallax error for a cloud altitude of 12 km is displayed on Fig. 4.4. Table 4.1 lists the exact parallax error distances for 12 km cloud height.

SSP distance angle γ	0°	10°	20°	30°	40°	50°	60°
parallax error Δs (in km)	0	2.5	5.2	8.4	12.5	18.7	29.6

Table 4.1: Parallax distance error Δs for a cloud altitude of 12 km for various distance angles γ from SSP.

4.2 Correction of the Parallax Error

To correct the parallax error one can use the correct inversion formula or, alternatively, subtract the error term from the incorrect value γ

$$\gamma_0 = \gamma - \Delta\gamma(\gamma, h) \tag{4.15}$$

In any case the correction method requires the input of the cloud top height which is generally not symmetric to the SSP but a function of the geographic location $h(\lambda,\phi)$. Therefore, in reality the correction will be applied to the longitude and latitude values

$$\lambda_0 = \lambda - \Delta\lambda(\lambda,\phi;h(\lambda,\phi)) \tag{4.16}$$

$$\phi_0 = \phi - \Delta\phi(\lambda,\phi;h(\lambda,\phi)) \tag{4.17}$$

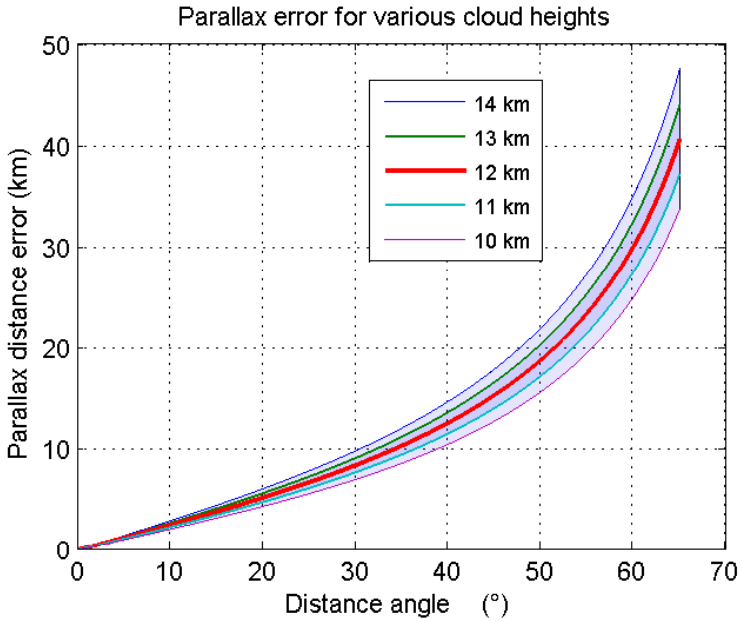


Figure 4.3: Parallax distance error for various emission altitudes as function of the distance angle γ from SSP.

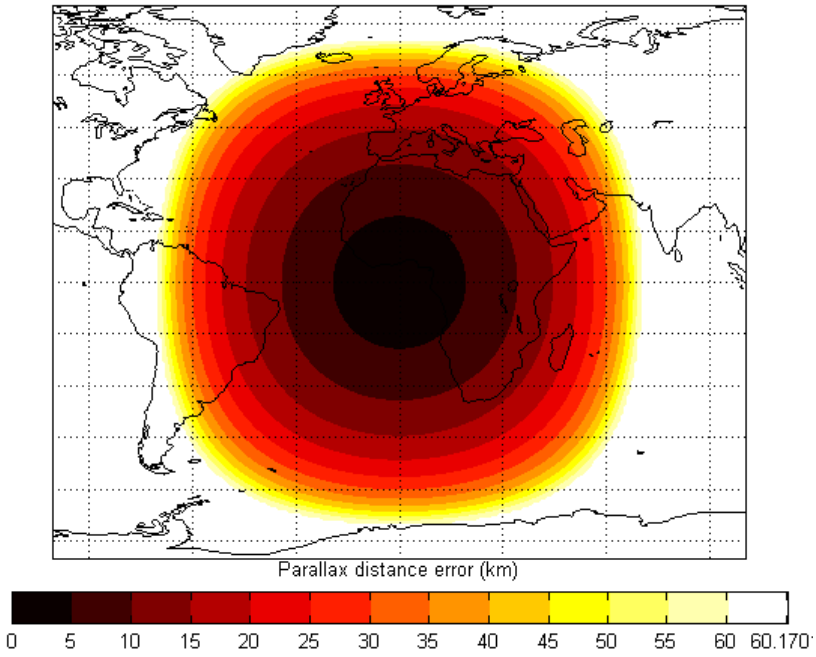


Figure 4.4: Parallax error distance as function of longitude and latitude for a cloud altitude of 12 km.

Parallax correction is an important topic in satellite based cloud observation and is in general not a trivial task. Image data are gathered originally on equally spaced grids. The parallax correction changes the position assignment of the grid cells for the whole image, i.e. the data values are mapped to an irregular grid. This may require special effort in the data interpolation.

For point data, such as lightning positions, a parallax correction is conceptually simpler since the data are not connected in space. In principle, for each of the points the longitude-latitude coordinates have to be corrected by the formulas (4.16-4.17) to new positions.

For the LI lightning data parallax correction is again more complicate since the data are not exact point data but have spatial extent, i.e. any group represents a small image.

4.2.1 Correction Strategy for Level-2 Products

The detected events, groups and flashes have all their geographic coordinates (longitude, latitude). Also the accumulated products are related to a geographic grid. Therefore, both product groups need a parallax correction which could be performed before or after the clustering procedure.

It is proposed to perform the parallax correction *after* the L2 product creation algorithms by the following reasons:

Group-Flash Clustering can be performed with uncorrected data:

- The group clustering restores the pulse, which was split by the detection process, to the shape as it was 'seen' by the detector. This should be done in the projection space by the criterion of adjacency on the discrete detection grid, thus before parallax correction.
- The flashes are built by clustering the grouped events by their ground projected distances. Hence, actually the flash clustering would have to be done with corrected position data. However, the flash clustering algorithm output will not change much for corrected or uncorrected positions since only differences in distance are considered.

The parallax correction procedure has to correct the longitude, latitude values of the events. Additionally, all position related group and flash parameters, such as centroid position and footprint area, have to be recalculated.

Accumulated products are provided as regular grids in the satellite view projection. They represent in fact 'images', which should be treated by the same strategy as the other satellite born cloud images. Consequently, a parallax correction is applied after product creation by the same methods and in consistence with the parallax correction of satellite images.

4.2.2 Discussion of Altitude Functions

The ATBD specifies 2 options for obtaining the source height information h for the parallax correction:

1. an assumed standard cloud top height which can be a constant or depend on time and geographic location. Also a land-ocean dependency can be considered.
2. the actual cloud top heights as determined from the FCI data.

Both approaches will be discussed in the following.

Standard Altitude Function

A standard altitude function can be constant or dependent on various parameters. This function has continuous support, i.e. can be calculated for any geographic position.

In the order of complexity the function can

1. h use a global constant height value which can be derived as an average value from statistics of thunderstorm cloud top heights.
2. $h(\phi)$ model a dependency on latitude, e.g. taking into account the higher cloud top altitudes in the tropics.
3. $h(\phi, t)$ introduce a dependency on time, e.g. considering seasonal changes.

4. $h(\lambda, \phi, t)$ depend in a more complicated way from geographic position, e.g. discriminate between land and ocean.

All these functions can be derived from cloud climatologies and/or conceptual models. Except for the last case, the function will be smooth, so the parallax correction will preserve the connectivity of the lightning pattern. The correction results in a shift of the whole pattern to a new position with slight distortions of the geometrical shape.

Despite the variety of possible altitude function, it is recommended to use as a basic approach a globally constant altitude (1), or a smooth dependency on latitude (2). This reduces the parallax error already significantly, while still keeping the parallax correction simple and comprehensible. More sophisticated functions may be used in special situations and case studies.

Figure 4.5 shows the remaining parallax error after correction with an assumed cloud top height of 12 km. The strong reduction of the parallax error is obvious (cf. with Fig. 4.3). The scatter of real thunderstorm cloud top heights around this standard value will produce only a small remaining parallax error. A cloud top of 12 km is a typical value for the mid-latitudes. For tropics and subtropics with higher clouds (≥15 km) the parallax error is less reduced, but it was already small.

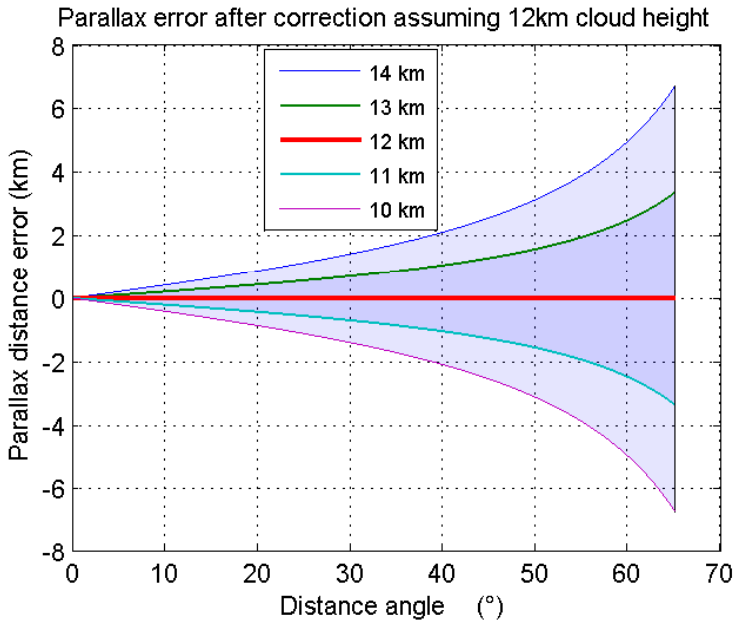


Figure 4.5: Remaining parallax distance error for various emission altitudes after parallax correction with a constant altitude function of 12 km.

FCI Cloud Top Height Product

The FCI will produce a cloud top height (CTH) product which can be naturally used for the parallax correction of the lightning products. Since the CTH is estimated for the real cloud field it will be generally a discontinuous function. Consequently, the parallax correction may lead to stronger distortion of the lightning group pattern.

The function will also have a discrete support $h(x_i, y_k)$, since it is defined on the FCI-grid. This allows for a direct parallax correction for each grid cell of the LI detection array. Due to the different sampling size and rotation angle an interpolation procedure will be necessary similar to the subsampling for the accumulated products discussed in section 3.1.3. Also interpolation with respect to time has to be conducted, since the lightning data time occur randomly in between the cloud image time frames.

Simulated Parallax Correction Satellite Derived Using Cloud Top Height A parallax correction of the lightning data using real cloud top heights is demonstrated in the following for a thunderstorm situation in Central Europe on the 28 July 2006. The MSG infrared cloud image in the IR 10.8 channel detected at 171241 UTC was used. For each pixel the brightness temperature was transformed into an altitude value assuming an adapted standard atmosphere temperature profile with a constant lapse rate of 6.5 K/km, a ground temperature of 20°C and no limiting tropopause. The resulting cloud top height image is shown on Fig. 4.6.

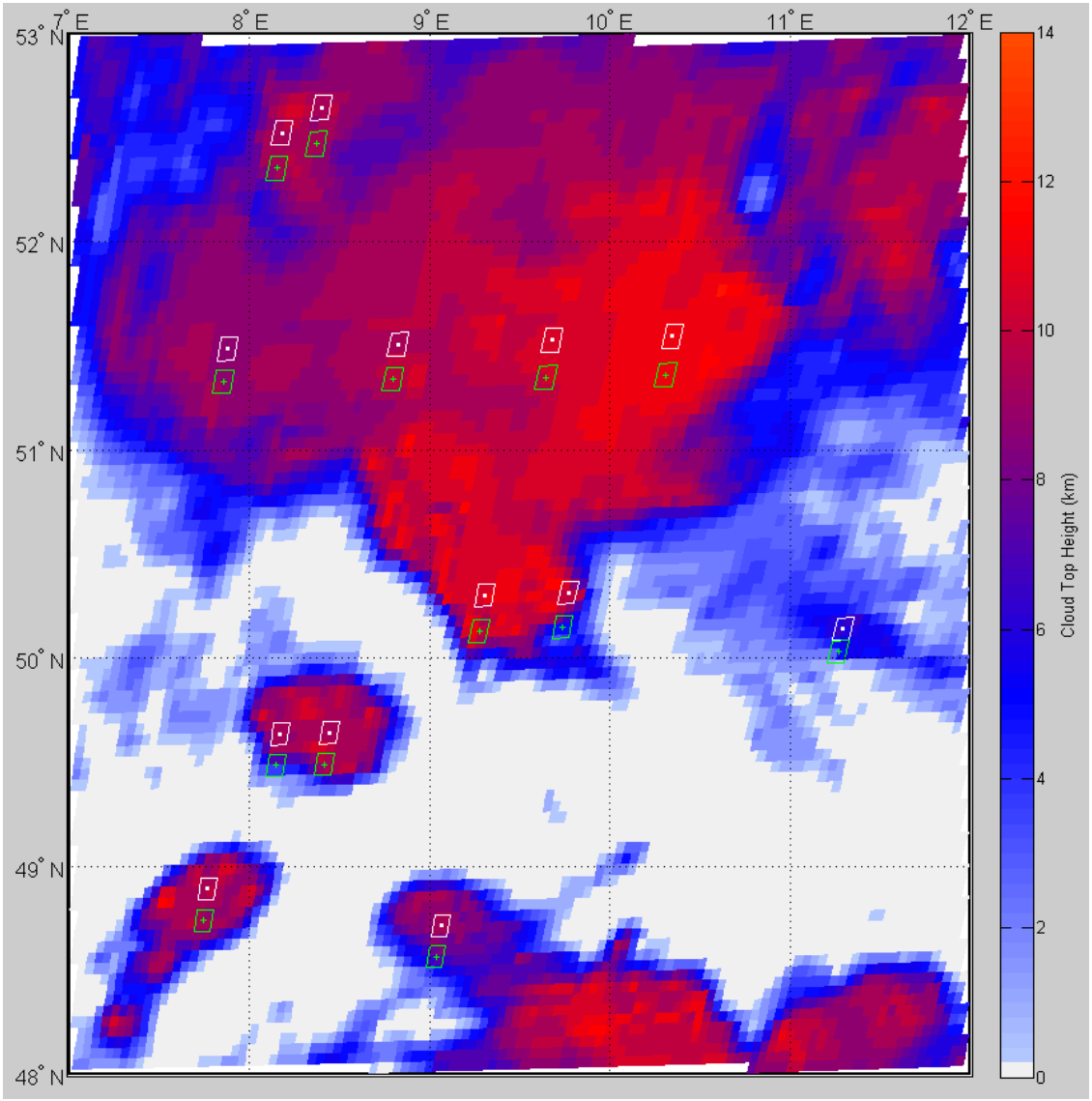


Figure 4.6: Demonstration of parallax correction using real cloud top heights for simulated lightning positions. For a storm scene over Central Europe on 28th July 2006 171241 UTC the MSG IR 10.8 image data was transformed into cloud top heights. The lightning positions are shown before (white) and after parallax correction (green). The quadrangles correspond to the pixel footprint of the LI, while the dots mark the pixel centers.

Lightning was overlaid to the image at artificial positions in clouds with different heights. For these lightning positions the parallax correction was performed with the corresponding altitude data derived from the IR 10.8 image. The resulting correct positions are shifted by different amounts. For lightning in the lowest cloud (easternmost with cloud top at 7.6 km) the correction shift is small, while for lightning in the large storm cloud with tops up to 12 km the shift is 20 km.
Note, that the satellite image is not parallax corrected, thus after correction the lightning positions do not fit to the clouds on the image.

4.3 Summary of Parallax Correction

The LI detects lightning on the top surface of clouds, therefore the derived lightning position data are shifted by the parallax effect towards larger SSP distance angles. A correction of this parallax error has to use information on the cloud top altitude. This information can be retrieved from another instrument on the MTG - from the Flexible Combined Imager, which will deliver a cloud top height product. Since this method uses the real cloud data, it can be expected to achieve the best parallax correction results after applying the necessary interpolation in time and space

Alternatively, for a stand-alone correction of the LI lightning data, an altitude function can be modeled. Here the assumption of a constant height of e.g. 12 km will already produce a significant correction. The remaining parallax error corresponds only to the variation of cloud top around the assumed mean height and will be much smaller than the original parallax error.

With respect to the level-2 product generation it is recommended to perform the parallax correction after the creation of the products. This makes both procedures independent and transparent. Users of the level-2 products can conveniently add their own algorithms.

The parallax correction of the LI lightning data is essential for any interpretation in the relation to the ground position. For the combination with other satellite borne products many applications will work completely satisfying with both uncorrected data, as well as with both corrected (by the same model) data.

5 Summary and Conclusion

The present study evaluated the algorithm strategies for the Lightning Imager level-2 products - Clustered point data and accumulated data. Basing on the products specification and algorithm description in the ATBD the various versions and parameter variations have been theoretically discussed and explored for test data. Additionally, possible parallax correction methods for the level-2 data were discussed.

The clustering of the lightning event data is divided in two steps: 1. pixel events are clustered in groups, which represent an optical pulses, 2. groups are clustered in flashes.

Adjacency in both time and space is the criterion for group clustering, since optical pulses are continuous in time and space. For flash clustering the short sequence in time and the closeness in space suggests the nearest neighbor criterion, i.e. the gap between groups has to be below a threshold value. Both time and space thresholds have to be satisfied independently. This clustering algorithm strategy was realized in simple flexible software modules for testing with various test data using detected LIS data and also simulated proxy data.

Statistical analysis of test data derived from LIS events confirm the validity of these clustering criteria. Compared to LIS the number of groups reduces significantly, since groups adjacent in time are joined now. As a consequence of joining groups also by adjacency in time a new data field has to be introduced, which describes the group duration. Also for some of the existing fields the interpretation has to be changed e.g. for footprint area and child count.

In flash clustering the variation of the threshold values (above a minimum of 330 s and 5 km) changes the number of resulting flashes only slightly. Hence most groups of a flash are well localized and separated from other flashes, so the established thresholds can be used for cluster building.

Both group and flash clustering procedures operate only on event data. Hence, both clustering procedures can be in principle performed independently.

The level-2 accumulated lightning products are created from the events on a per flash basis for time intervals of 30 s and spatial sampling size of 2 km. Thus, these products represent aggregated lightning information on a grid and on scales which are compatible to the FCI cloud image products. The 3 products accumulate different parameters of the lightning flashes:

The flash area index product counts how frequent a cell was covered by the area of a lightning flash. This product gives a clear picture of current lightning distribution which is easy to interpret and to use in now-casting.

The flash number product counts for each cell the portions of flashes it was covered by. This product integrates to the total flash number and will be well suited in analyses of lightning distribution on storm scale and in climatology.

The flash radiance product accumulates for each cell the radiance. It thus contains the optical energy emitted by lightning flashes from the cloud tops. It will be useful in lightning physics and chemistry applications.

The 3 accumulated products reflect different aspects of lightning distribution on a short time scale. They can be further accumulated for to climatologies of lightning distribution or other derived products.

Due to the observation geometry a parallax error appears symmetrically around the SSP, which can reach large values for larger nadir angles. This error can be corrected applying the correct inversion formula with the emission height.

Since the lightning optical radiation is always originating from the high cloud tops, a correction procedure which uses a constant reasonable mean height will already produce a significant correction. The remaining parallax error, will be small and in the order of the sample footprint.

A standard grid or parallax correction function/table can be supplied for each longitude-latitude pair basing on a standard altitude function. More sophisticated algorithms for parallax correction can be developed and tested using the cloud top height product of the FCI.

With respect to the level-2 product generation it is recommended to perform the parallax correction after the creation of the products. This makes both procedures independent and transparent. Users of the level-2 products can conveniently add their own algorithms.

Bibliography

ATBD-LI (2012). Algorithm Theoretical Basis Document (ATBD) for L2 processing if the MTG Lightning Imager data. Technical report, EUMETSAT. 100 pp.

CGMS (1999). LRIT/HRIT Global Specification. Technical report, EUMETSAT, CGMS 03. 61 pp.

Christian, H. J., R. J. Blakeslee, and S. J. Goodman (1989). The detection of lightning from geostationary orbit. *J. Geophys. Res. 94*, 13,329–13,337.

Christian, H. J., R. J. Blakeslee, S. J. Goodman, D. A. Mach, M. F. Stewart, D. E. Buechler, W. J. Koshak, J. M. Hall, W. L. Boeck, K. T. Driscoll, and D. J. Bocippio (1999). The Lightning Imaging Sensor. In *11th International Conference on Atmospheric Electricity*, June 7-11, Guntersville, Alabama, pp. 746–749.

Finke, U. (2010). liproxy - Software for the generation of artificial lightning proxy data for the MTG Lightning Imager. Technical report, EUMETSAT. 22 pp.

Snyder, J. P. (1987). *Map Projections–A Working Manual.* U. S. Geological Survey Professional Paper 1395. Washington, DC: U. S. Government Printing Office.

List of Figures

2.1 Optical signature of lightning observed on the cloud top 6
2.2 Scheme of the simple clustering algorithm steps 7
2.3 Lattice distances chessboard and manhattan 9
2.4 Histogram of time differences between groups and group duration 10
2.5 Comparison of group statistics for LIS and for the new processing 10
2.6 Scheme of flash clustering 14
2.7 Frequency distribution of time differences between groups 14
2.8 Flash Examples: Flashes at close spatial distance and large time distance 15
2.9 Flash Examples: Flashes overlapping in time at different positions 15
2.10 Distribution of group number in flashes for variation of the thresholds for distance and time 16
2.11 Flash clustering for proxy data 17

3.1 Mapping of the lightning data from LI-grid to the FCI-grid 21
3.2 Generation of accumulated products 23
3.3 Accumulated flash area product mapped to cylindrical projection 24
3.4 Accumulated flash area product interpolated to a regular longitude-latitude-grid 24
3.5 Accumulated products for a 30 s time period 25
3.6 Accumulated products for a 300 s time period 25
3.7 Accumulated flash area product for proxy data derived from Nordlis lightning observation 26

4.1 Vertical perspective view from satellite 29
4.2 Parallax appears for signals originating from elevated cloud layers 31
4.3 Parallax distance error for various emission altitudes 32
4.4 Parallax error distance as function of longitude and latitude 32
4.5 Remaining parallax distance error for various emission altitudes 34
4.6 Demonstration of parallax correction using real cloud top heights 35

Abbreviations

ATBD	Algorithm Theoretical Basis Document
CCD	Charge Coupled Device
CTH	Cloud Top Height
DE	Detection Efficiency
FCI	Flexible Combined Imager
FOV	Field of View
GEO	Geostationary Orbit
GLM	Geostationare Lightning Mapper
L2	Level 2
LI	Lightning Imager
LIS	Lightning Imaging Sensor
LLS	Lightning Location System
MSFC	Marshall Space Flight Center of NASA in Huntsville, Al
MTG	Meteosat Third Generation
NASA	National Aeronautic and Space Administration
NOAA	National Oceanic and Atmospheric Administration
OTD	Optical Transient Detector
SSP	Sub-Satellite Point
WED	Weighted Euclidean Distance
TRMM	Tropical Rainfall Measurement Mission