Master Thesis

Universiteit Utrecht

Koninklijk Nederlands Meteorologisch Instituut

Intercomparison of the FLITS and KLDN with ATDnet Lightning Detection Systems in the Netherlands, with a case study on the potential of infrasound lightning detection

---

Author:
Lotte de Vos

Supervisors:
Dr. Aarnout van Delden
IMAU
Dr. Hidde Leijnse
Dr. Maurice Schmeits
KNMI

April 29, 2015

Royal Netherlands
Meteorological Institute
Ministry of Infrastructure and the Environment

University of Utrecht

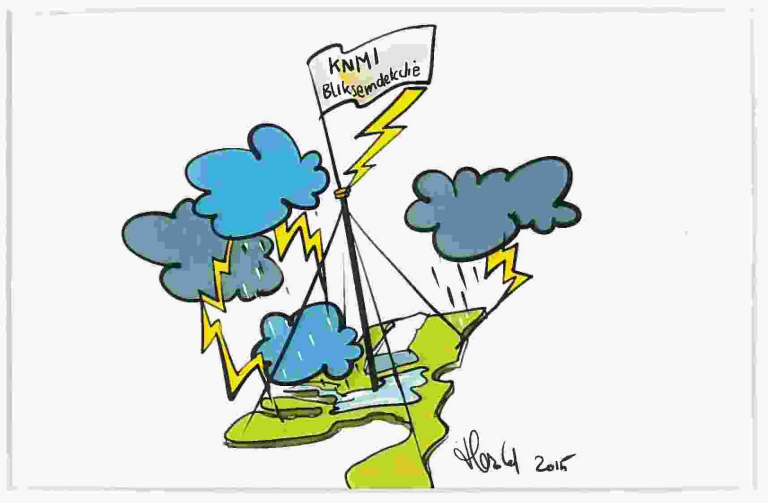
Contents

Abstract

Because of the risks to public safety and property, meteorological services actively monitor lightning events. The generated lightning detections are used by insurance companies, air traffic controllers and other interested parties. The detection databases can also be used for scientific purposes and as input for weather models, since lightning can be considered as a proxy for associative severe weather conditions like heavy rain, gusts and hail.

KNMI's FLITS lightning detection system will be replaced by the KLDN system from the external company Météorage. In order to anticipate on differences in lightning detection due to the use of a new system, historical datasets are analyzed. The relative performance of FLITS, KLDN and ATDnet, a lightning detection system used by UK Met Office, is determined. The results will form a scientific base for KNMI to decide on the use of KLDN for existing applications like lightning monitoring in general, severe weather alarms and lightning activity detection in the Netherlands, especially around airports.

The results show large differences between the systems, especially on the sensitivity for cloud to cloud vs cloud to ground lightning. The use of KLDN will affect the flight warnings around airports. The weather alarm criteria are recommended to be revised.



1 Introduction

Lightning is one of the most perceptible weather phenomena. The incidental occurrence of it offers stark contrast to the damage and associated hazards it can cause. Direct impacts can be lethal and/or cause fires. Even nearby lightning strikes can cause damage to electrical equipment. The increase of electrical equipment per household is one of the reasons financial damages by lightning in the Netherlands has increased to an estimated 35 million euro per year, according to Nationale Nederlanden [1]. Lightning can also be a significant hindrance to infrastructure, as airport regulation prohibits aircrafts to refule, take off or land during lightning storms. Additionally, lightning damages and resulting delays to the railroad system are not uncommon; in the period between 2001 to 2006, ProRail reported 1,556 lightning related damage incidents to the Dutch railroad system [7].

Because of the risks to public safety and property, meteorological services actively monitor lightning events. The generated lightning detections are used by insurance companies, air traffic controllers and other interested parties. The detection databases can also be used for scientific purposes and as input for weather models, since lightning can be considered as a proxy for associative severe weather conditions like heavy rain, wind gusts and hail [17].

KNMI has used the SAFIR network to monitor lightning in the Netherlands since 1995, which carries the name FLITS since improvements and soft-ware upgrades in 2004. KNMI has decided that the FLITS system can no longer be maintained, and plans to make use of the KLDN lightning data delivered by the external partner Météorage. These detections are derived from output of sensors of the EUCLID network.

Performance of lightning detection can vary between systems, especially if the systems make use of different types of sensors as is the case with FLITS and KLDN. For KNMI and its customers, it is important to understand differences in lightning detection output resulting from the switch to another system. In this report a comparison will be made between the lightning detections of FLITS and KLDN. The biggest challenge in validating a lightning detection system, is that it is nearly impossible to have a reliable ground truth. In a comparison between two systems, only the relative performance can be determined as both are approximations of the actual reality which is not known. There are several options to circumvent this in validating lightning detection systems. A ground truth can be found from railroad damage reports [7], satellite-based lightning imaging [24] or high speed camera images

in a lightning storm [20] and an intercomparison can be made with additional detection systems [10][19][18].

In order to examine the relative performance of KLDN compared to FLITS for the applications of lightning detection by KNMI, this last method is applied in this report. The performance of KLDN and FLITS will be compared to one another, as well as with the additional dataset of ATDnet, a UK lightning detection system covering a large area, including the Netherlands. The aim of this report is to anticipate on the differences in data output resulting from the use of a new lightning detection system. The results will form a scientific base for KNMI to decide on the use of KLDN for existing applications like lightning monitoring in general, severe weather alarms and lightning activity detection in the Netherlands, specifically around airports.

For this purpose, first the characteristics of lightning will be discussed in section 2.1. In section 2.2, the technique of detection, that makes use of some of these characteristics, by each detection system is discussed. Their relative strengths and weaknesses have been extensively described in literature, and will be reviewed here. In section 3 the approach of comparing datasets is discussed. This is followed by section 4 with the results regarding relative lightning detection in general, detection during severe weather and detection of lightning activity, leading to the conclusions in section 5.

A recurring theme in lightning detection validation is the challenge to find a reliable ground truth reference. In a stand-alone chapter 6, the potential of infrasound lightning detection is discussed by analyzing data from two infrasound array stations and the electromagnetic lightning detections of FLITS, KLDN and ATDnet during a case of severe weather in the Netherlands. This research is related to the report, but can be read separately.

2 Theoretical background

2.1 Lightning characteristics

The physical processes in lightning is extensively described by Poelman (2010) and Baba Rakov (2009), and is recapped in this section. Polarization in clouds occur because of mechanisms related with strong vertical currents and temperature gradients with height. Small positively electric charged particles at the bottom of the cloud move upward, while heavy negatively electric charged particles fall downward from the top of the cloud,

thus creating a static buildup. Only in sufficiently moist and large clouds, most often occuring in summer storms, the charge separation can build up enough to overcome the threshold value for air to become an electric charge conductor [17].

A charged area in the cloud can induce an opposite charge build up in a different (area of a) cloud or the earth's surface. One type of lightning that can occur is negative cloud to ground (CG) lightning. Figure **??** represents the whole process of a negative CG flash schematically. A stepped leader, a branched zigzag path from the charged cloud towards the ground surface, deposits negative charge along the way. The reasons for the chosen path are not well known but may be influenced by impurities/dust in the air resulting in slight variations of conductivity. The air may glow purplish as the stepped leader grows; this is the color of ionized air molecules. Note that the stepped leader is not the lightning strike itself but it defines the path the lightning strike will take.

When the stepped leader approaches the ground, the electrons in the already positively charged surface will be further repulsed. The induced charge becomes sufficiently large to overcome the threshold value and the upward leader from the ground can attach to the downward stepped leader. The connection of the extending plasma channels of the downward stepped leaders and the upward leader marks the beginning of the break-through phase. All the charge within the channel is floating at high speeds towards the ground. This so-called return stroke neutralizes the leader charge, though it may deposit excess positive charge onto the bottom of the cloud, where the stepped leader initiated, as well as along the leader channel. The current of the return stroke at the ground rises to values around 30 kA and decreases to half-peak value in tens of microseconds. This electric transfer heats up the channel up to over 30,000 K and the resulting expansion of air (over 10 atm) causes the thunderous sound.

Often, the first return stroke is followed by another leader that moves more gradually downward following the same path (ignoring the branches). This so called dart leader deposits a charge of approximately 1 Coulomb along the path, with a dart leader current peak of about 1 kA. Similarly as before, the dart leader connects with an upward moving return stroke, though this happens generally closer to the ground surface as the charge is smaller. The peak value of the subsequent return strokes are 10 to 15 kA and decay to half-peak in a few tens of microseconds. A downward negative lightning flash typically consists of three to five strokes, in which a stroke is the term

for when a downward leader connects with a return stroke, neutralizing the charge separation. Strokes of a single lightning flash follow the same path through the air, though it is possible for flashes to consist of only 1 stroke [4].

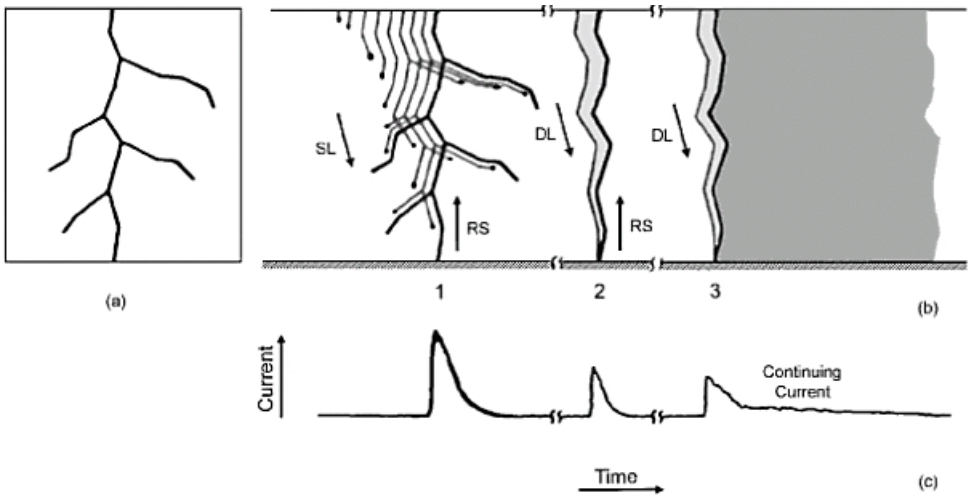


Figure 1: Schematic representation of all phases in a three-stroke CG lightning flash: a) still-camera image, b) streak-camera image, c) current change in time. SL = stepped leader, RS = return stroke, DL = dart leader. Figure from Poelman (2010) [17].

Subsequent return strokes in a flash are generally a factor of 2 to 3 smaller than the first return stroke, and follow with time intervals typically tens of milliseconds long, though any value between 1 and hundreds of milliseconds is possible. The duration of the flash in total is typically some hundreds of milliseconds, in which a charge of some tens of coulombs is exchanged [4]. The return strokes from the ground produce low or very low frequency signals (LF or VLF) of <300 kHz. Downward negative CG lightning flashes make up 90% of all CG discharges [17].

The counterpart of the negative lightning described here is positive lightning. In this case the positively charged stepped leader originates from the positively charged top of the cloud. The transfer of electrons is upward, exchanged between the cloud and the ground [17].

It is also possible for lightning to occur within or between clouds in so-called cloud to cloud (CC) lightning. These can be both positive and negative, i.e. positive or negative stepped leader. Instead of a return stroke

originating from the ground, it originates from an oppositely charged area of the same, or a different cloud. In this study, no distinction will be made between intercloud (discharges between two clouds) and intracloud (discharges within the same cloud) lightning. Detection of CG lightning is relevant for information on impacts and related risks in terms of damages and safety, though CC detection can be used as indicator on severe weather development, and is relevant for air traffic.

2.2 Lightning detection systems

2.2.1 Detection techniques

The previous section described the various motions of charge in a lightning event, all of which generate electric and magnetic fields associated with that motion. Besides the generally recognized optical and acoustic (i.e. flash and thunder) signals, lightning can be registered by electromagnetic pulses emitted by the processes in a lightning strike. The waveforms of return strokes and their propagation in time follow a recognizable pattern that has been studied extensively [4].

The frequency range of the electromagnetic pulses emitted during a lightning flash/stroke lie between 1 Hz and near 300 MHz. Both CC as CG lightning emit signals in the very high frequency or VHF (30 to 300 MHz) and the low frequency or LF (30 to 300 kHz), though CC is more easily discernable by VHF [23]. In LF lightning detection CC and CG signals can be distinguished based on decay time which is typically shorter in CC lightning [16]. In the very low frequency or VLF (<30 kHz), mainly CG lightning is detected. The expectation is that VHF sensors are best equipped to detect CC lightning, though it requires a higher density sensor network than LF and VLF sensors, and can suffer from blind spots, especially in mountainous areas [10].

A lightning detection system consists of several sensors, placed close enough together that multiple sensors register the signal originating from a stroke occurring in the covered area. The signal will arrive at different times at each sensor due to the time it takes to traverse the distance from the point of origin to the sensor. The combined information from the sensors yields the location. This is generally done with one of three methods.

Magnetic field direction finding (MDF) makes use of two vertical magnetic loops that are positioned orthogonally to each other. A magnetic field

passing through the coil induces a current in the loop. The strength of the currents through the two loops gives the direction from where the signal originated. The point where the lines from the sensors in the direction towards the signal intersect gives the location of the lightning discharge. In most cases two sensors are sufficient to determine the location of the stroke, with the exception of a lightning flash occurring anywhere on the line between two sensors. Therefore at least three sensors placed in a triangle are required for an accurate location. There is a back-azimuth error of approximately 1 degree in these types of sensors. This means that the location accuracy decreases for larger distances between sensor and stroke [17].

A different form of localizing lightning with direction finding is interferometry, that determines the direction of the source by solving the incoming wave signal at five pentagonally placed sensors. The combined information of multiple sensor arrays yields an intersection point. This is the location of the stroke, see figure 2. Interferometry measures in the VHF spectrum, and can therefore not measure reliably over large distances [17]. The most important difference between MDF and interferometry is that MDF measures the magnetic flux signal while interferometry makes use of the electrical field signal of a lightning stroke.

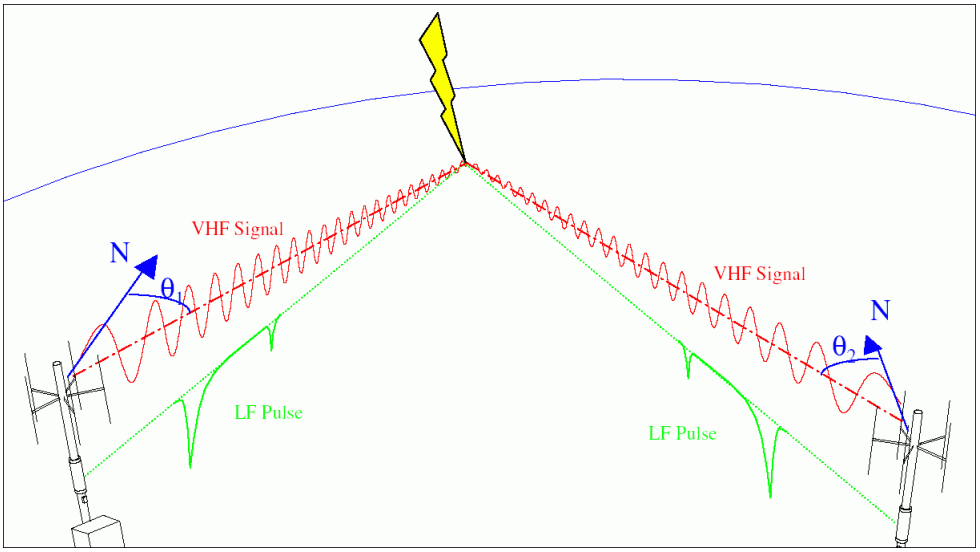


Figure 2: Location finding with interferometry. The direction of the lightning signal source is derived from the phase difference of the VHF wave signal in the five sensors placed close together. The combined information of the two sensors provides the intersection point, thus the location. Figure taken from Beekhuis [6].

The third method to locate lightning is the time of arrival (TOA) lightning location retrieval. The time it takes for the signal to arrive at sensors is dependent on the distance between stroke and sensor. Though the exact time the stroke transpired is unknown, the difference in time of arrival of the signal between the sensors around the stroke can be determined. This time interval gives the difference in distance the signal has travelled from the stroke location towards both sensors. Between two sensors, a hyperbola can be drawn that connects all locations with the correct distance difference to the two sensors. The intersection of the hyperbolas between all sensors that register the signal yields the correct location. As three hyperbolas can sometimes generate two intersection points (see figure 3), at least four hyperbolas, thus four sensors, are required to accurately determine the location of the stroke with TOA. TOA predominantly makes use of LF signals [17].

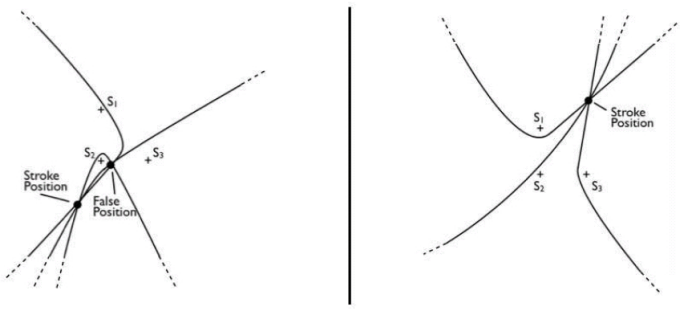


Figure 3: Location finding with Time of Arrival technique. Left is the situation in which three sensors generate multiple possible lightning locations, while right is the example of an unambiguous location with three sensors [17].

There is a large variability in lightning detection systems, making use of these techniques depending on the requirements of lightning data and availability of sensors in a network. Besides the type of sensors in detection systems, differences between detections result from data processing as well. The (often empirical) choice for a certain signal to noise threshold value determines the number of signals classified as lightning. Lower threshold values will lead to higher detection efficiencies at a cost of more false alarms.

2.2.2 FLITS

The data of the SAFIR/FLITS system is collected using four measuring stations in the Netherlands, together with three additional stations in Belgium.

The area in which it detects lightning is called the NL21-region, see figure 4, though in reality the coverage reaches not far beyond the KOUW-region, the area just spanning the Netherlands.

Lightning localization occurs with a combination of VHF (around 110 MHz) interferometer of five antennas in pentagonal shape and long-wave-antenna measuring LF (<4 MHz) for TOA [6].

FLITS can discriminate between CC and CG lightning based on wave form analysis of the LF with criteria concerning the amplitude, rise time and decay time of the system, where decay time is the most important factor in lightning type characterization. Though theoretically the VHF signals of CC and CG lightning differ from one another, the sensors used by FLITS are not able to use VHF data for correct CC and CG distinction [7].

FLITS detects the individual signals from segments in a CC flash, distinguishing first signal from sequential and last signals. Normally these individual discharges are grouped into one CC lightning event by FLITS software. From this grouping the mean location of the segments is given, as well as the length and duration of the sequence. The length of a CC flash segment sequence can be over 15 km. FLITS gives electric current values only for CG strokes.

The location accuracy of FLITS is approximately 2 km, with a detection efficiency (meaning the ratio of number of detected events divided by the number of real events that have occurred) of approximately 60% [17].

2.2.3 KLDN

Météorage is a French commercial organization that has been detecting lightning in Europe since 1987. The data selection as delivered to KNMI will be referred to as KLDN and spans the so-called NL25-region. The locations of the EUCLID sensors contributing to KLDN data and the NL25 region are given in figure 4.

KLDN makes use of several types of LF sensors, namely IMPACT, LPATS and LS7001. the IMPACT and LPATS are older and do not exhibit the same performance as LS7001, especially regarding CC detection [19]. All three types of sensors make use of TOA and Magnetic Direction Finding (MDF) in the LF spectrum (1 kHz - 350 kHz).

The waveform of the electro-magnetic signal is used to classify CC and CG discharges. The number of CC detections in 2011 to 2013 have been higher in the north of France compared to the Netherlands. This is attributed to the updated LS7002 sensors in France. Similar updates in two stations in Belgium and Denmark are scheduled. KLDN gives values for electric currents of CC as well as CG discharges.

The EUCLID network has been evaluated in Belgium with the use of high speed camera images as ground truth. Based on the comparison between lightning detection output of CG lightning and the known locations from the camera images of 210 strokes, the performance of the system has been evaluated. The EUCLID stroke detection efficiency was 84%. The median location accuracy varied between 0.2 km and 1.9 km [20]. A different validation in France yielded a detection efficiency of 90 to 95% and a location accuracy within 1 km everywhere [17].

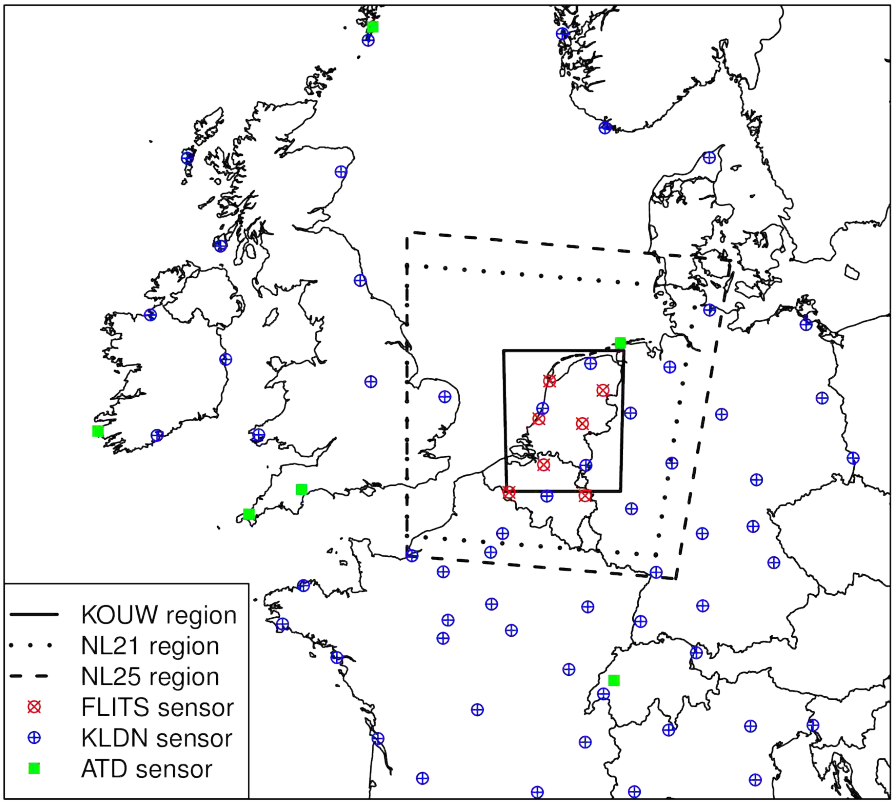


Figure 4: Map of sensor location of FLITS, KLDN and ATDnet, as well as the regions KOUW [22], NL21 and NL25.

2.2.4 ATDnet

ATDnet measures in the VLF range of 10 to 14 kHz [15], currently with 14 sensors worldwide. These VLF waves propagate in the ionosphere over very long distances (thousands of km), thus requiring few sensors in the detection network, see figure 4. The coverage extends over Europe, South America, Africa and central Asia. Because of the frequency range in which the sensors measure, ATDnet detects mainly CG lightning strokes, usually only the first return stroke. Many subsequent return strokes in a flash occur

too fast for the system to register, because of a 15 ms dead-time [17].

ATD stands for Arrival Time Difference since location finding is consequently done with TOA. The location accuracy has been estimated on 5 km over the United Kingdom as well as Africa, with an accuracy of 20 km in the rest of Europe. The detection efficiency in Western Europe and the UK is up to 90%, decreasing to 80% and 20% in respectively Northern Africa and Southern America [17]. In comparison studies in Belgium and France it was found that lightning detections drop during the night, more so than other detection systems. It was suggested this is caused by diurnal variations in the ionosphere, interfering with the wave propagation [19].

2.2.5 Previous performance studies

Lightning detection systems are challenging to validate as a large ground truth dataset reference is relatively difficult to obtain. In order to give insight in the value of lightning detection systems, various studies have been done in which the relative performance of several systems are determined by comparing datasets to one another or a differently obtained ground truth dataset. In order to get an idea of the relative performance of FLITS, KLDN and ATDnet, some relevant previous studies are discussed here. The same terminology as in the papers is used. The detection efficiency (DE) is the percentage of the truth reference (often obtained in a ground truth campaign) that is detected by the lightning detection system. Note that relative detection efficiency (RDE) and probability of detection (POD) are the same, and both indicate the percentage of the lightning detections in the reference system that is detected by the lightning detection system. A more detailed explanation of POD is given in section 3.3.

In a study carried out at KNMI, SAFIR/FLITS CG data has been validated with lightning related damage reports in the Netherlands from the NS (Dutch railways) over a period from 2001 to 2006, where the location and time of lightning damage is used as ground truth reference of a CG flash. Any case where the NS flash coincides with a stroke as reported in FLITS in the defined matching area and time range is labeled as a hit. The POD was calculated by dividing the number of hits by the total amount of damage reports (and multiplied by 100 for percentage). The averaged corrected POD was calculated as 57% with an uncertainty of 3% for a matching area with radiuses' in the range of 2-6 km [7].

Druë et al. (2007) examined the relative performance of SAFIR (of which

FLITS is derived) to BLIDS, which is part of the EUCLID and used by the German weather service for nowcasting of thunderstorms until 2006. The region covered in the study constitutes a large area of Germany, and the data that was used dates from May 2003 until December 2004. The (flash) total lightning (TL) POD of SAFIR out of BLIDS was 60%, with a (flash) TL POD of BLIDS out of SAFIR of 21%. This suggests that SAFIR is more sensitive to total lightning than BLIDS. This difference is mostly due to the fact that SAFIR detects more CC lightning. The SAFIR performance on CG flashes was lower than BLIDS. Also, SAFIR often seems to misclassify CC lightning as CG lightning [10].

Poelman et al. (2013) made use of high speed camera images as a ground truth for the validation of SAFIR, EUCLID, ATDnet and GLD360, an MDF based global lightning detection system operating in the VLF. Flashes were recorded on 3 days in August 2011 in Belgium. With high speed camera pictures, individual strokes of the same flash are registered. As strokes of a single flash follow the same path, the location accuracy can be determined as those strokes should be localized in the same location. As the ground truth is found with the pictures, the detection efficiency can be found by calculating what percentage of the actual strokes are found by the detection system. Note that FLITS is an updated version of the SAFIR system and KLDN data is derived from the sensors in the EUCLID network.

The SAFIR system making use of TOA with magnetic direction finding has a detection efficiency (i.e. the percentage of actual lightning detected by the system) of strokes of 64% and of flashes of 92%, meaning that while it misses some of the strokes, the detection of lightning activity i.e. flashes is quite good. The location accuracy of detection is 1.0 km. When considering only the output from the VHF sensors of SAFIR, the detection efficiency is approximately the same, but the location accuracy of detection is significantly worse, namely 6.1 km. ATDnet has a detection efficiency of strokes of 58% and of flashes of 88%. This relatively poor performance has been attributed to the low detection efficiency on one of the three days on which the thunderstorm occurred at night, which has a negative effect on the performance of ATD. The best performance was that of EUCLID with a detection efficiency of strokes of 84% and of flashes of 100%. The location accuracy of both ATDnet and EUCLID are good, 1.0 km and 0.6 km respectively [20].

Over the same period as the previous study, a comparison is done by Poelman (2011) between ATDnet and the operational and total lightning processor of the Belgium lightning detection system, which consists of SAFIR with

VHF interferometry plus LF sensors for CC and CG discrimination. The total lightning processor is extended with additional Vaisala sensors and the system localizes strokes with VHF interferometry and TOA. It was found that the POD values from ATDnet out of both these total lightning and operational detection systems were highest with respectively 53% and 62% [18].

In a different study of Poelman et al. (2013) the performance of a Météorage dataset (thus the same lightning detection network as KLDN) is compared with SAFIR and ATDnet. A distinction is made between the TL detection of Météorage and only the low frequency detection of SAFIR (which consists mostly of CG strokes). With the detection in Belgium between May and September in 2011 and 2012, hits are defined when two detection systems register lightning within a predefined temporal and spatial distance to one another. All systems are used as a ground truth in turn. The relative detection efficiency (RDE) in this study is what was earlier described as POD; the percentage of hits between two datasets as part of the total detections of the set which is used as truth reference. The results are given in table 1 below, in which TL Météorage is referred to as MTRG+ and only the CG detection is referred to as MTRG [19]. Note that low RDE values of a system could be caused by its poor detection, but also by a high false alarm ratio in the other system which is used as ground truth [19].

	Stroke RDE(%)	Flash RDE(%)
MTRG out of SAFIR	28	37
MTRG out of ATDnet	26	34
ATDnet out of SAFIR	27	60
MTRG+ out of ATDnet	40	57
SAFIR out of MTRG	34	46
ATDnet out of MTRG	47	80
SAFIR out of ATDnet	18	32
ATDnet out of MTRG+	39	69

Table 1: RDE (same as POD) values from Poelman et al. (2013) [19].

The values as presented in table 1 are not quite near 100%, which suggests a large variability in performance of detection systems. In the same study, the networks are compared on their ability to detect lightning activity. A hit is defined when two networks detect 1 or more strokes in a predetermined area of 10 * 10 km^2 over the span of a day. With this method, the resulting POD values range between 83% and 95%, with a false alarm ratio (fraction

of detections when ground truth found nothing) between 5% and 14% [19].

Though these values are all determined in a different region and time period than the current comparison, they can help provide insight in the performance of the detection systems in the Netherlands examined in this report.

3 Methodology

3.1 Flash conversion

It is important to take into account how detailed FLITS, KLDN and ATDnet each detect discharges connected to lightning. FLITS is capable of detecting individual segments of CC discharges, that are automatically combined into a single detection by FLITS software. Though KLDN detects CC lightning, it does not do so on a level where it can distinguish segments. Similarly, FLITS and KLDN can detect the first return stroke in CG lightning as well as several subsequent return strokes originating from the same lightning event, while ATDnet is assumed to mainly detect first CG return strokes [6] [17].

The terminology used in this report is as follows; all detections in the original datasets of FLITS in which the CC segments have automatically been combined by software, and the original datasets of KLDN and ATDnet are referred to as strokes. Strokes of the same flash are detected within a small time frame and distance from one another. By grouping the strokes within a time and distance limit the dataset is converted to a flash dataset, which can then be used to accurately compare system detection performance.

A methodology to investigate the grouping in lightning datasets was presented by Finke (1999), and later applied in similar lightning system comparisons by Drüe et al. (2007) and Poelman et al. (2013) [13] [10] [19]. The temporal and spatial distance between all events during the span of 24 hours with all other events in that period is calculated and plotted in a so-called single-point autocorrelation plot, see figure 5 and 6. These density plots of spatial and temporal distance between all strokes in a day are given for all three datasets in figure 6. Figure 5 shows the single point signal autocorrelation of FLITS before the grouping of the individual CC segments into strokes by the FLITS software.

The points in the region dt $<$0.01 s correspond with CC lightning, which

have discharges, i.e. segments, that follow each other rapidly. The region 0.01 s <dt <1 s shows activity that corresponds with CG lightning. All lightning activity in the region dt >1 s represent the correlations between strokes that are not part of the same flash [10]. In figure 5, a large amount of correlation is found in dt <0.01 s for the segments that are part of a single CC stroke, which are no longer visible in figure 6 as they are then combined in a single stroke.

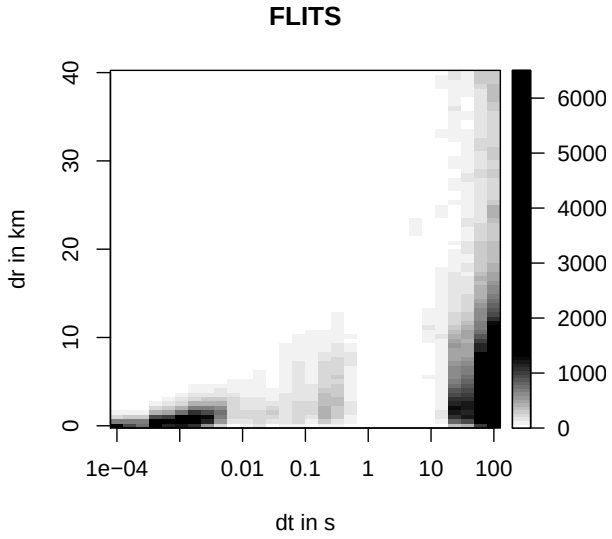


Figure 5: Single-point signal autocorrelation plots of strokes recorded by FLITS on 11 August in the NL21 region over the whole 24 hours, before grouping of CC segments.

In the FLITS dataset used in the report the CC segments are grouped into single strokes.

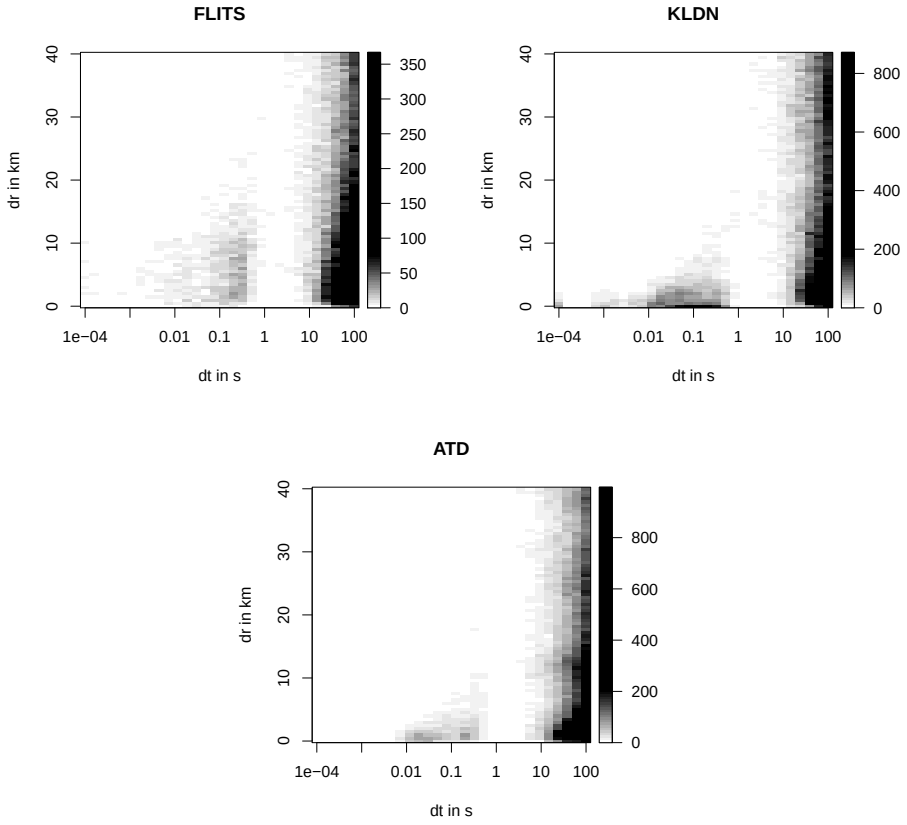


Figure 6: Single-point signal autocorrelation plots of strokes recorded by FLITS after grouping of CC segments, KLDN and ATDnet on 11 August in the NL21 region over the whole 24 hours.

The autocorrelation plots show that the spatial threshold may change over a relatively wide range without excluding strokes that show autocorrelation. This corresponds with literature, where there is agreement regarding the temporal threshold (dt) of 1 s, though the spatial thresholds (dr) varies between 10 and 50 km [10], 15 km [19] and 0.15° longitude and latitude [24], which in the investigated region was approximately 20 km. All these articles agree that over a certain range, changes in dr do not significantly change the grouping of the strokes.

The spatial and temporal limits chosen in this report to group strokes into flashes are 15 km and 1 s respectively. These are not only based on the autocorrelation plots, but also on the location accuracy of SAFIR found in previous research. The spatial limit is then approximately twice the location

accuracy of 6.1 km [20].

Strokes are grouped when they occur within these values for dt and dr from the (first) reference stroke. This reference stroke gives the location and time of the flash. If any of the strokes within the flash was classified as CG-stroke, the whole flash is considered to be CG. If not, the flash is classified as a CC-flash. This is in accordance to the flash conversion as done by Poelman et al. (2013) and Drüe et al. (2007) [19][10].

3.2 Matching of datasets

Whenever two flashes from both datasets are detected within 1 s and 15 km of each other, this is classified as a match. If there are multiple flashes of the other dataset that qualify, the one with the smallest dr is chosen for the match. Flashes cannot be matched more than once. Any flashes that remain are labeled as unmatched; category 'b' or 'c' in table 2. In this table system B is the reference system.

	System B: **yes**	System B: **no**
System A: **yes**	**a** (hits)	**b** (false alarms)
System A: **no**	**c** (misses)	**d** (correct negatives)

Table 2: Contingency table example.

3.3 Contingency table scores

The probability of detection (POD) is calculated for all combinations of datasets, in which all datasets take turn in being the ground truth reference. A POD of dataset A out of dataset B is calculated as follows (see table 2):

$$\text{POD(A out of B)} = \frac{\text{a}}{\text{a} + \text{c}} * 100\%$$

Another common contingency table score is the false alarm ratio (FAR) with the following equation:

$$\text{FAR(A out of B)} = \frac{\text{b}}{\text{a} + \text{b}} * 100\%$$

The POD and FAR values are a measure of overlap between system A and B, where perfect overlap corresponds with a POD of 100% and FAR of 0%. The critical success index (CSI) is a measure of overlap which varies between 0% (no overlap) and 100% (perfect overlap) and is calculated with:

$$\text{CSI(A out of B)} = \frac{\text{a}}{\text{a + b + c}} * 100\%$$

These equations are very informative when validating forecasts with the ground truth [25]. However, when comparing datasets with one another in which one dataset is used as reference, the FAR value can be misleading. FAR implicitly assumes that any unmatched detection in A is not real since the reference B does not detect anything there as well. High false alarm ratio values could also be caused by poor detection in dataset B. In this report, the POD and FAR values are to be interpreted as indicators of how similar lightning detections of two systems are, more so than how well they detect lightning.

4 Results

4.1 Dataset characteristics

The results in this section are aimed to provide insight in the characteristics of the datasets used in this report. These characteristics concern the area of detection, the number of CC and CG strokes and flashes in FLITS and KLDN, the stroke/flash ratio and temporal variation over the time period.

4.1.1 Coverage

The area where lightning discharges are detected varies between the systems due to the placement of the sensors. In order to examine this in the datasets used, the accumulated number of strokes and flashes of FLITS, KLDN and ATDnet during the whole time period between 01-01-2010 and 31-10-2014 are compared. In figure 7 the number of total lightning, CC and CG strokes are counted per area box of approximately 10 * 10 km^2.

The scales are similar for maps of the same stroke type. As ATDnet does not distinguish between lightning types and is expected to only detect CG lightning, the total set is plotted twice with the same values. Once with an axis range similar to the TL maps of FLITS and KLDN, and once with an

axis range of the CG maps. ATDnet has global coverage, thought only the NL21 subset has been archived at KNMI.

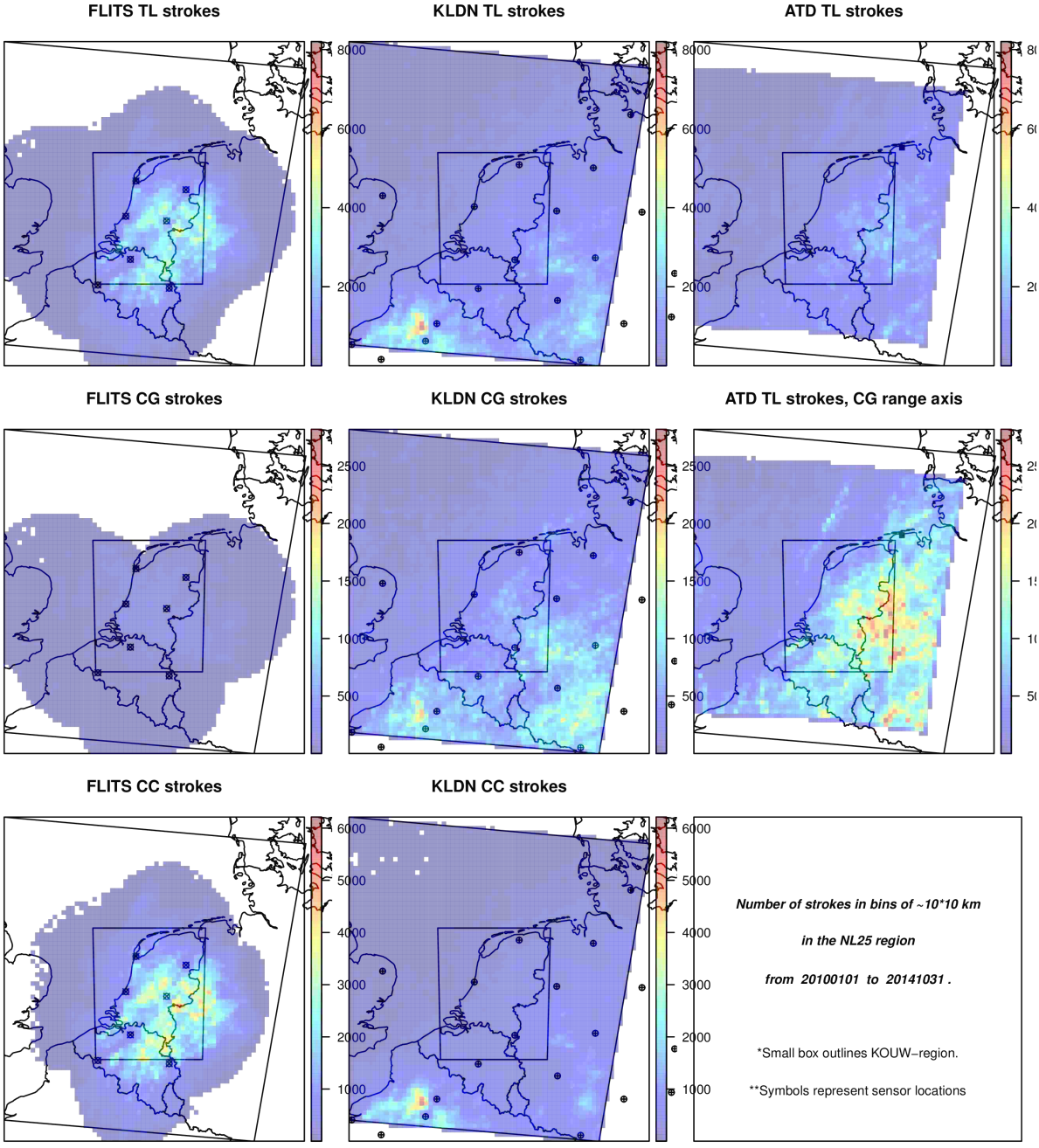


Figure 7: Accumulated number of strokes in area boxes of approximately 10 * 10 km^2 for total lightning (TL), CG and CC strokes of FLITS, KLDN and ATDnet, together with sensor locations of the system.

From figure 7 it is clear that the systems have a different coverage range. FLITS has the most limited coverage. It is interesting to see that all strokes FLITS detects north of the Netherlands over the whole time period are of type CC, suggesting that the most northern sensor's CG detection is lacking or misclassified as CC lightning. FLITS also detects far more CC strokes than CG strokes, and more CC strokes than KLDN. The sensor type in the Météorage network is not consistent. The CC detection of KLDN is highest above France, which is likely due to the improved CC detection after Météorage updated the sensors to LS7002. If all strokes detected by ATDnet are considered CG lightning, this system detects most CG strokes. All systems cover the KOUW-region, which is therefore the most suitable region over which comparisons between systems can be made.

The same figure as figure 7 is made for flashes, where strokes are grouped based on a maximum spatial distance of 15 km and a temporal distance of 1 s. The accumulated TL, CG and CC flashes of FLITS, KLDN and ATDnet are represented in the bottom 8 maps in figure 8. The map of ATDnet is again represented twice with different axis. Additionally, in the top three maps, the number of 10 minute intervals in which there is nonzero detection is given for each system. If during a 10 minute period, at least 1 flash is detected in the 10 * 10 km^2 box, this interval is a nonzero detection. The accumulated number of such intervals is a measure of the capacity of the system to detect if there is any lightning activity. This is particularly important near airports where the exact number of detections are less important than the omission of lightning altogether.

In the top three maps in figure 8, it becomes clear that FLITS most often detects at least 1 lightning flash. These are likely often CC flashes, of which FLITS detects significantly more than KLDN. It is important to note that a high detection rate can both be caused by a higher sensitivity as well as a larger number of false alarms. The nonzero detection values over the whole period of KLDN and ATDnet are quite similar in the KOUW-region. Similarly as for strokes, when all detections of ATDnet are considered CG, this system detects most CG flashes. FLITS detects the smallest amount of CG flashes.

4.1.2 Flash count dependence on dr-criteria

The number of flashes resulting from the dataset over a period of 1 year in the KOUW-region is calculated for different values for dr in the flash conversion. Figure 9 therefore represents the measure in which the flash count in

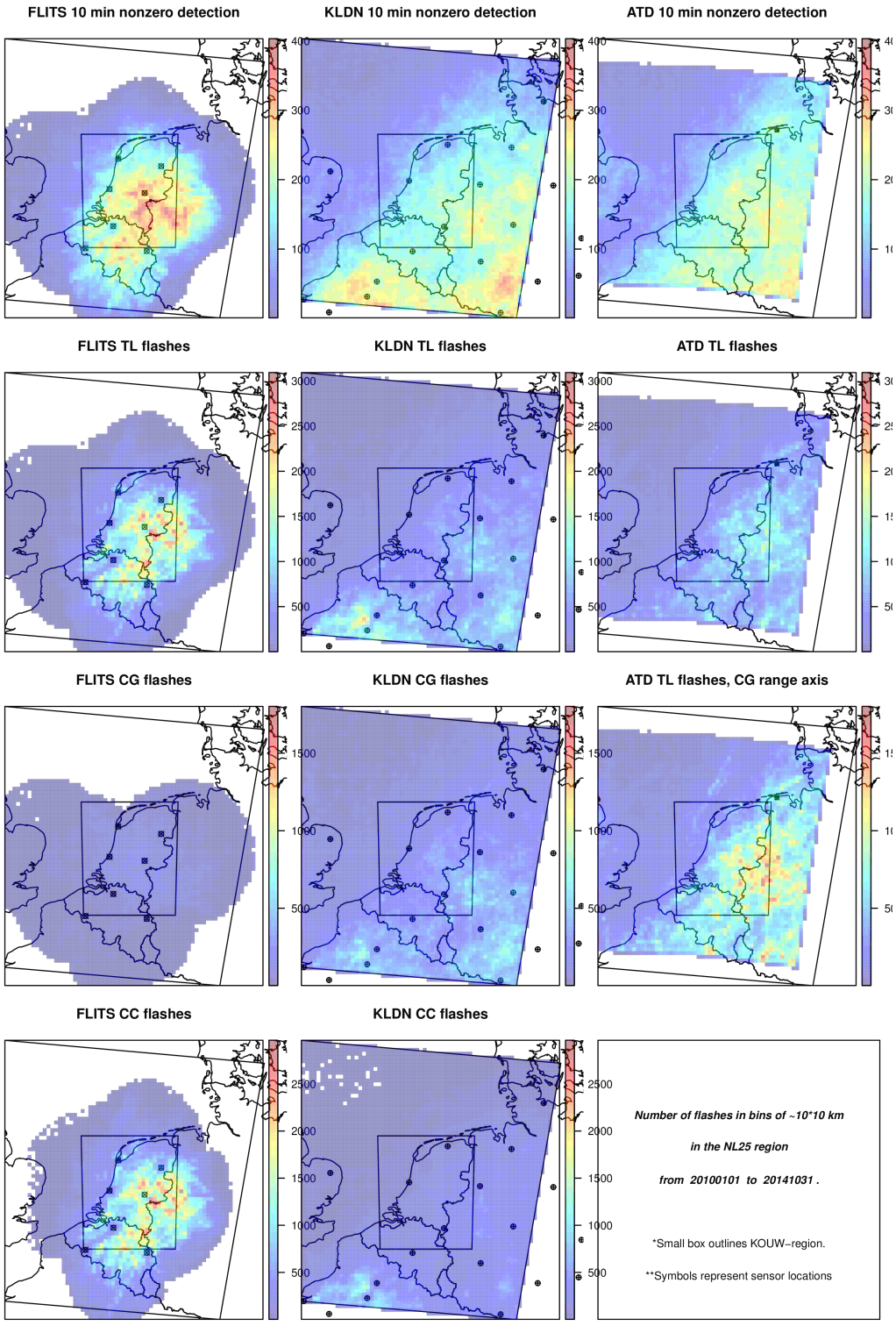


Figure 8: Accumulated number of flashes in area boxes of approximately 10 * 10 km^2 for total lightning (TL), CG and CC strokes of FLITS, KLDN and ATDnet, together with sensor locations of the system, where flashes are grouped strokes with dt = 1 s and dr = 15 km. The top maps represent the number of nonzero detection 10 minute intervals of each system; the number of 10 min periods in which there was at least 1 lightning flash registered by the system.

dependent on dr-criteria. The dt is constant at 1 s.

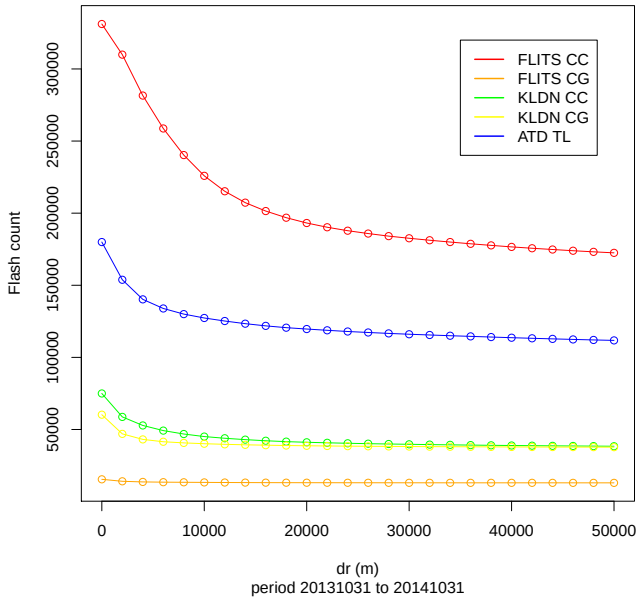


Figure 9: Variation in number of flashes for different dr values for FLITS, KLDN and ATDnet strokes in the KOUW-region from 31-10-2013 to 31-10-2014.

The flash count in figure 9 seems to show no more significant decrease in number of flashes at dr >25 km. A dr of 15 km would also be a safe value as a grouping criterion for strokes to flashes. The graph that levels off at the largest value of dr is FLITS CC. This is an indication of bad location accuracy (LA), i.e. the distance error margin of detection. CC segment discharges that are grouped together originate from different locations in the streamer, which spread further in the horizontal direction than the mostly vertically oriented CG flashes. The projection of the detection on the earth's surface by the detection system is therefore less accurate for CC lightning. This corresponds with a larger LA and a slower decrease of flash count with increasing dr.

Figure 9 shows that there are approximately 10 times more CC flashes in FLITS than CG flashes. The CC/CG ratio is close to 1 for KLDN in this period.

4.1.3 Temporal variation in the datasets

The period considered is 01-01-2010 to 31-10-2014. Temporal variations of several aspects in the datasets have been evaluated. See appendix A for more information.

- Most flashes occur between 16 and 19 UTC. A detection decrease at night is found for ATDnet, most likely due to the diurnal influence on VLF propagation. (Figure 29)
- FLITS and KLDN have an average stroke to flash ratio of approximately 1.5, and ATDnet has a ratio of approximately 1.3. These ratios fluctuate greatly over the time period though no clear trend is found. (Figure 30)
- The majority of lightning discharges occur between May and September. The relative detection between the systems does not seem to vary significantly over the time period. (Figure 31)

4.2 POD calculations

In this section the relative performance of the datasets is examined. The POD values are calculated over the whole KOUW-region for different dr values in order to verify if the choice of dr = 15 km is justifiable. Next the regional variation of POD is examined by calculating the POD in sub-areas of the KOUW-region, as well as temporal variations by calculating the monthly POD in time. Next the POD is calculated with a reference of the overlapping detections of the other two systems.

4.2.1 POD dependence of dr-criteria

The POD values between FLITS, KLDN and ATDnet are calculated with different values of dr for stroke to flash conversion and dataset matching.

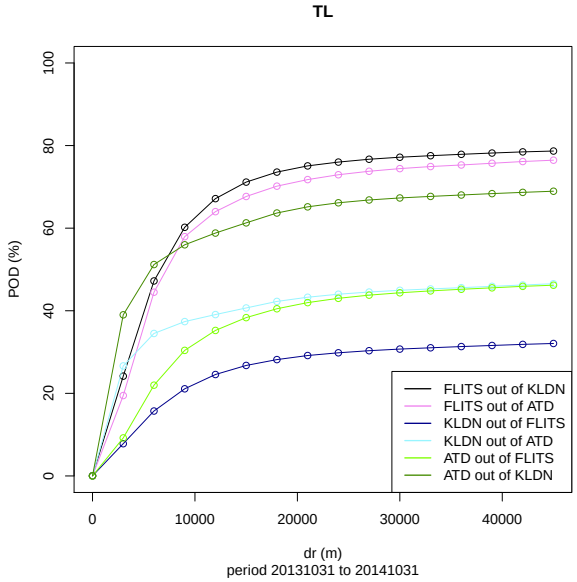


Figure 10: POD for various values of dr in the flash conversion of TL strokes and matching of datasets in the KOUW-region with dt = 1 s in the period 31-10-2013 to 31-10-2014.

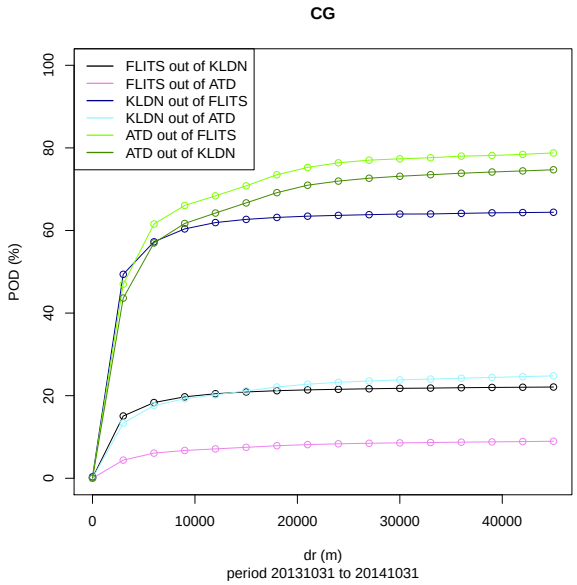


Figure 11: POD for various values of dr in the flash conversion of CG strokes and matching of datasets in the KOUW-region with dt = 1 s in the period 31-10-2013 to 31-10-2014.

Figures 10 and 11 show that the dr is justifiable as for dr >15 km no significant increases in POD occur. The POD of CG flashes seems to level off at lower dr values than the TL flashes. This is in agreement with the expected better location accuracy in CG detection. The POD outcomes of TL are significantly different from the CG POD values, as will be discussed in further detail in the following section.

4.2.2 Regional variations in POD

Figure 12: KOUW-region borders shown on map of the Netherlands [22].

In the so-called KOUW-blocks, 12 fixed rectangular regions (80 * 90 km) that together span the area of the Netherlands (see figure 12), the POD is calculated between FLITS, KLDN and ATDnet over the period between 01-01-2010 and 31-10-2014. The same is done when only CG strokes in FLITS and KLDN are considered, with TL of ATDnet. The POD values are calculated with flash conversion and matching done with the criteria dt = 1 s and dr = 15 km. If a flash in the KOUW-region matches with a flash outside this region, the flash is still labeled as a match. No faulty unmatched flashes result from the border locations.

Again big differences between CG and TL POD values are found, as well as a large variability between KOUW-blocks. The POD over the whole KOUW-region as well as the range of the 12 POD values in the KOUW-blocks are given in table 3.

	TL POD	CG POD
FLITS out of KLDN	67.5 %: 39 to 85 %	20.0 %: 12 to 40 %
FLITS out of ATD	52.0 %: 32 to 63 %	6.2 %: 4 to 12 %
KLDN out of FLITS	25.4 %: 22 to 34 %	71.5 %: 62 to 84 %
KLDN out of ATD	31.3 %: 28 to 40 %	20.4 %: 17 to 26 %
ATD out of FLITS	38.8 %: 33 to 44 %	72.5 %: 55 to 85 %
ATD out of KLDN	62.1 %: 48 to 72 %	66.6 %: 51 to 78 %

Table 3: POD values in KOUW-region.

The first conclusion that can be drawn from table 3 is that the POD-range between KOUW-blocks is quite large. Evidently the POD is very variable per region.

These values can be compared with the flash findings of Poelman (2013) et al.'s comparison between Météorage, SAFIR (LF only) and ATDnet in Belgium as represented in table 1 in section 2.2.5. The color signifies whether the POD value has been found higher, lower or similar to the value found in table 1. From this we find that FLITS performs less well in this comparison than SAFIR in Belgium during the 2011 and 2012 lightning season; The CG POD from FLITS out of the other two system is lower, while the other systems detect a larger fraction of FLITS CG flashes. KLDN detects a larger amount of FLITS flashes in our study (though note that the VHF data of the system is used here). However, the overlap between KLDN and ATDnet is lower, as the POD values in both combinations are lower.

Without taking the previous study in Belgium into account, we can see that the highest relative performance is found by comparing CG detection of KLDN and ATDnet with FLITS. The reverse CG POD's are the lowest, meaning that the FLITS CG detection could be lacking. However, the TL POD values of FLITS compared to KLDN and ATDnet are higher, while the reverse values are smaller than the CG case, suggesting that FLITS detects a lot of CC lightning which is either not detected or classified as CG lightning by both other systems.

ATDnet seems to detect a significant number of flashes not detected by KLDN. This could be because Météorage sets the noise thresholds for labeling a signal as lightning higher than the thresholds used in ATDnet.

4.2.3 Temporal variations in POD

The POD value has been proven to change significantly when considering a different area or type of lightning. In order to get an idea on the variability of the POD in time, the POD has been calculated in the KOUW-region in time bin sizes of 30 days during the 5 lightning seasons of 2010 to 2014.

The graphs are given in figures 32 and 33 in appendix B. They show a large variability with shifts of up to 30% of the POD values in time, especially where ATD is involved. There does not seem to be a clear annually recurring trend in the POD value. For the TL graphs, the POD values of FLITS out of KLDN, FLITS out of ATD and ATD out of KLDN are often the highest. For CG POD values the highest values belong to ATD out of KLDN, ATD out of FLITS and KLDN out of FLITS.

4.2.4 POD and FAR with combined reference

The POD and FAR are again calculated for each KOUW-block, where the dataset is compared with a dataset constructed from the matched values of the two other datasets to create a more robust reference. The mean values of time and location are used to construct matches. As before, dt = 1 s and dr = 15 km in flash conversion and matching. This is also done when only CG strokes of FLITS and KLDN are taken into account together with TL of ATDnet. The calculations are made over de period 01-01-2010 to 13-10-2014. As before, no faulty unmatched flashes are due to the border locations.

Remember: FAR(A out of B) = number of flashes found in A but not in B / total amount detected by A.

There is a large variability between the KOUW blocks and between CG and TL calculations, see figure 13. In these figures, lines of equal critical success index (CSI) is given as a reference. CSI is 100% for systems that overlap completely, and 0% for systems that do not overlap at all.

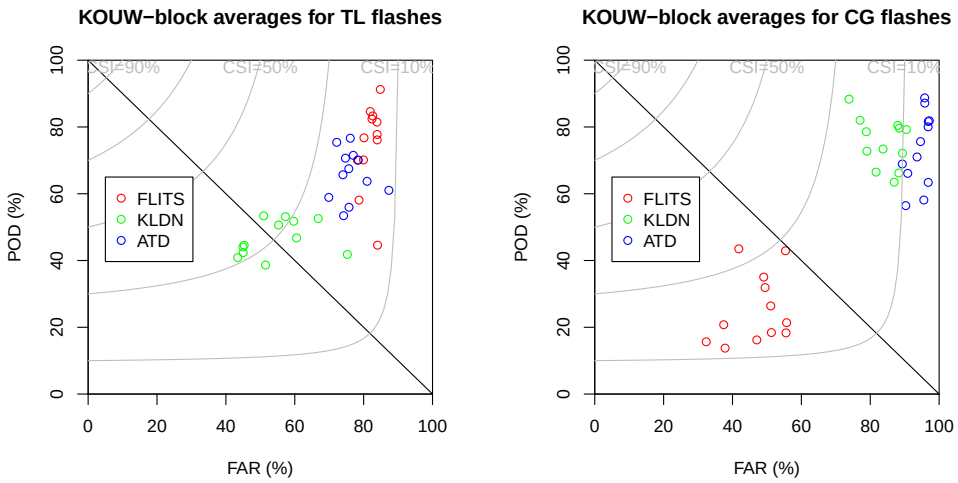


Figure 13: Plot of POD and FAR values of each of the twelve blocks in the KOUW-region calculated in the period from 01-01-2010 to 31-10-2014, with the diagonal line representing bias = 1, and values calculated for total lightning and for CG strokes only. Curved lines of equal CSI show critical success index.

Figure 13 shows the variation in POD and FAR values within the blocks of the KOUW-region, as well as the relative performance with the other datasets. For perfectly matching detection performances the figure would show the values in the upper left corner. We see in figure 13 that high POD values often are related to high FAR values. The values right above the diagonal where the bias = 1 shows that a large fraction of the detection match with the reference, though in addition a large number of extra detections are made.

FLITS detects a larger number of CC flashes than the others while the detection of CG is lower than that of ATDnet and KLDN. This means that the reference becomes quite small for KLDN and ATDnet in the CG calculations, resulting in a large FAR. FLITS proves itself the most limited system in CG detection with values left below the bias = 1 diagonal. It misses approximately 70% of the reference and finds relatively few flashes outside this reference.

KLDN detects approximately 50% of the TL flashes in the reference of KLDN and ATDnet. Similarly, half of its detections are not part of the TL reference. Both POD and FAR increase in the CG calculation because of

small CG detection of FLITS limiting the reference.

The large increase of POD of FLITS for TL compared to CG shows that CC flashes are responsible for a large part of the overlap with the other system's matches, though this also results in a higher FAR, i.e. it detects flashes not seen by both other systems.

From the higher CG POD and FAR values in KLDN and ATDnet compared with FLITS it becomes clear that a significant number of CG flashes are not found as CG flashes by FLITS.

4.3 Severe weather detection

On June 9th 2014, KNMI gave out a red weather alarm for the province of Limburg because of extreme lightning. Because of the risk of lightning strikes, it was advised to stay clear of open fields and away from high trees. This day coincided with Pinkpop, an outdoor festival with 67,000 visitors in Landgraaf, Limburg. The storm went in North-East direction, passing exactly over Landgraaf between 18 and 19 UTC, when the festival was still ongoing. During the most severe part of the storm the festival was put on hold and the public was advised to stay on the field with emergency operators stand-by. Despite KNMI's warning, the Pinkpop organization implored the public not to panic claiming that the lightning storm would not hit the location before the end of the festival. Though the storm passed the festival terrain without any casualties, the organization of the festival were reproached for the way they handled the red alert.

KNMI plays an important part in warning the public for severe weather. The weather warning in the case of Pinkpop festival was deemed justified in hindsight, as the number of discharges were extensive, putting the public at considerable risk. The use of a different lightning system should not affect the situations in which KNMI gives out a lightning related weather warning. Figure 14 shows the CG and CC flashes as detected by FLITS and KLDN on June 9th 2014. The first thing that stands out is the differences between the number of detections. During this event, FLITS detected far more CC flashes and far less CG flashes than KLDN. Also the number of total lightning flashes detected by FLITS was higher than KLDN. With the historical datasets of KLDN and FLITS, we can investigate all moments in the period between 2010-01-01 and 2014-10-31 where FLITS would have detected severe weather, and we can compare that to the detections by KLDN.

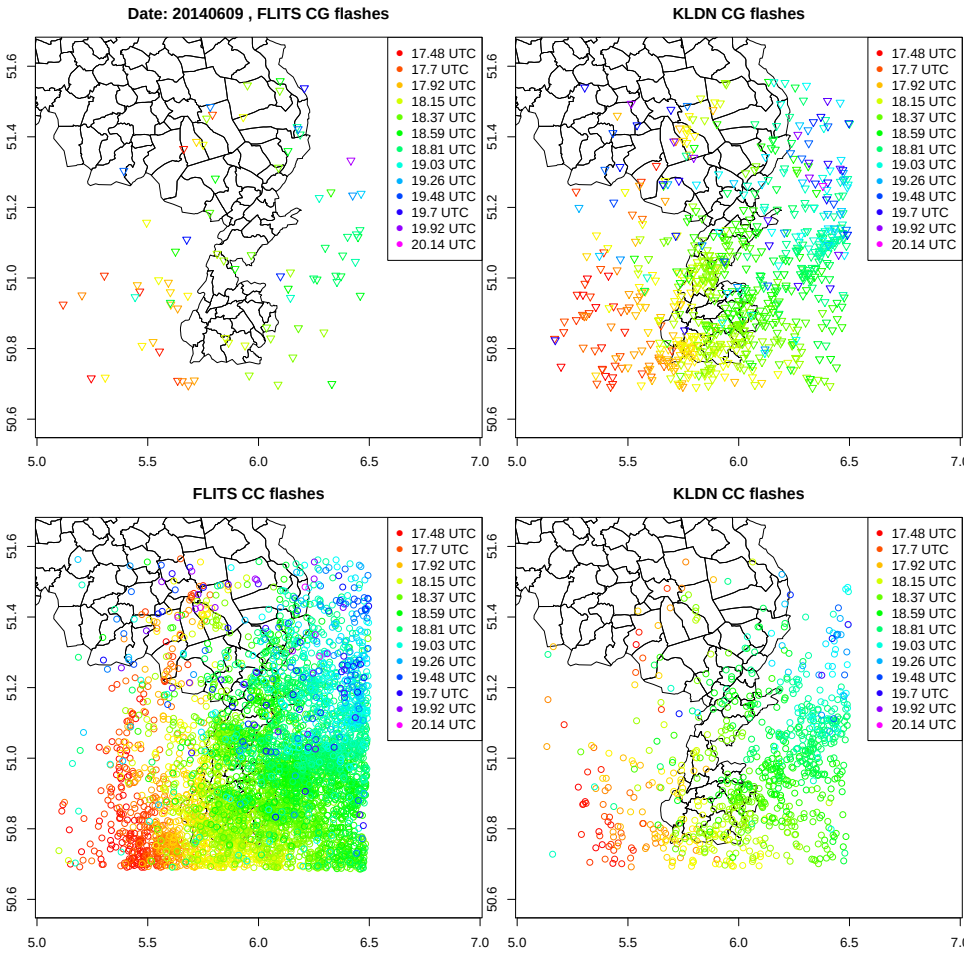


Figure 14: Detection of CC and CG flashes by FLITS and KLDN during Pinkpop 2014.

The so-called KOUW-system (Kansverwachtingen voor Onweer ten behoeve van Uitgifte Weeralarm), that predicts the probability of severe thunderstorms, has made use of FLITS stroke output (i.e. CG strokes and grouped CC segments) [22]. The criterion for a weather alarm for severe thunderstorms is defined as at least 500 TL strokes within 5 minutes in an area of 50 * 50 km^2.

When analyzing the historical datasets of FLITS, KLDN and ATDnet from 01-01-2010 to 31-10-2014, the number of times the criteria for a weather alarm are reached within the KOUW-region can be counted. In the evalua-

tion of the stroke datasets this occurs on 14 days for FLITS, 2 days of KLDN and on 7 days for ATDnet. We consider the moments that FLITS has detected more than 500 strokes within an area of 50 * 50 km^2 within 5 minutes as reference. We define severe weather peaks by grouping 5 minute intervals that meet the criteria when they are part of the same storm; If there is not more than 600 seconds and 50 km distance between these intervals where the criteria are not met they are considered part of the same severe weather peak. The resulting dataset consists of 25 severe weather peaks, occurring on 14 unique days of the historical dataset.

Table 4 shows the number of KLDN severe weather peaks that coincide with the reference of FLITS severe weather peaks ('Hit') as well as the times KLDN detects a severe weather peak where FLITS detects none ('False alarms'). A hit is defined if the severe weather peak of KLDN overlaps (no matter how briefly) in time and area with a severe weather peak of the FLITS reference. The minimum stroke count criterion for severe weather peaks is lowered for KLDN while the reference remains the same. The criteria of 50 * 50 km^2 area and the period of 5 minutes to define an interval of severe weather remain constant, as well as the grouping of time intervals into severe weather peaks.

KNMI issues a weather warning if there is a high probability of severe weather. The most right column in table 4 states the number of days that KLDN detects severe weather that coincide with a day where the KNMI has given a thunder storm related weather alarm. The period considered here is 01-01-2010 to 31-10-2014.

Minimum stroke criterion	Hits	False alarms	KLDN alarm days	..coinciding with KNMI alarm days
500 strokes	1 out of 25	1	2	2
400 strokes	1 out of 25	1	2	2
300 strokes	2 out of 25	6	5	3
200 strokes	4 out of 25	20	15	6

Table 4: Severe weather detection.

From table 4 it becomes clear that a reduction of the minimum stroke criterion in KLDN is not the way to detect the same severe weather peaks as FLITS with the original criterion. Not only is KLDN not even close to

detecting the same storms as FLITS (4 out of 25 for a minimum stroke criterion of 200 strokes), but the number of false alarms increases dramatically. Similarly, the number of days where weather alarms are registered that do not coincide with a KNMI weather alarm increases too.

When the historical datasets from 01-01-2010 to 31-10-2014 are examined, the number of 5 minute intervals where the stroke detection in a 50 * 50 km^2 area exceeds 500 are counted. Often severe weather lasts longer than 5 minutes, and in that case the intervals follow one another. FLITS detects 155 such intervals, KLDN 4 and ATDnet 45. In figure 15 the relative detection of each system is represented in boxplots. In the 50 * 50 km^2 area in the 5 minute interval in which there are over 500 stroke detections, the number of detections of the other systems are counted and divided by the exact number of detections of the system that gave the alert. This is done for number of strokes (left) and flashes (right) where flashes are defined by a maximum distance between strokes in time and space of dt = 1 s and dr = 15 km.

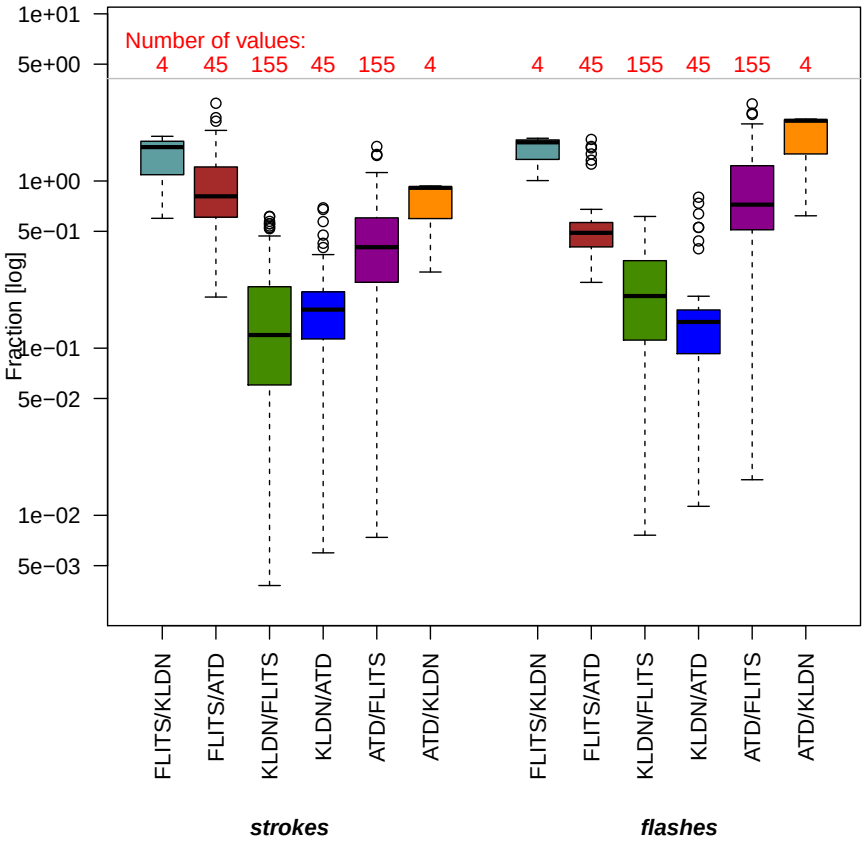


Figure 15: The stroke fractions (left) and flash fractions (right) between non-alerting systems and alerting system for all intervals the alerting system detects more than 500 strokes in a region of 50 * 50 km^2 within 5 minutes. The fraction values are represented in a boxplot on a logarithmic vertical axis. On top the number of values in the boxplot, namely the number of intervals where the weather alarm criteria are met for that particular system in the period between 01-01-2010 and 31-10-2014, are given.

If the difference in detection would be mainly due to differences in relative sensitivity, there would be a more or less constant factor difference in detection within the same region and time. In that case the boxplots would show little variation, as there would always be a similar factor difference between the systems' number of detections. From figure 15 this does not seem to be the case for strokes as well as for flashes. The number of stroke detections of KLDN during weather alarm intervals of FLITS vary between 0.5 and 0.005 of the stroke detections in FLITS (green boxplot).

In some cases the boxplot span the area above and below one. For instance the number of FLITS stroke detections when ATDnet detects more than 500 strokes in 50 * 50 km^2 and 5 minutes is sometimes more and sometimes less than the exact number of stroke detections by ATDnet in this interval. This suggests that the difference in detection between systems is more complex than merely a factor difference due to sensitivity towards total lightning.

We can examine this further by comparing KLDN detections with FLITS during the severe weather intervals. If we consider all 155 of the 5 min intervals between 01-01-2010 and 31-10-14 in which FLITS detects more than 500 strokes within 50 * 50 km^2 in the KOUW-region as a reference, we can count the number of strokes (TL and CG only) detected by KLDN in the same period and 50 * 50 km^2 area. These values are plotted in figure 16.

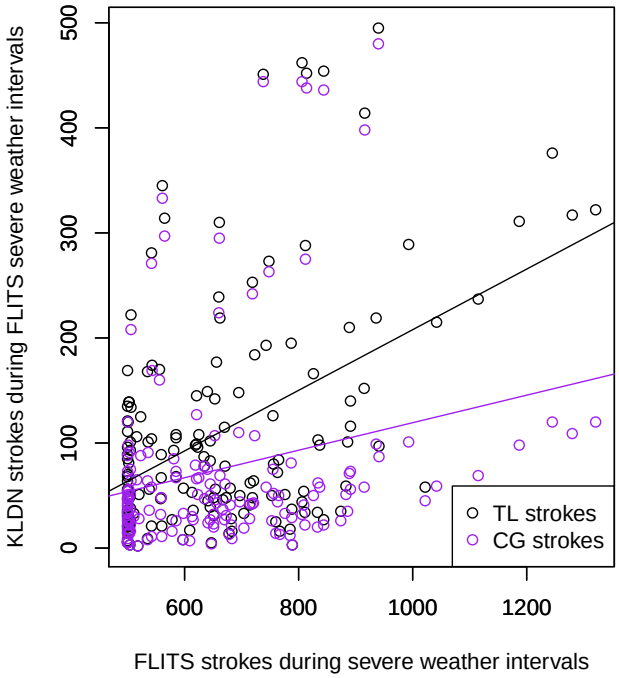


Figure 16: Number of KLDN and FLITS strokes during the 155 FLITS severe weather intervals with linear fits.

In order to detect the same severe weather intervals as FLITS, the threshold for KLDN should be lowered to only 2 detections per 5 minutes in a 50 * 50 km^2 area. This would result in enormous amounts of false alarms, as was hinted in table 4 in the beginning of this section.

Upon closer examination of the severe weather intervals of FLITS, a large amount of the strokes detected in these intervals are CC strokes. The number of CG strokes per interval in which there are at least 500 TL strokes detected in a 50 * 50 km^2 area and 5 minutes by FLITS are shown in the histogram in figure 17.

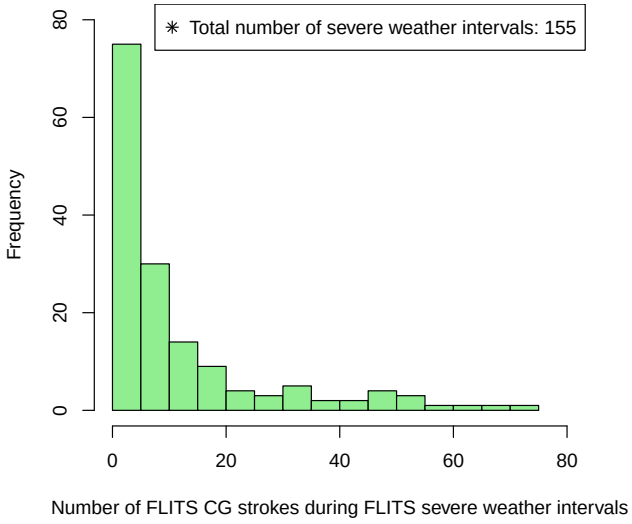


Figure 17: Histogram of the number of CG stroke detections of FLITS during which the TL detection was at least 500 strokes per 50 * 50 km^2 area in 5 minutes.

As KLDN detects significantly less CC strokes compared to FLITS and the weather alarm intervals consist mainly of CC strokes, entire severe weather intervals are missed by KLDN.

The danger posed by severe thunderstorms is due to more than just the lightning discharges themselves. Heavy lightning coincides usually with heavy rainfall and strong wind gusts, both of which can cause significant damage as well. As previous results pointed to the conclusion that the same severe weather peaks in the historical dataset cannot be found by KLDN

stroke detection alone, the weather alarm criteria could be extended to a more general weather warning. Not only number of lightning discharges, but other meteorological data are considered as well.

The hourly meteorological variables of precipitation amount (RH) and maximum wind gusts (FX) at the nearest KNMI automatic weather station are examined for the FLITS severe weather peaks in the historical dataset. Though in some cases the RH and FX values at the nearest KNMI station are representative of severe weather, in almost all of the hours where measured FX and RH are high a severe weather peak of FLITS occurs nearby.

We compare severe weather peaks of (combined) 5 minute intervals where KLDN detects at least 70 CG strokes within an area of 50 * 50 km^2 with the hourly meteorological data at the nearest meteorological station. Only CG strokes are considered because they pose more risk to national safety. All RH >6 mm and FX >1.6 m/s coincide with the severe weather peaks. Nevertheless, many severe weather peaks occur when the other meteorological variables do not report severe weather. This means that the severe thunderstorm probably did not pass the station. Also, of the 15 unique days where KLDN detects at least one severe weather peak, only 9 coincide with the 14 alarm days of the FLITS reference. However, a broader weather alarm could be made based on the criteria that there should be a minimum of strokes detected, as well as heavy rainfall and strong wind gusts. By deriving these values from radar it becomes less likely that a storm is missed because of a too large distance to the nearest stations. See figures 34 and 35 in appendix C for detailed analyses.

4.4 Lightning Activity Detection (LAD)

The users of lightning detection systems are interested in the detection of all lightning flashes. Though more importantly, it is vital that lightning storms are not missed by the system. If a lightning detection system only detects a fraction of the flashes in a certain area and time, the user still has information on the occurrence of the lightning storm. However, if all flashes of a lightning event are undetected, the user does not know lightning has occurred in that period and region. In general, a lightning detection system is more useful if it misses a fraction of the flashes, though detects all lightning storms, compared to a system that misses entire storms. Especially when lightning detections are used for severe weather warnings, the detection of the first discharges in a storm are particularly important.

In this study the POD is calculated for by matching each flash to a flash detected by a different system. In the so-called Lightning Activity Detection (LAD), we compare the datasets to one another in their ability to detect any lightning activity over a certain time span in a predefined area. A match is defined if both systems detect at least one flash in a fixed 10 * 10 km^2 area over the time interval. Note that it does not matter if one system detects many more flashes than the other. As long as at least one lightning flash is found by both, the datasets match in this period in this area. If one system detects lightning while the other doesn't, this is labeled as an unmatched detection. This method is derived from Poelman et al. (2013) [19].

$$\text{LAD(A out of B)} = \frac{\text{matched area boxes between A and B}}{\text{total detected area boxes of B}} * 100\%$$

Again, the false alarm ratio can be determined. This can be calculated here with the following equation:

$$\text{FAR(A out of B)} = \frac{\text{unmatched detection area boxes of A}}{\text{total detected area boxes of A}} * 100\%$$

The following LAD values are calculated (area box size is 10 * 10 km^2 in all):

- TL flashes with dt = 5 min
- CG flashes with dt = 5 min
- CG strokes with dt = 5 min
- TL flashes with dt = 1 day
- CG flashes with dt = 1 day
- CG strokes with dt = 1 day

The LAD values are again calculated for any combination of systems where one system is considered the reference. The LAD and FAR value for each calculation is represented in the top panel of figure 18.

Because of the damage risks to aircrafts, airport policy forbids refueling, takeoff and landing during lightning. The alert level for at least one discharge within 10 km of the airport is amber, and red for 5 km [2]. The air can be declared safe when there have been no discharges in these regions for 10 minutes. To compare the performance of the lightning detection systems,

the LAD is calculated in the region with radius 5 km and 10 km around Schiphol airport near Amsterdam, and Rotterdam airport. The considered time period bins are 10 minutes. The resulting LAD and FAR values are represented in the bottom plot in figure 18. In these figures, lines of equal critical success index (CSI) is given as a reference.

Several things become evident from the top figure 18.

- LAD values are higher and FAR values are lower when a longer period is considered. See figures 36 and 37 in appendix D for the dt dependency of LAD and FAR.
- LAD and FAR are slightly better, i.e. higher CSI, for stroke detection compared to flash detection.
- All detection system combinations (for strokes as well as flashes) have similar TL LAD en FAR values, though these values differ greatly for CG.
- Under the optimal conditions, where CSI is highest (TL, 1 day), the systems agree in approximately 70% of the cases on the presence of lightning.

The positions of the colors in the bottom graph in figure 18 are most in agreement with the 5 min TL flash case in the top figure, as is expected. From the positions of the pink and black colors in the upper right and the positions of the dark blue and light green in the bottom left, we can conclude that FLITS detects most often non zero lightning near airports. FLITS strokes are not detected by the other systems in approximately 60% of the cases, since the LAD of light green and dark blue are approximately 40%; KLDN and ATDnet would give out an amber flight alert (LAD of capital letters) in approximately 50% of the times FLITS would, and 35% in case of the red flight alerts (LAD of small letters). The FAR of approximately 20% of KLDN and ATDnet out of FLITS means that in 20% of the KLDN and ATDnet alerts, FLITS does not detect any lightning.

When considering only the months May-September (i.e. the lightning season), the outcomes are slightly better than when the whole year is considered: the LAD is slightly higher and the FAR is slightly lower in all cases (results not shown in graph).

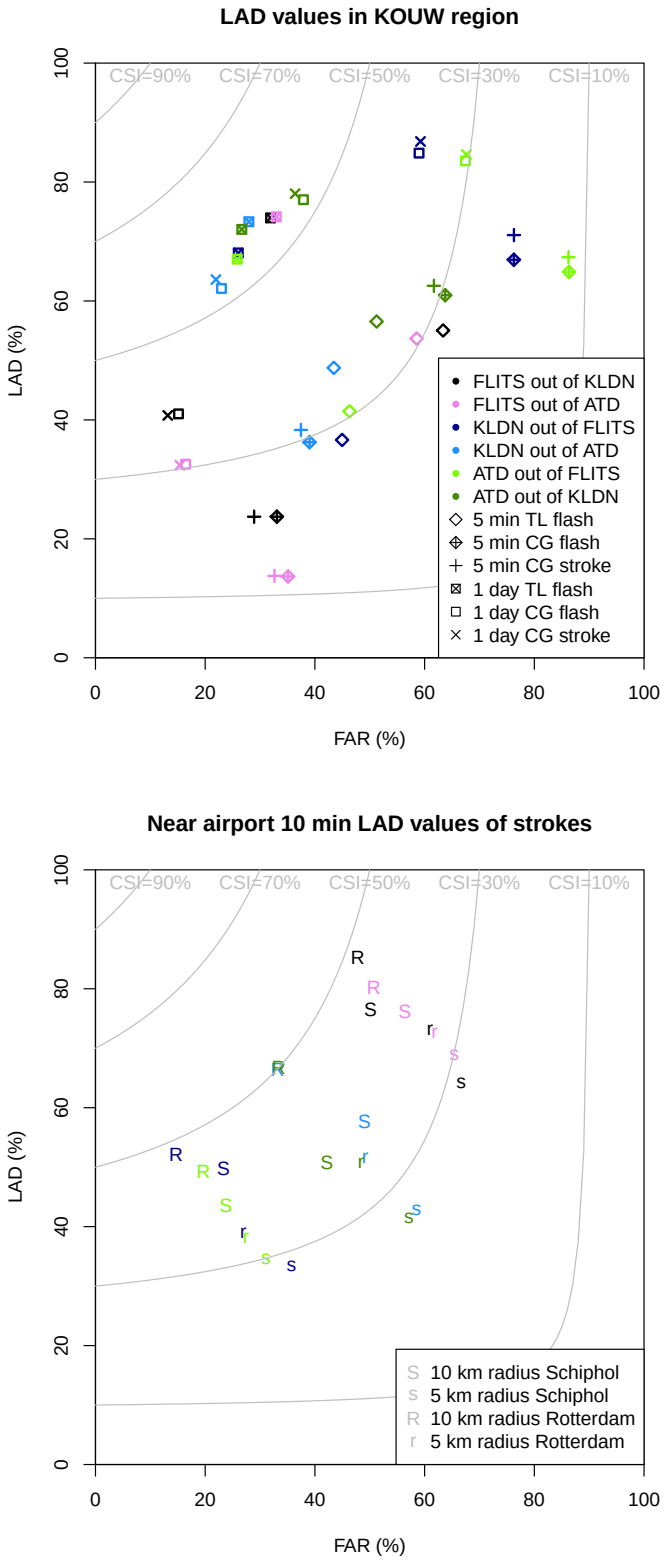


Figure 18: LAD and FAR values in KOUW-region in 10 * 10 km^2 boxes (top) over the period between 01-01-2010 and 31-10-2014, as well as LAD and FAR values of stroke detections within 5 and 10 km of the airports Schiphol and Rotterdam (bottom). Curved lines of equal CSI show critical success index.

A similar comparison between SAFIR (LF only), Météorage and ATDnet in Belgium over the lightning seasons of 2012 and 2013 by Poelman et al. (2013) yielded LAD values between 83% and 95% and FAR values between 5% and 17% with same area box sizes and time bin size of 1 day [19]. Apparently the system detections are more in agreement over Belgium than over the Netherlands. This can also be due to the fact that we use a different SAFIR dataset (with VLF included) than Poelman et al.

5 Discussion and conclusions

In order to gain insight in the anticipated changes of KNMI lightning detection by the use of KLDN instead of FLITS, 5-year historical datasets of FLITS, KLDN and ATDnet in the Netherlands have been analyzed. The relative performances of FLITS, KLDN and ATDnet have been examined with special focus on severe weather detection and detection of lightning activity.

Though the relative performance has been shown to fluctuate in time and location, we can draw conclusions from the average POD values in the KOUW-region. As the POD of FLITS out of KLDN and out of ATDnet decreases significantly when only considering CG lightning, we can conclude that FLITS detects only a small amount of the CG flashes detected by other systems, although it compensates this by higher CC detections. This is in accordance with FLITS's CC:CG detection fraction of approximately 10:1. It is likely that FLITS misclassifies CG detections as CC.

From the relative detection of KLDN and ATDnet, we see that the POD of ATD out of KLDN does not differ much between TL and CG. However, the fraction of ATDnet flashes that KLDN detects is larger in TL compared to CG. This means that a considerable amount of KLDN CC detections overlap with ATDnet detection. As ATDnet is expected to mainly detect CG discharges, this result indicates that either ATDnet detects more CC flashes than assumed, or that KLDN misclassifies CG detections as CC.

The increase of POD and FAR of FLITS TL vs CG calculated with a reference of matches between KLDN and ATDnet reaffirms the importance of CC flashes in the overlap with the other systems. It could be interesting to calculate the FAR with a reference where at least one of the two other systems detect a lightning flash as the reference of flash matches results in very high FAR values that give limited information.

Even though misclassification may occur in both systems, the number of flashes in FLITS far exceeds KLDN. This diminishes the number of times KLDN would reach the weather alarm criterion of 500 TL strokes within 50 * 50 km^2 within 5 minutes. Upon examination however, the situations in which FLITS reaches the weather alert criterion are not found by KLDN by just diminishing the number of discharges in the weather alarm criterion. The number of strokes detected by KLDN during FLITS severe weather peaks vary between several hundreds to values below 10, therefore making it impossible to distinguish these severe weather peaks from non-severe weather.

There has been discussion about the usefulness of the weather alarm criterion, as it considers TL strokes. The number of strokes in a flash are not directly relevant for safety considerations. Though TL has been suggested to be indicative of the severity of thunderstorms, CG lightning poses more risk for the general public than CC lightning. It could therefore be argued that the weather warning criterion should address CG flashes instead of TL strokes. Though KLDN is better in detecting CG, the severe weather peaks of FLITS consists mostly out of CC detections. The CG flash count weather alarm criterion should therefore be extended with criteria on meteorological conditions like rainfall and heavy wind gusts.

The lightning activity detection calculations show that all systems agree on the occurrence of lightning in approximately 70% of the cases for TL flashes over the span of a day, but disagree significantly when CG lightning and shorter timespans are considered. The LAD and FAR values found in the Netherlands are worse compared to the results with similar systems and the same calculations in Belgium [19].

The amber and red alerts that communicate negative refueling, takeoff and landing advise are based on the occurrence of lightning in a radius of respectively 10 and 5 km around an airport. The LAD values around Rotterdam airport and Schiphol show that KLDN would agree with approximately 50% of the amber alerts and 35% of the red alerts given out by FLITS. Similarly, FLITS agrees with approximately 80% of the amber alerts and 65% of the red alerts given out by KLDN. This shows that the change from FLITS to KLDN results in significantly different flight alarms.

In summary, the lightning detection systems of KLDN and FLITS differ significantly. The coverage of KLDN extends much farther than FLITS, and FLITS detects no CG discharges in the north of the Netherlands (because

of a faulty sensor). Though both KLDN and (especially) FLITS are likely classifying CG discharges as CC, KLDN detection performance regarding CG flashes is probably better. The large number of CC discharges measured by FLITS results in the detection of a relatively large number of so-called severe weather peaks while KLDN falls short in this respect. Adjusting the weather alarm criterion for severe thunderstorms in a manner that the same severe weather peaks are found by KLDN has been proven unsuccessful. A weather alarm criterion for severe thunderstorms based on meteorological variables like wind gusts and precipitation as well as lightning discharges is suggested. Similarly, negative flight advice alerts for thunderstorms/lightning will be significantly different with KLDN compared to FLITS.

The conclusion that FLITS and KLDN detect different aspects of a lightning storm with a poor agreement of historical datasets has severe impacts on the applications that were made for the FLITS detection system. Météorage could do a ground truth campaign of high-speed camera observations in the Netherlands similarly as the one in Belgium [20], especially since the POD and LAD values found in the Netherlands are lower than those found in Belgium with similar systems [19]. This would give information on the possibility that KLDN CG lightning detection is an improvement compared to FLITS. An on location lightning validation campaign could give conclusive evidence on the ability of KLDN to detect CG lightning near airports. However, the fewer CC detections by KLDN (due to the LF sensors) could result in reduced capacity to warn airports for lightning activity, regardless of good CG detection.

Severe weather detection with KLDN will likely not be done with lightning stroke count alone. In a similar fashion, solutions for lightning monitoring and airport warnings could be found by an extending on KLDN lightning data with other meteorological variables. If these solutions are not sufficient, KNMI has to decide whether the current application on the lightning detection system should be maintained. If so, a prolongation of the FLITS system should be considered. If the maintenance of the applications warrants the additional costs of the upkeep of FLITS for CC detection KNMI could combine the data of FLITS and KLDN (and possibly even ATDnet) for CG detection and thereby monitor total lightning in the Netherlands.

6 Epilogue: The potential of infrasound for lightning detection

6.1 Introduction

As large reliable ground truth datasets of lightning discharges are difficult to obtain, validation of detection systems are often done by comparing several systems to one another [19][10]. A new approach is to use data from infrasound arrays that are installed for nuclear test-ban treaty verification purposes, thereby possibly providing a validation technique without additional costs. By comparing the measured infrasound data with three ground based electromagnetic lightning detection systems in a case study of a thunderstorm in the Netherlands, the potential of using infrasound for lightning detection can be examined.

6.1.1 Infrasound lightning detection

Besides electromagnetic signals lightning events emit acoustic signals as well. Transport of charges through the lightning channel heats it to temperatures of up to 30,000 K and the resulting expansion of air (over 10 atm) causes the thunderous sound [4]. Besides audible sound from channel expansion, (for humans) inaudible infrasound (<20 Hz) is produced in the cloud base [21]. Dessler (1973) describes a model of this process, first suggested by Wilson (1920) [9]. In a charged area of the cloud, charged water droplets repel each other. The electrostatic field reduces the atmospheric pressure in the charged cloud area. The repellence suddenly stops when a lightning stroke neutralizes the cloud region. The sudden compression as the atmospheric pressure restores is believed to produce infrasound pulses. This signal has a strong vertical orientation [11]. However, the exact mechanism causing infrasound in thunderstorms is often debated [8] [5]. Nevertheless, infrasound from lightning is easily discriminated from infrasound originating from other sources, with a dominant frequency of 3.9 Hz for CG lightning. [3].

Infrasound can be detected with a microbarometer. In an array of microbarometers the arrival time difference of the wave signal between elements is calculated. This determines the direction and apparent velocity of the signal. With the apparent horizontal sound velocity measured by the array the elevation angle of the signal can be determined, i.e. apparent speeds larger than the speed of sound indicate a source at higher altitudes (see figure 19). By combining angles of arrival (azimuth and elevation) and arrival times from at

least two arrays, the infrasound-sources can be localized. The effects of wind and temperature on sound velocity are ignored. A direct path of propagation is assumed, as the signal-to-noise ratio (SNR) is not sufficient for infrasound propagating inside the atmospheric waveguide (i.e. an indirect non-straight path) [12].

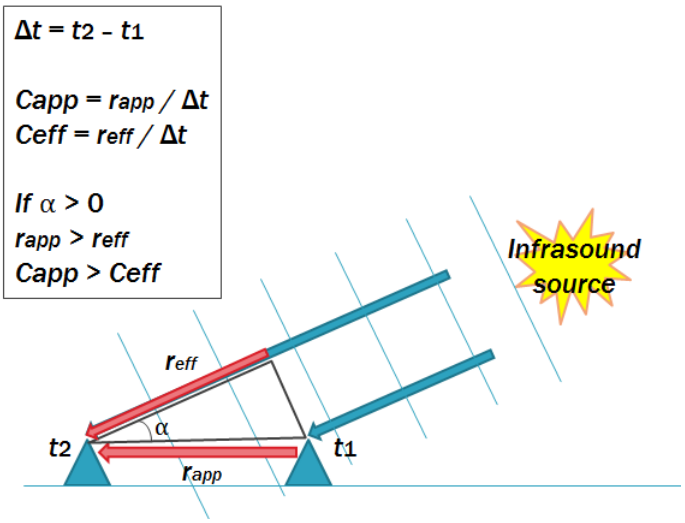


Figure 19: Schematic representation of infrasound detection with two array instruments of sources at higher altitudes, where r_{eff} is effective path, r_{app} is apparent path, t_1 and t_2 are arrival time at the instruments and $Ceff$ and $Capp$ are effective and apparent velocity respectively.

6.1.2 Validation

Assink et al. (2008) aimed to link combined VHF and LF electromagnetic lightning detections from SAFIR to infrasound data at two stations. Localization of lightning discharges from cross bearing the two infrasound datasets has been unsuccessful due to too large distances between stations (145 km apart). Good correlations between CG discharges as measured by SAFIR and infrasound data were found up to 50 km distance from the infrasound array. The deviation in theoretical and observed back-azimuth was found to be $4.6 \pm 3.2°$ [3].

Farges and Blanc (2010) found a good correlation between data from one infrasound array and Météorage's LF electromagnetic detections of CG discharges within 150 km. The low detection efficiency of 5% was attributed to the notion that subsequent strokes in a flash are measured as one, the long

duration of an infrasound signal (10 s) resulting in missed discharges, and thirdly the low signal to noise ratio in signals that travel through the lightning storm [12]. Distant lightning (>60 km) could not directly be associated with infrasound signals [11].

Chum et al. (2013) match infrasound data with flashes within 40 km distance consisting of grouped electromagnetic EUCLID lightning detections close together in space and time. Infrasound pulses were successfully associated with CC discharges, where the ability of infrasound to detect CC lightning was shown. The research also supports the theory that the infrasound is produced by a change in the electrostatic field (and resulting sudden contraction in the cloud base) instead of the heating channel [8].

In previous papers electromagnetic lightning detections of CC as well as CG have successfully been linked to infrasound data at relatively close distances. Cross bearing infrasound has the potential to generate localizations of lightning. In this chapter a case study is examined to determine the relative performance of infrasound detections with electromagnetic lightning detection systems.

First, infrasound localizations from two stations spaced 22.3 km apart are compared to the localizations from electromagnetic detection. Secondly, the incoming infrasound signals at the individual arrays are linked to lightning localizations as detected by three electromagnetic detection systems. The predisposition towards CC/CG detection in infrasound is evaluated. Special focus lies on the use of elevation angle as measured in infrasound to determine the height of CC and CG sources.

6.2 Method

6.2.1 Infrasound detection

For this research two infrasound array stations in the Netherlands located at 22.3 km distance from each other are used. The dbn array is located at de Bilt and consists of 6 instruments in a triangular shape spaced approximately 30 m apart. The cia array located at Cabauw also consists of 6 instruments that are placed in a pentagonal shape spaced approximately 100 m apart (see figure 20).

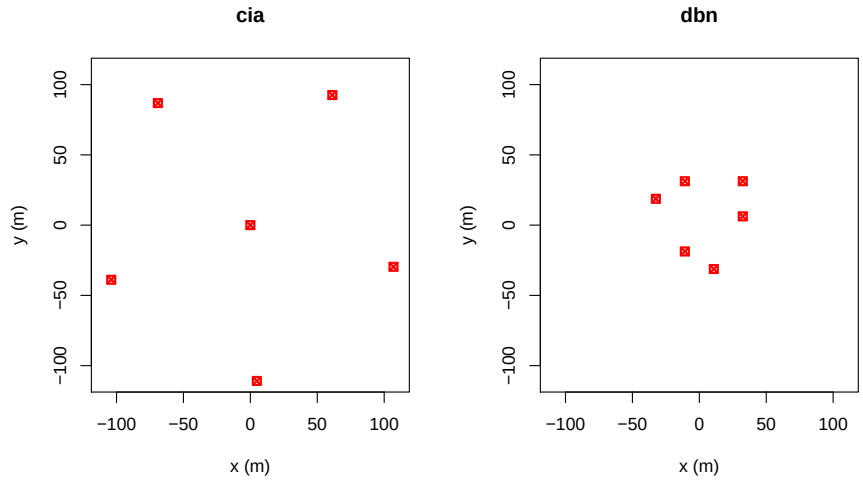


Figure 20: Infrasound array lay-out of dbn station in de Bilt and cia station at Cabauw.

With the time aligned incoming signals of all instruments a 'best beam' is found that gives the direction and velocity (energy) corresponding to the maximum coherence of the time aligned signals. Different beams directions and trace velocities are evaluated, defined by a square slowness grid from -0.01 s/m to 0.01 s/m existing of 150 steps (22500 beams). Coherency is determined by Fisher statistics, in which the Fisher ratio is used as a measure for coherency [14].

The distance between array instruments determines the range of infrasound that can be detected. If the wavelength is too long compared to the distance between instruments (i.e. low frequency), the infrasound signal can be detected but beamforming, or other array processing, no longer works because of the weak/no obervable time difference between instruments. If the wavelength is too short compared to the distance between instruments (i.e. high frequency), there is the chance of aliasing, meaning that non-related signals are correlated, which results in wrong best beam solutions. The instruments in dbn are spaced more closely together than the instruments in cia. Besides the differences in detection range, this shorter distance in dbn also means that there is less atmospheric influence on the signal during travel time between elements in dbn station. The array data contains arrival time of events, Fisher ratio (measure of signal-to-noise), back-azimuth (direction angle towards the signal with North as 0°) and the apparent velocity.

It is assumed that the infrasound propagates with a constant effective speed of sound ($Ceff$) from source to the array in a direct path between source and array. This is assumed because the high frequency infrasound signal from lightning travels shorter distances, and by focusing on direct propagation the attenuation is minimized. If the apparent velocity ($Capp$), i.e. the velocity of the signal as if projected on the horizontal surface, exceeds the effective speed of sound, the signal travels from an elevated angle. This angle is calculated with the following equation.

$$\alpha = cos^{-1}(\frac{Ceff}{Capp})$$

However, wind and atmospheric effects like pressure and temperature influence the effective sound speed as well as distort the signal itself. It is assumed that wind and atmospheric effects on propagation are integrated in the apparent velocity.

Because of the difference in array lay-out, dbn is more sensitive to the high infrasound frequencies that originate from lightning, making the raw dbn dataset larger than cia. The raw datasets are filtered to minimize noise with minimal values for apparent sound speed and Fisher ratio. The threshold for apparent sound speed is 300 m/s, which is lower than the assumed tropospheric effective sound speed of 340 m/s (for atmospheric temperature of 13.5° C) to account for wind, with wind speeds up to 40 m/s. The Fisher ratio (F) is related to signal to noise ratio with the relation: $SNR^2 = (F-1)/C$ [3] where SNR is the signal to noise ratio and C is the number of instruments in the array. For a signal to noise ratio of 1, the Fisher ratio for the array stations in Cabauw and de Bilt should be 7. The actual chosen Fisher ratio thresholds are 3 for cia and 6 for dbn, in order to remain a number of events comparable to the number of detections in the electromagnetic lightning detection systems. These different thresholds likely result in a cia dataset with more non-lighting originated signals than in the dbn dataset.

6.2.2 Lightning localization with infrasound arrays

The apparent velocity is the projected horizontal velocity of the incoming wavefront across the array. It describes the horizontal velocity of the signal traveling towards the array from the location of the source as projected on the earth's surface. This 2D approach is used to localize lightning events from cia and dbn data. A localization is defined if the back-azimuths of the detected signals at dbn and cia intersect and the following equation is true.

$$ta_{cia} - \frac{r_{cia}}{Capp_{cia}} = ta_{dbn} - \frac{r_{dbn}}{Capp_{dbn}} + \delta t$$

Here ta is the signal arrival time at that station, r is the distance between array station and back-azimuth intersection, $Capp$ is the apparent velocity and δt is the time offset. If the stroke time calculated from both arrays does not deviate more than the maximum allowed $\pm\delta t$, the stroke time is assumed to be the mean of the two outcomes. Note that it is assumed that the signal travels with the constant $Capp$ value, as determined at the infrasound array, from source to receiver.

As the location is obtained by the intersection of back-azimuths, maximum location accuracy is obtained when the intersecting back-azimuths are perpendicular. Similarly, intersections at large distances are determined with a lower location accuracy. If multiple combinations between cia and dbn qualify, the combination with the smallest time offset is chosen. No double matches can be made.

All combinations of incoming infrasound signals within 74.3 seconds of one another are evaluated, as this is the maximal arrival time difference of a signal between dbn and cia at the slowest allowed apparent velocity ($\frac{22{,}300\ \mathrm{m}}{300\ \mathrm{m/s}}$). The localization may be the projection of a source at higher altitudes on the earth's surface. The matches result in a dataset of 2D locations and stroke times as found by the infrasound data in dbn and cia.

6.2.3 Matching electromagnetic localizations with single-array infrasound signals

The localizations in FLITS, KLDN and ATDnet can be matched with the infrasound data from respectively dbn and cia in a similar manner. Infrasound signals are less accurate detected when the signal has travelled farther distances, so detection areas of increasing radius up to 120 km are investigated. An infrasound event is compared to all electromagnetic detections occurring in the previous 400 seconds, i.e. the time it takes an infrasound signal to travel 120 km at the lowest allowed apparent velocity. A match is defined if the calculated and observed back-azimuth angle deviates less than an allowed angle offset (δdeg) and the following equation is true.

$$ta - \frac{r}{Capp} = ts + \delta t$$

Here ta is the arrival time of the signal at the array station, r is the distance between the discharge location as found by FLITS/KLDN/ATDnet and the array station, ts is the stroke time as found by FLITS/KLDN/ATDnet and δt is the time offset that lies between the maximum $\pm\delta t$. The apparent velocity is used as it is assumed that the 2D location is the projection on the earth's surface if the lightning source is at an altitude.

If multiple combinations comply to the criteria, the combination with the smallest time offset is chosen. No double matches are allowed.

6.3 Results

The case of July 27th 2013 between 9 and 11 UTC was chosen for its high amount of discharges in the region around the dbn and cia station. The storm moved in the Northeast direction over the two stations. During this time interval the number of strokes detected in the area within 120 km of the point between cia and dbn was 4,736 by FLITS, 3,749 strokes by KLDN and 4,922 strokes by ATDnet.

Multiple strokes could be part of the same lightning flash. By grouping the lightning strokes based on temporal and spatial distance, the datasets can be converted into datasets of flashes and can thereby be more accurately compared to other flash datasets [13] [10] [19]. The resulting number of flashes by grouping based on a spatial separation <15 km and a temporal separation <1 s within 120 km of the center becomes 2,967 for FLITS, 1,661 for KLDN and 2,677 for ATDnet.

The raw infrasound data has been filtered as described on page 50. The remaining number of events in the infrasound datasets (including the 400 seconds after 11 UTC to allow for delayed arrival time) is 932 for cia and 2,414 for dbn.

6.3.1 Cross bearing cia with dbn

First the allowed time offset between signal arrival time at cia and dbn is determined. In order to do so, the time offsets are calculated for any match that is found with the maximum time offset of 25 s. The result is plotted in the histogram in figure 21.

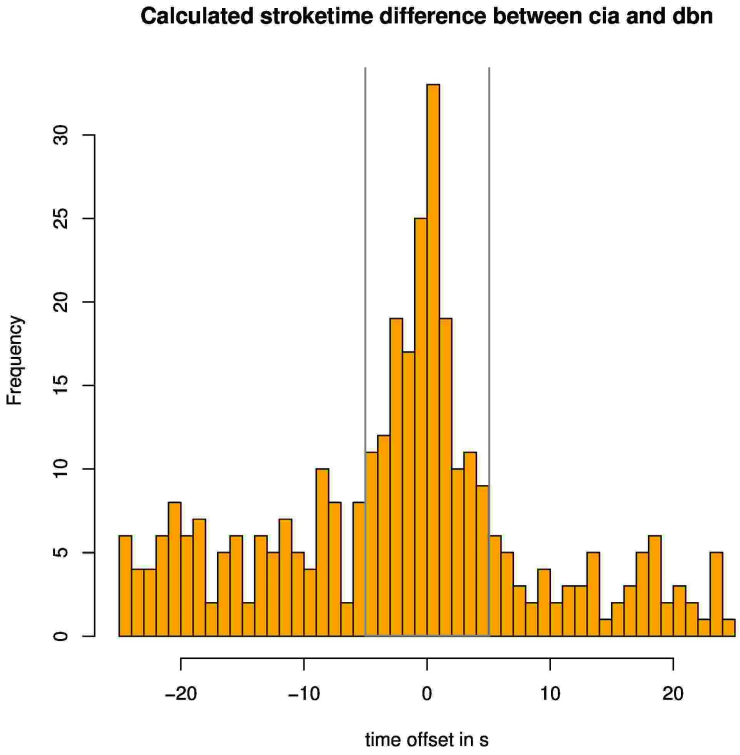


Figure 21: Arrival time difference at cia station compared to arrival time at dbn of all array matches found with maximum allowed time deviation of 25 s, with vertical lines at ± 5 s.

A perfect propagation model would have zero time offset. However, the model used and its assumptions introduce propagation errors and uncertainties which can lead to false matches. The time offset histogram in figure 21 can help define sensible limits of good matches. Figure 21 shows that most matches have a time offset within ± 5 s. A broader matching range will generate more false matches. The maximum time offset is therefore set to 5 s. The 200 matches found by cross bearing of cia and dbn with this maximum $\pm\delta t$, are given in figure 22. See the detections of the electromagnetic lightning detection systems in figure 23 for comparison.

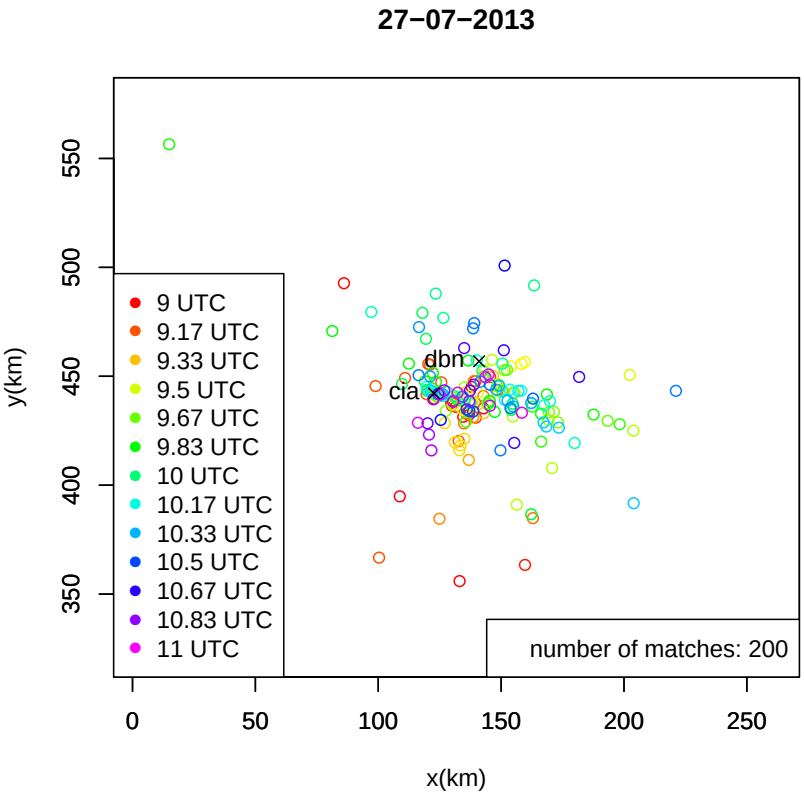


Figure 22: Matches found by cross bearing cia with dbn infrasound data of 27-07-2013 between 9 and 11 UTC, where color represents time of discharge.

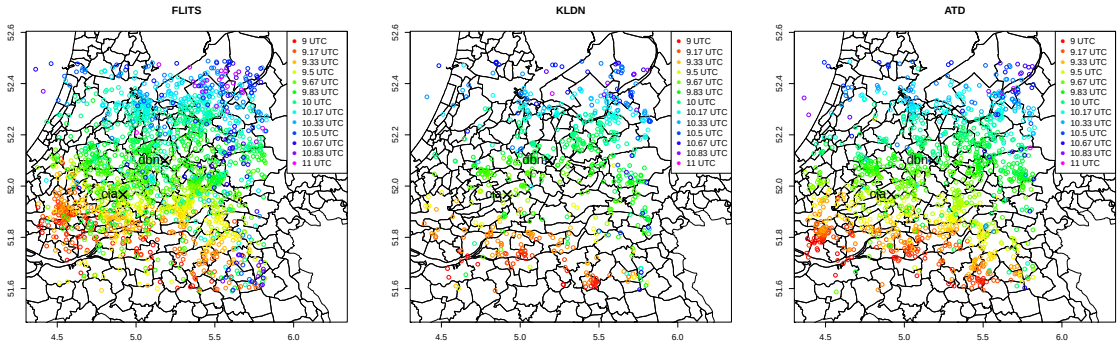


Figure 23: Lightning discharge detections of respectively FLITS, KLDN and ATDnet on 27-07-2013 between 9 and 11 UTC with the same color scheme as figure 22.

	ELDS reference: yes	ELDS reference: no
Infrasound dataset: yes	hits	false alarms
Infrasound dataset: no	misses	correct negatives

Table 5: Contingency table.

Assuming that the electromagnetic lightning detection systems (ELDS) accurately represent the lightning discharges, we can calculate the relative performance of the infrasound dataset with the probability of detection equation, based on the values in the contingency table 5.

$$\text{POD} = \frac{\text{hits}}{\text{hits} + \text{misses}}$$

The flash datasets are used in this comparison, and matches ('hits') are defined if two detections occur within 5 seconds and 15 km of each other. The POD is calculated for areas of ever increasing radius from the center point between the dbn and cia stations.

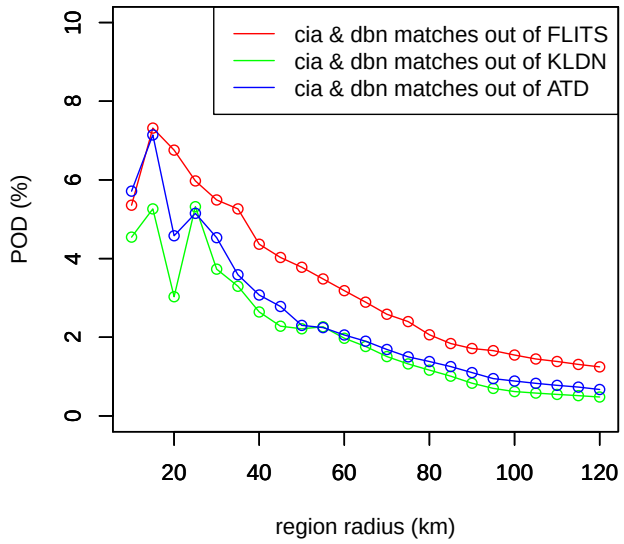


Figure 24: POD of cia and dbn matches with FLITS, KLDN and ATDnet flashes at ever increasing area size around the center point between cia and dbn station.

FLITS	yes	no
yes	28	172
no	482	-

KLDN	yes	no
yes	5	195
no	129	-

ATDnet	yes	no
yes	15	185
no	316	-

Table 6: Contingency tables of FLITS, KLDN and ATDnet with infrasound dataset in 30 km radius area around center point.

Figure 24 shows that the POD values of cia and dbn matches out of FLITS, KLDN and ATDnet flash datasets are highest at closer distance to the station. Note that the POD values on the smaller areas are based on a smaller number of detections compared to the calculations over the entire region and are therefore not as robust. To give an impression on how well the datasets match, the contingency tables 6 with FLITS, KLDN and ATDnet are given below. The considered area is the area with radius = 30 km from the center point.

Taking into consideration the equations for False Alarm Ratio FAR = $\frac{\text{false alarms}}{\text{false alarms + hits}}$ and POD = $\frac{\text{hits}}{\text{hits + misses}}$, we see that FLITS as the reference truth results in the highest POD (5.5% against 3.7% in KLDN and 4.5% in ATDnet) and lowest FAR (86% against 97.5% in KLDN and 92.5% in ATDnet). KLDN as reference truth gives the lowest POD and highest FAR, and is therefore the poorest match with the cia dbn dataset.

6.3.2 POD of individual arrays

The cia station detects significantly less events than dbn. In order to see how well they each match with the electromagnetic systems, the individual POD is calculated. For this purpose, the infrasound datasets are compared to FLITS, KLDN and ATDnet respectively with an allowed deviation in observed vs calculated back-azimuth angle of 10° and maximum time difference of 5 s. Note that the angle deviation results in lower location accuracy at larger distances. The matching is done with the stroke datasets. This is because the infrasound detection is on a larger timescale than electromagnetic lightning detection, and will not match with multiple strokes of the same flash [12]. The localizations are represented in figure 25.

FAR (%)	FLITS	KLDN	ATDnet
cia	75.1	81.9	76.3
dbn	72.7	81.1	74.3

Table 7: FAR values of cia and dbn out of FLITS, KLDN and ATDnet respectively.

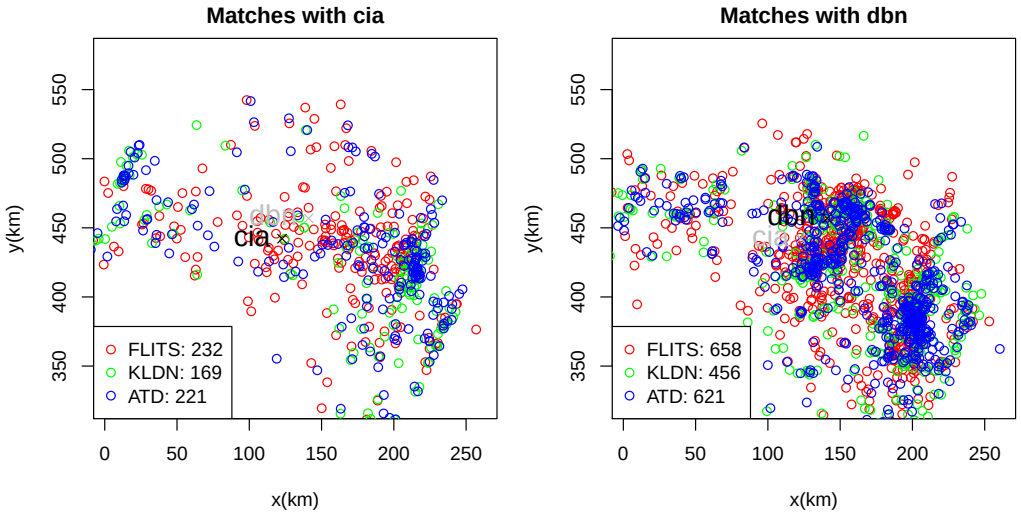


Figure 25: Matches found between cia and dbn with ungrouped datasets of respectively FLITS, KLDN and ATDnet on 27-07-2013 between 9 and 11 UTC.

The FAR and the POD give a quantitative description how well cia and dbn match with lightning detection systems. The FAR results given in table 7 show that approximately 3 out of 4 detections in cia and dbn are not attributed to a flash found by FLITS, KLDN or ATDnet.

The POD values describe the percentage of lightning events also detected by cia and dbn. Because of the timescale of the infrasound detection, all matches from figure 25 are considered flashes. This flash dataset is then compared to the total flash datasets of FLITS, KLDN and ATDnet to calculate the POD values.

As the accuracy of infrasound detection decreases for localizations at large distances, and the use of an allowed degree deviation as matching criterion becomes less accurate further away from the array station, the POD is calculated for area sizes of increasing radius from the center point between the

array stations. This is shown in the upper left graph in figure 26. The same is done when only the CC or CG flashes in both datasets are considered in the bottom two graphs of the same figure.

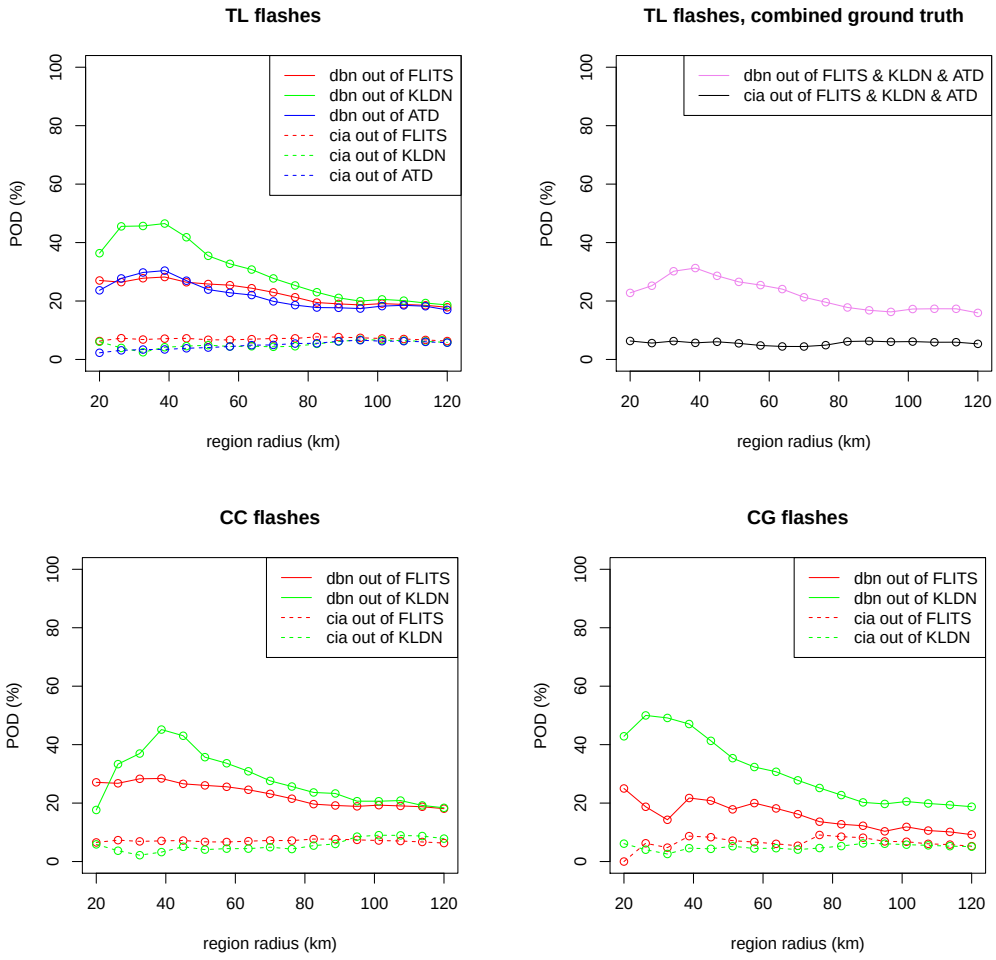


Figure 26: Area size dependent values of POD for TL, CC and CG flashes of cia and dbn with respectively FLITS, KLDN and ATDnet on 27-07-2013 between 9 and 11 UTC. The graph in the right upper corner describes the POD values when the reference is a dataset of discharges that are found by all three datasets.

The FLITS dataset of this storm has a CC:CG ratio of approximately 35:1. KLDN has a CC:CG ratio of 1:3. These detection ratios are constant over the whole region. Thus the POD calculations with KLDN CC flashes and especially FLITS CG flashes are based on a small dataset, and are there-

fore less robust than the other POD values.

6.3.3 Source height determinations

Matching infrasound array data with localizations from electromagnetic lightning detections systems provides a distance and an elevation angle (α) from the array to the CC and CG discharges. Here α was calculated as described in section 6.2.1 with $Ceff = 340$ m/s. With the assumption that the infrasound signal travels in a straight path the height of the source can be calculated with $h = r * tan(\alpha)$ where h is the height and r is the 2D distance between array and the discharge location as projected on the earth's surface.

Special distinction is made between CC and CG discharges in order to see if the origin of the infrasound location differs between lightning types.

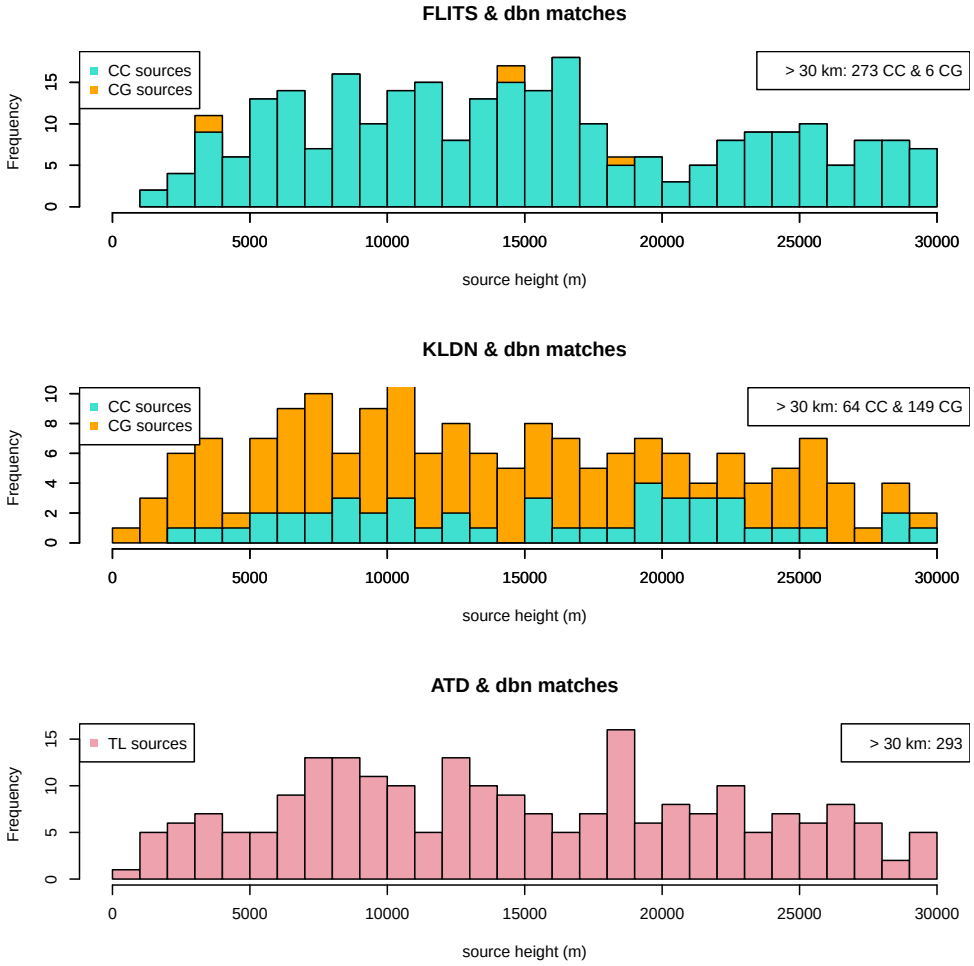


Figure 27: The heights of matches between dbn and FLITS, KLDN and ATDnet. In KLDN and FLITS the stacked CC and CG discharges are represented with different colors. Values above 30 km altitude are mentioned in the top right corner.

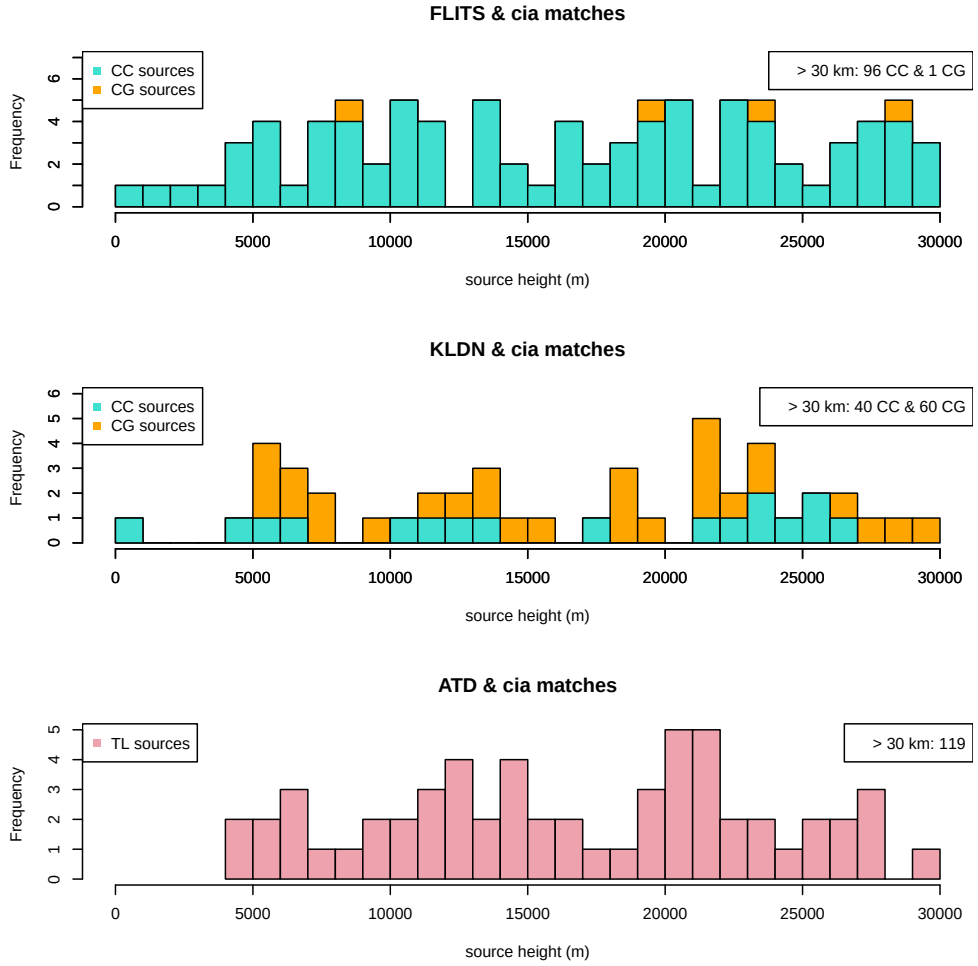


Figure 28: The heights of matches between cia and FLITS, KLDN and ATD-net. In KLDN and FLITS the stacked CC and CG discharges are represented with different colors. Values above 30 km altitude are mentioned in the top right corner.

Figures 27 and 28 show the calculated heights of infrasound sources. There are significantly more matches with FLITS of type CC, though the CG heights vary over the same range as the CC heights. CC and CG also do not differ in height from each other in the matches between the arrays and KLDN. Also, the majority of the calculated altitudes exceed by far the expected source height, namely the cloud base. Values of several tens of kilometers are highly unlikely for lightning related infrasound sources, thus clearly indicating the shortcomings of the linear propagation model.

6.4 Discussion/conclusions

The cross bearing of the array stations dbn and cia provides a dataset of possible discharges. The number of discharges is far less than the number of lightning detections from electromagnetic systems, resulting in a low POD. This is mainly because the cia array detects far less events than dbn, which is likely caused by the large distance between array elements in cia. For a better detection of infrasound signals from lightning discharges, an array with more closely spaced instruments is required. The highest POD (5.5%) and lowest FAR (86%) were found with FLITS as reference. The cross bearing compared with KLDN generated the poorest match.

The individual matches between dbn and the flashes detected by the electromagnetic systems results in far higher POD values, especially for KLDN with a POD over 40%. Highest POD values were found in the region with a 40 km radius. However, the FAR values are also highest with KLDN ($\approx$ 81%). As cia consists of a lot less detections, the cia POD values are low.

Questions can be posed regarding the accuracy on CC and CG distinction by KLDN and FLITS. The number of CC detections by FLITS is very high. The calculated POD values are not significantly different for CC and CG flashes. If the source of infrasound would be the heating channel during a discharge, a lower infrasound source altitude would be expected from CG lightning. The heating channel in CC lightning is generally at higher altitudes.

Both CC and CG discharges have been attributed with high infrasound sources. However, these results are not conclusive regarding the question whether the cloud base is the location of infrasound sources for all lightning types. The height calculation seems too simplistic, as the outcomes are too high to be realistic. Instead of assuming linear propagation of the infrasound signal, a more advanced wave propagation model is required. This will result to a smaller time offset and lower false matching. Improved 3D localization makes a more accurate heigth estimate possible, as well as probably result in a larger overlap with the electromagnetic lightning detections.

Though these results rely on a single case study, the following conclusions can be drawn. Lightning detection with infrasound arrays could be possible for two arrays at close distance ($\approx$ 25 km) with instruments spaced not too far apart ($\approx$ 30 m). The first step towards lightning detection with infrasound is to improve the wave propagation model used in the analysis.

Acknowledgements

I wish to thank various people for their contribution to this MSc project; my KNMI supervisors Dr. Hidde Leijnse, Dr. Maurice Schmeits and Ing. Hans Beekhuis, whose advice, enthusiasm and guidance have been indispensible throughout the project. I would like to express my very great appreciation to Dr. Läslo Evers and Ir. Pieter Smets for their assistance with the infrasound analysis. Advice given by Dr. Dieter Poelman has been a great help in the interpretation of the results. Assistance provided by Dr. Stéphane Pedeboy and Ir. Marc Bonnet of Météorage was greatly appreciated, as well as UK Met Office for providing the ATDnet dataset. I also wish to acknowledge the help provided by Dr. Aarnout van Delden, my supervisor from IMAU.

In addition, I would like to thank the people at KNMI and IMAU for their welcoming and constructive attitude towards me. Finally, I wish to thank my parents and friends for their support and encouragement throughout my study.

A Temporal variation in the datasets

A.1 Temporal distribution of flashes

The number of flashes is calculated for KLDN, FLITS and ATDnet are divided in bins of 1 hour in order to get insight in the distribution of flashes for each database. The chosen region is KOUW (as figures 7 and 8 show that all systems have good coverage here) over the period from 01-01-2010 to 31-10-2014, with a dt = 1 s and a dr = 15 km.

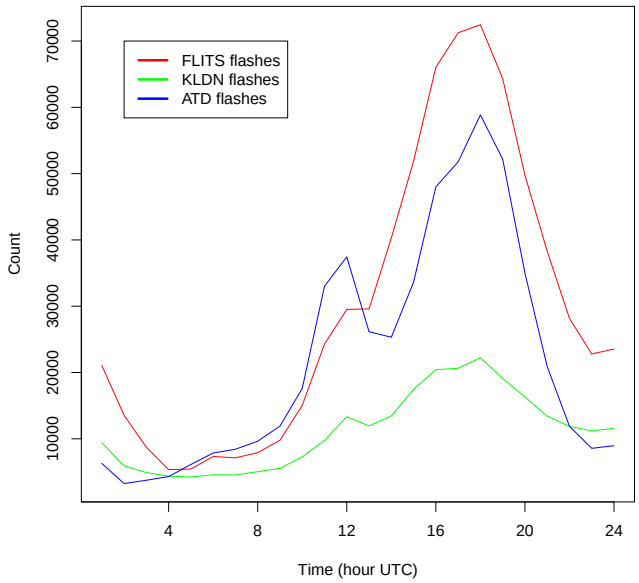


Figure 29: Flash count of FLITS, KLDN and ATDnet distributed in one-hour bins representing the hour (UTC) in which they occur. Strokes in the KOUW-region are grouped with dt = 1 s and dr = 15 km.

The detection system with most flash detections is FLITS. FLITS has the highest detection at almost all hours of the day, with the exception between 5 and 12 UTC, where ATDnet shows more detections.

KLDN is most constant in time and shows the least diurnal variation. The secondary peak at 12 UTC which is present in all three datasets is exclusively due to summer thunderstorms occurring around this time between June and August in the years 2010, 2011 and 2013. In winter seasons, light-

ning predominantly occurs later in the day.

ATDnet shows a diurnal decrease in performance at night, though detects more flashes than FLITS and KLDN during the secondairy peak at 12 UTC, a time in which the performance of ATDnet is not inhibited by variations in the ionosphere.

A.2 Stroke/flash ratio

The number of strokes and flashes are counted for FLITS, KLDN and ATDnet in 30 day bins. The values are the average number of strokes per flashes in the KOUW-region in 30 days.

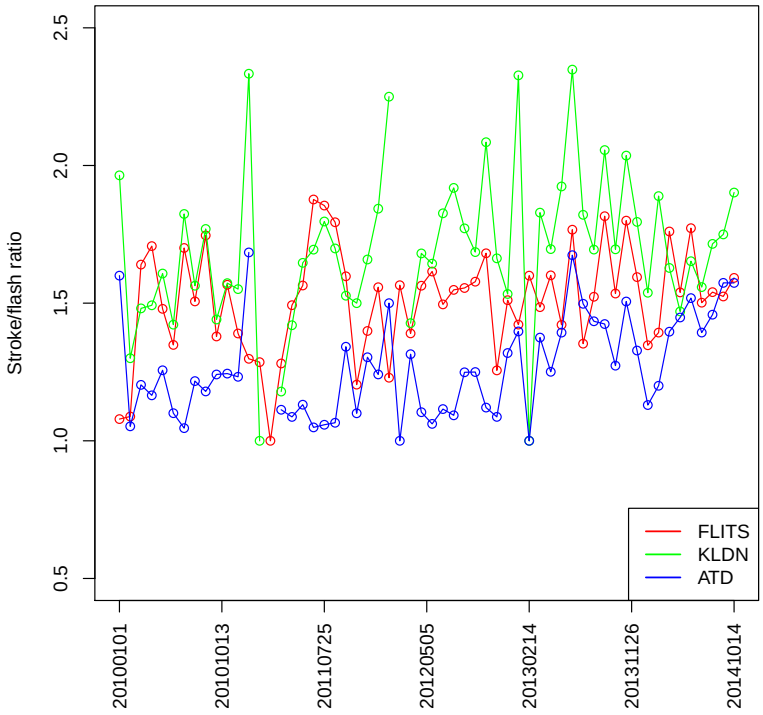


Figure 30: The ratio of the amount of strokes and flashes in periods of 30 days in the KOUW-region for FLITS, KLDN and ATDnet, with flash conversion that uses dt = 1 s and dr = 15 km. The graph shows no value in 2 30 day periods, where the accumulated detection of KLDN was zero.

KLDN and FLITS in figure 30 show similar average strokes per flash,

followed by ATDnet. This corresponds with the notion that ATDnet mostly only detects the first return stroke in a flash. There seem to be no dramatic systematic changes in stroke/flash ratio during the time span of the datasets.

A.3 Event count

The number of strokes accumulated in bins of 30 days is given over the whole period between 01-01-2010 and 14-10-2014 for FLITS, KLDN and ATDnet in the KOUW-region. Next, the number of flashes is calculated with dt = 1 s and dr = 15 km and given in the same figure. These values are given on a linear and logarithmic axis in figure 31. Changes in the (relative) detection counts over time would suggest a change in the network that influences the detection performance.

In the top panel in figure 31 there seems to be no obvious variations in time in the relative detection of the datasets. In thunderstorm seasons, FLITS detects a large number of strokes. From the bottom panels with logarithmic axes, it becomes clear that in periods of scarce lightning events FLITS detects more strokes and flashes than KLDN and ATDnet. In several instances KLDN and ATDnet detection counts are (near) zero while FLITS detects several tens of strokes or flashes. The shape of the stroke curves and the flash curves are similar in all datasets.

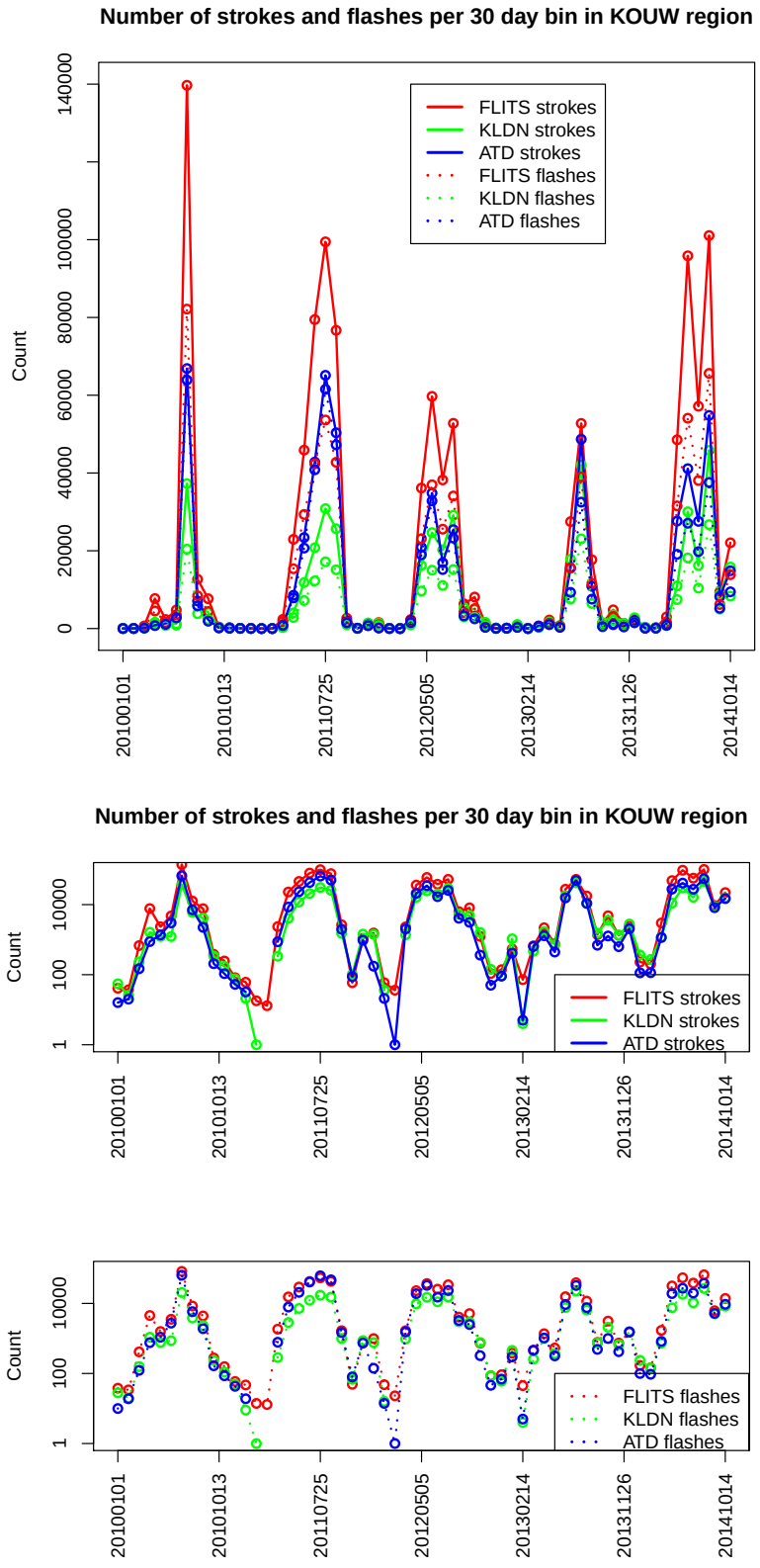


Figure 31: Number of strokes and flashes per 30 days in the KOUW-region in the period between 01-01-2010 and 14-10-2014 plotted on a linear axis (top) and a logarithmic axis (bottom).

B Temporal variations in POD

The POD values of any combination of FLITS, KLDN and ATDnet have been calculated for 30 day bins in the lightning season from May to September (i.e. lightning season) in the period between 01-01-2010 and 31-10-2014. In these calculations the total lightning flashes in the KOUW-region are matched with dr = 15 km and dt = 1 s. Over the whole timeline, 8 days are excluded because of missing days in one of the datasets. The POD calculations of both TL and CG flashes are calculated, see figures 32 and 33.

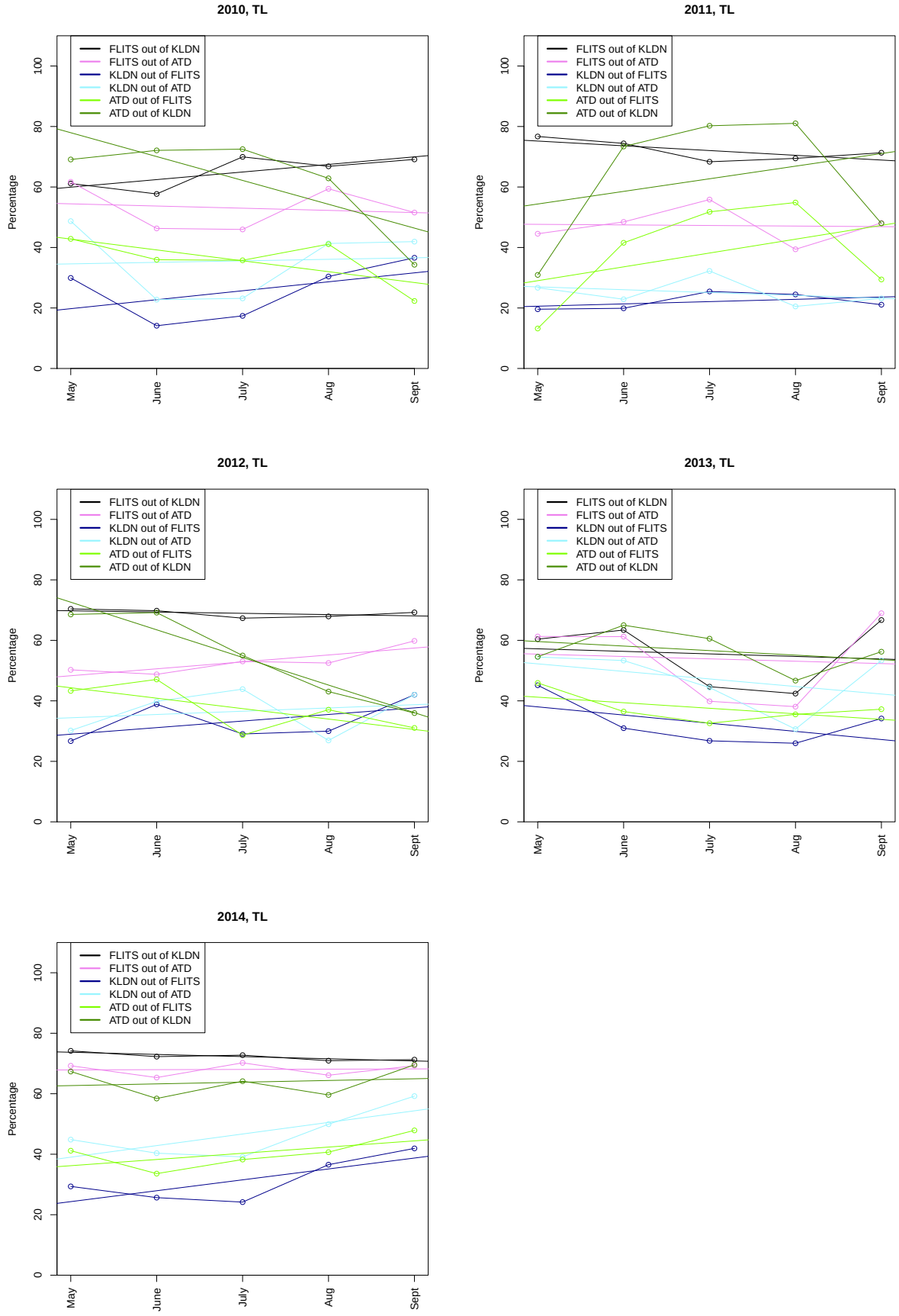


Figure 32: POD of TL flashes calculated per 30 days with dr = 15 km and dt = 1 s in the KOUW-region.

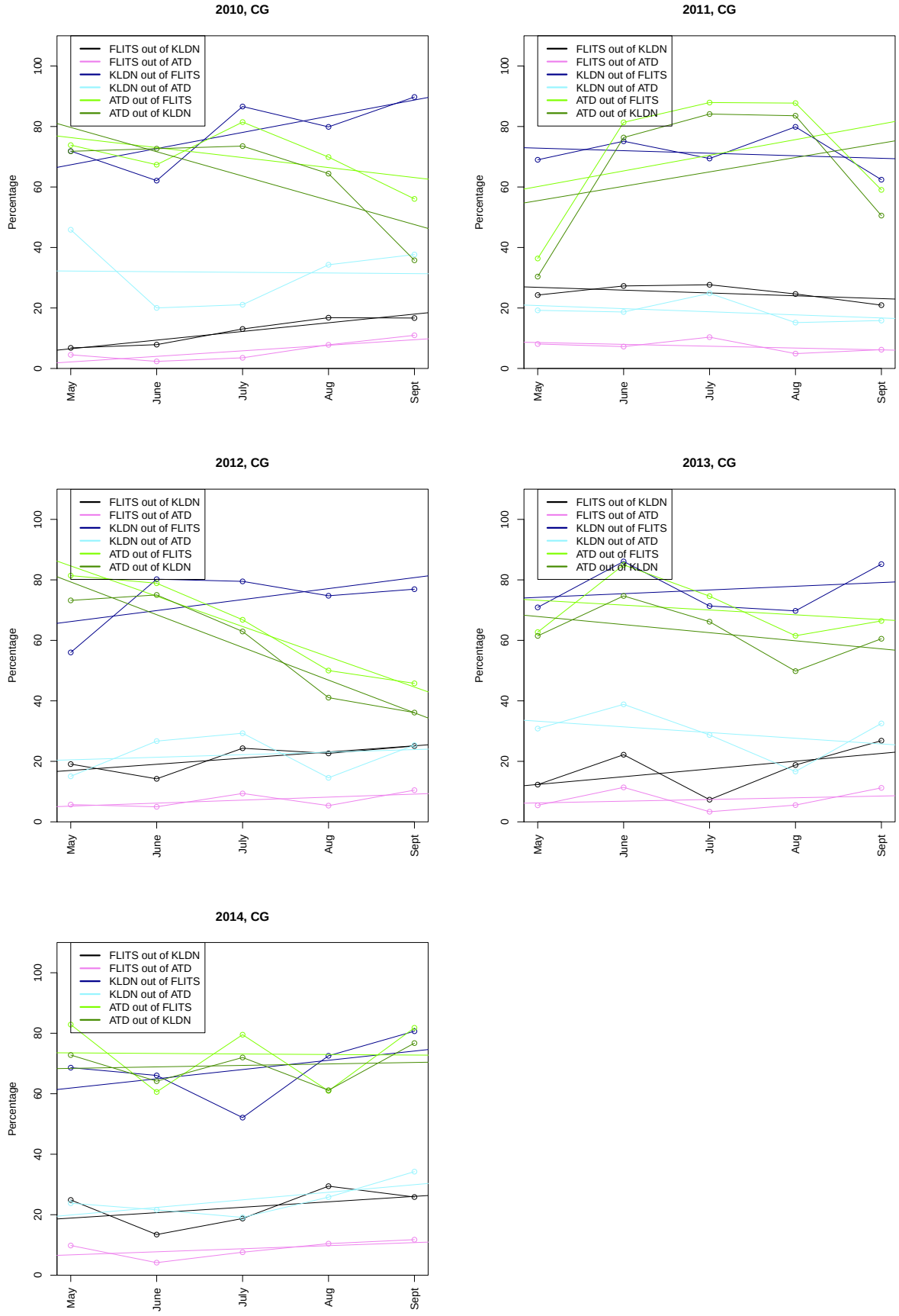


Figure 33: POD of CG flashes calculated per 30 days with dr = 15 km and dt = 1 s in the KOUW-region.

C Meteorological conditions during severe weather peaks

In the following two plots, the meteorological data of the nearest station is plotted for all severe weather peaks. In figure 34 these severe weather peaks are all moments where FLITS detects at least 500 TL strokes within 5 minutes in an area of 50 * 50 km^2. In figure 35 the severe weather peaks are all moments where KLDN detects at least 70 CG lightning strokes within 5 minutes in an area of 50 * 50 km^2.

In figure 34 the hourly precipitation amount (RH) and the maximum wind gusts (FX) are given for the severe weather peaks between 01-01-2010 and 31-10-2014. The values of the nearest station for the hour in which the severe weather peak occurs are used. Only the unique values are represented in the scatter plot (i.e. if the storm moves closer to a different station or occurs in multiple hours, all relevant FX and RH values are given). All hourly values in de Bilt are plotted as a reference. Labels represent date, time (UTC) and station number of (part of) the severe weather peak. Note: one outlier is excluded in the graph; at 28-06-2011, 19 UTC, station 356 (Herwijnen) registers FX of 20 m/s and RH of 79.0 mm.

In figure 35 information is added related to whether or not the severe weather peaks match the severe weather peaks of FLITS in figure 34; Similarly as in figure 34, the RH and FX values of the nearest station in that particulate hourblock is given with colored dots. Red dots are all KLDN severe weather moments that do not match with a severe weather moment by FLITS with the 500 TL stroke criterion (the colored dots in figure 34). If both KLDN and FLITS detect a severe weather peak with their respective stroke criteria, the dot is colored in blue. The date in the label is violet if it matches a date on which FLITS detects a severe weather peak as well. If the date is colored black, KLDN has detected a severe weather peak on a day that FLITS does not. Labels are only given for RH larger than 4.0 mm and FX larger than or equal to 15.0 m/s.

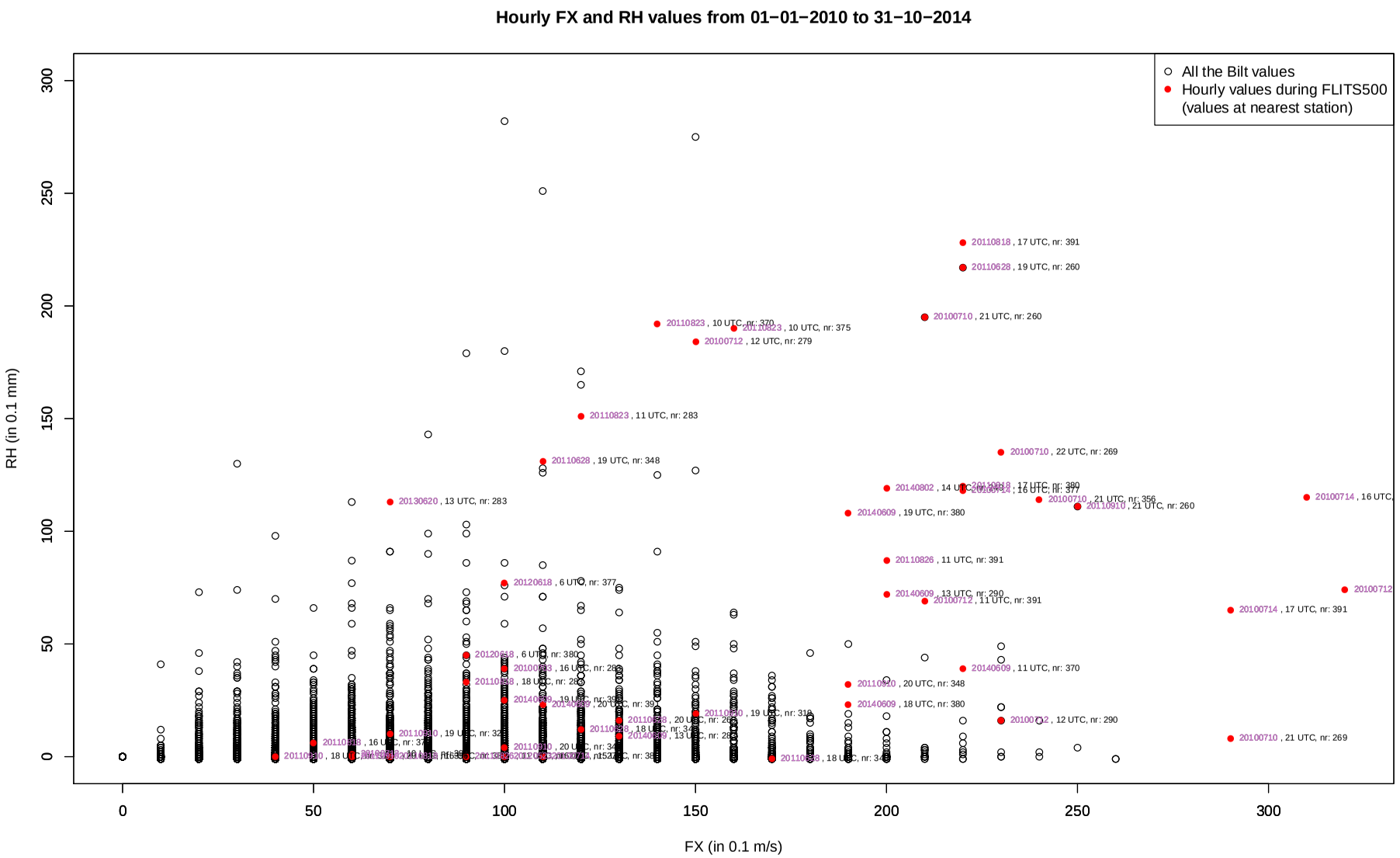


Figure 34: FLITS 500 TL severe weather peaks.

From figure 34 it becomes clear that while some severe weather peaks occur together with large values of FX and RH, this is not always the case. This could be because the station is too far from the storm to accurately capture the meteorological variables of the storm. In some cases, a part of the storm is captured in one hour block, but not in the other. An example of this is 18-08-2011, where station 391 registers high RH and FX values at 17 UTC, though the nearest stations at 18 and 19 UTC register FX and RH values that do not differ much from everyday values in de Bilt. Even though some severe weather moments based on the 500 TL stroke criterion within 5 min and 50 * 50 km^2, i.e. the FLITS500 reference, are missed in the meteorological station dataset, almost all high RH and FX values correlate with a severe weather peak.

Figure 35 shows that the meteorological values FX and RH of KLDN severe weather moments are also not always distinctive from the hourly values without storms. Also, the red dots in the upper right area in the scatter plot indicate that a large number of storm events are found that do not coincide with the FLITS500 reference. All days that KLDN detects severe weather (including the outlier) together with a RH larger than 4.0 mm and a FX larger or equal to 15.0 m/s, add up to 15 unique days. 9 of these days match the 14 unique detection days of FLITS500. Therefore, based on the criterion of at least 70CG strokes together with more than 4.0 mm precipitation in the hourblock, and wind gusts of at least 15.0 m/s for KLDN, almost the same amount of severe weather detection days are found in the historical dataset, though not completely the same days as found by FLITS with the original weather warning criterion.

The FX and RH threshold values mentioned are low compared to true severe weather, which entails wind gust speeds of 20-25 m/s and 30-50 mm of precipitation. As discussed in section 4.3, there is no exclusive relation between severe weather and hourly values of FX and RH at the nearest measuring station. This is likely because storms are sometimes missed by the measuring stations. The use of radar could offer improved measurements of current meteorological conditions. If a weather alert was based on high values of FX, RH and intense lightning based on and/or criteria instead of and/and criteria, a more complete alert can be given based on dangerous weather conditions in a broader sense than lightning only.

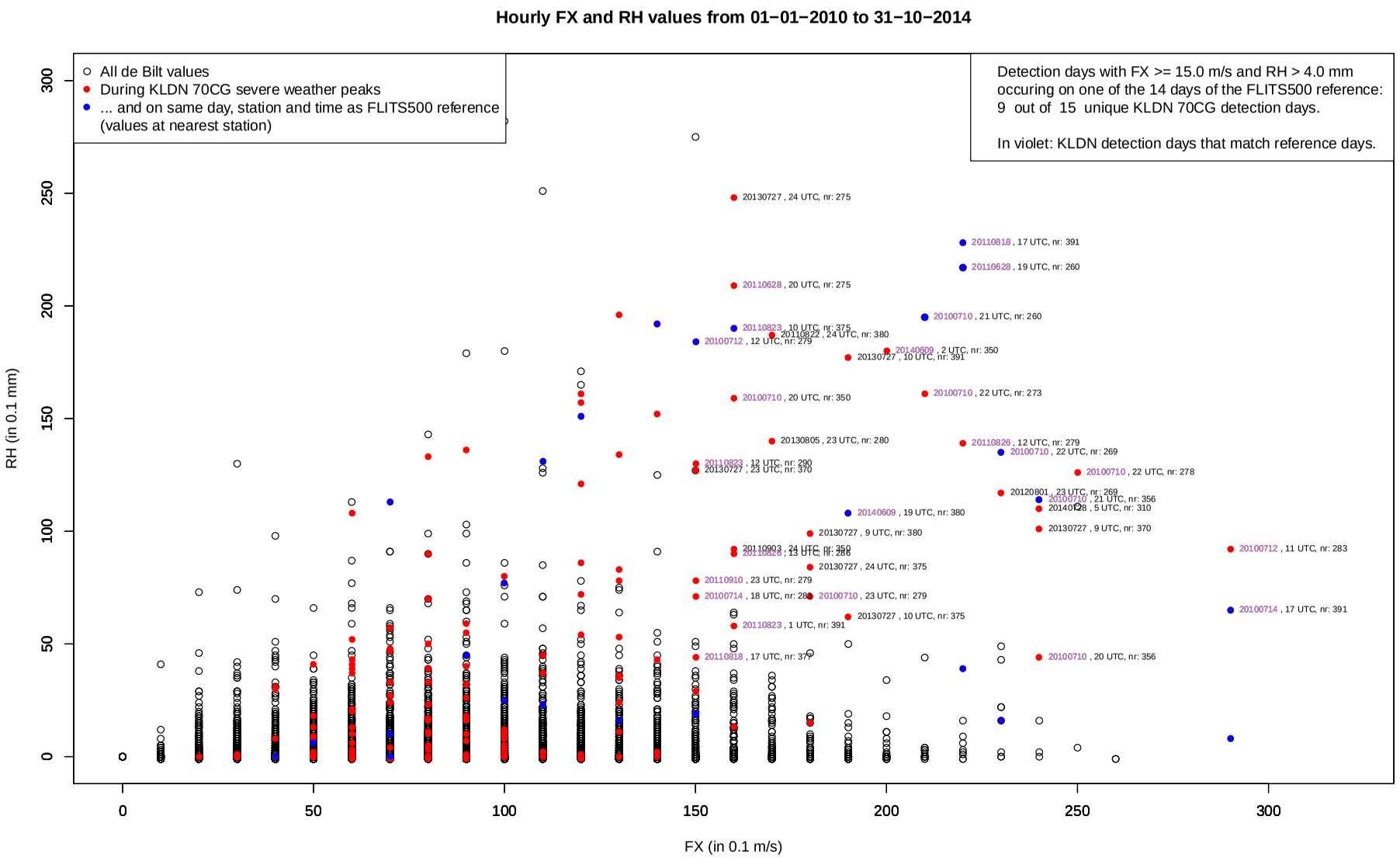


Figure 35: KLDN 70 CG severe weather peaks.

D Time bin size dependency of LAD and FAR

The lightning activity detection is a measure of the agreement between detection systems on the occurrence of any lightning activity. This agreement increases for longer matching period. Figures 36 and 37 represent the importance of time bin size for LAD and FAR. TL and CG flashes in the KOUW-region are considered.

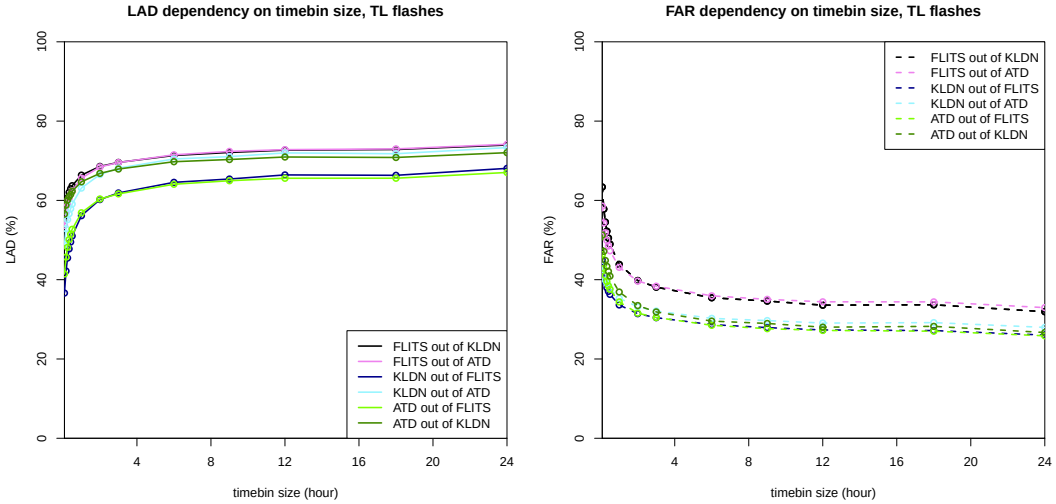


Figure 36: LAD (left) and FAR (right) values in KOUW-region in 10 * 10 km^2 boxes over the period between 01-01-2010 and 31-10-2014 for various values of time bin size, where TL flashes are considered.

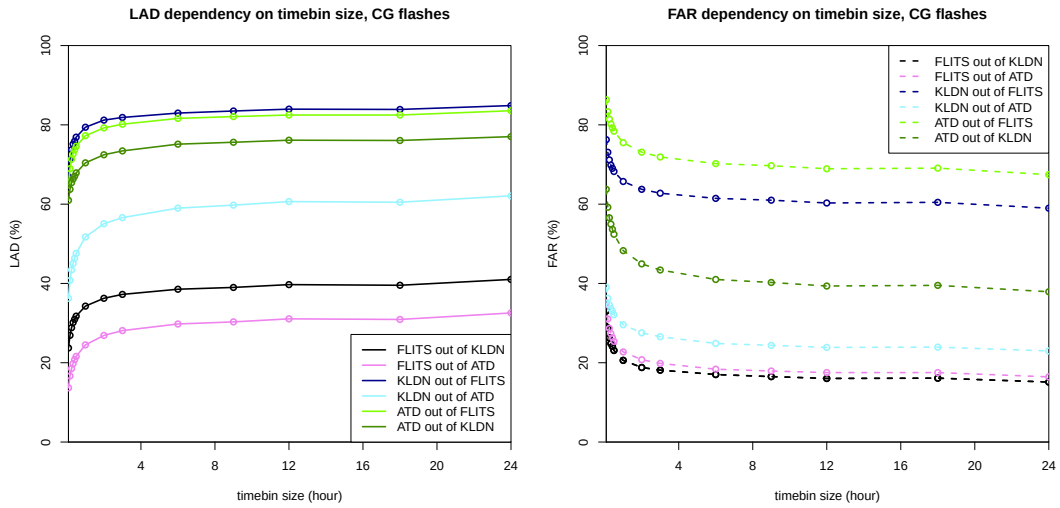


Figure 37: LAD (left) and FAR (right) values in KOUW-region in 10 * 10 km^2 boxes over the period between 01-01-2010 and 31-10-2014 for various values of time bin size, where CG flashes are considered.

As expected, from figures 36 and 37 we see a decrease in FAR and an increase in LAD for larger time bin sizes. Over a longer period, it is more likely to find a match in the 10 * 10 km^2 area box. This increase/decrease is no longer significant for a time bin size larger than 3 hours. This suggests that a storm generally generates lightning activity for a period of 3 hours in a certain area.

References

[1] Schade in beeld: Onweer en bliksem. https://www.nn.nl/Over-NationaleNederlanden/Pers-1/Persbericht/Schade-in-beeld-Onweer-en-Bliksem.htm. Accessed: 2014-12-01.

[2] Werkinstructie elektrische ontladingen op EHAM en EHRD. Retrieved from http://cwk-info.knmi.nl/documentatie/werkinstructies/Werkinstructie%20elektrische%20ontladingen%20op%20EHAM%20en%20EHRD.doc on 21-01-2015, July 2014.

[3] J.D. Assink, L.G. Evers, I. Holleman, and H. Paulssen. Characterization of infrasound from lightning. *Geophysical Research Letters*, 35(15), 2008.

[4] Y. Baba and V.A. Rakov. *Present understanding of the lightning return stroke*. Springer, 2009.

[5] W.H. Beasley, T.M. Georges, and M.W. Evans. Infrasound from convective storms: An experimental test of electrical source mechanisms. *Journal of Geophysical Research*, 81(18):3133–3140, 1976.

[6] H. Beekhuis. Sensoren en systemen, informatie over het bliksemdetectie systeem.

[7] R. Boonstra. Validation of SAFIR/FLITS lightning detection system with railway-damage reports. Technical report, KNMI, IR 2008-05., Available from KNMI, 2008.

[8] J. Chum, G. Diendorfer, T. Šindelářová, J. Baše, and F. Hruška. Infrasound pulses from lightning and electrostatic field changes: Observation and discussion. *Journal of Geophysical Research: Atmospheres*, 118(19):10–653, 2013.

[9] A.J. Dessler. Infrasonic thunder. *Journal of Geophysical Research*, 78(12):1889–1896, 1973.

[10] C. Drüe, T. Hauf, U. Finke, S. Keyn, and O. Kreyer. Comparison of a safir lightning detection network in northern Germany to the operational BLIDS network. *Journal of Geophysical Research: Atmospheres (1984–2012)*, 112(D18), 2007.

[11] T. Farges. *Infrasound from lightning and sprites*. Springer, 2009.

[12] T. Farges and E. Blanc. Characteristics of infrasound from lightning and sprites near thunderstorm areas. *Journal of Geophysical Research: Space Physics (1978–2012)*, 115(A6), 2010.

[13] U. Finke. Space-time correlations of lightning distributions. *Monthly weather review*, 127(8):1850–1861, 1999.

[14] B.S. Melton and L.F. Bailey. Multiple signal correlators. *Geophysics*, 22(3):565–588, 1957.

[15] J. Nash, N.C. Atkinson, E. Hibbett, G. Callaghan, P.L. Taylor, P. Odams, D. Jenkins, S. Keogh, C. Gaffard, and E. Walker. The new Met Office ATDNET lightning detection system. In *Proc. WMO Technical Conf. on Instruments and Observing Methods (Geneva, Dec.)*, volume 94, 2006.

[16] S. Noteboom. Processing, validatie, en analyse van bliksemdata uit het SAFIR/FLITS systeem. Technical report, KNMI, IR-2006-01., Available from KNMI, 2006.

[17] D.R. Poelman. On the science of lightning: An overview. *Royal Meteorological Institute of Belgium Report*, 526:56, 2010.

[18] D.R. Poelman. Present status and preliminary results of the Belgian lightning detection network. In *Proc. Sixth European Conf. on Severe Storms*, 2011.

[19] D.R. Poelman, F. Honoré, G. Anderson, and S. Pedeboy. Comparing a regional, subcontinental, and long-range lightning location system over the Benelux and France. *Journal of Atmospheric and Oceanic Technology*, 30(10):2394–2405, 2013.

[20] D.R. Poelman, W. Schulz, and C. Vergeiner. Performance characteristics of distinct lightning detection networks covering Belgium. *Journal of Atmospheric and Oceanic Technology*, 30(5):942–951, 2013.

[21] V.A. Rakov and M.A. Uman. *Lightning: physics and effects*. Cambridge University Press, 2003.

[22] M.J. Schmeits, K.J. Kok, D.H.P. Vogelezang, and R.M. van Westrhenen. Probabilistic forecasts of (severe) thunderstorms for the purpose of issuing a weather alarm in the Netherlands. *Weather and Forecasting*, 23(6):1253–1267, 2008.

[23] D.M. Suszcynsky, M.W. Kirkland, A.R. Jacobson, R.C. Franz, S.O. Knox, J.L.L. Guillen, and J.L. Green. FORTE observations of simultaneous VHF and optical emissions from lightning: Basic phenomenology. *Journal of Geophysical Research: Atmospheres (1984–2012)*, 105(D2):2191–2201, 2000.

[24] K.B. Thompson, M.G. Bateman, and L.D. Carey. A comparison of two ground-based lightning detection networks against the satellite-based lightning image sensor (LIS). *American Meteorological Society*, 31, 2014.

[25] D.S. Wilks. *Statistical methods in the atmospheric sciences*, volume 100. Academic press, 2011.